JMIR Medical Informatics

Impact Factor (2023): 3.1
Volume 5 (2017), Issue 4 ISSN 2291-9694 Editor in Chief: Christian Lovis, MD, MPH, FACMI

Contents

Original Papers

Viewpoints

Reviews

Original Paper

Computerized Decision Aids for Shared Decision Making in Serious Illness: Systematic Review

Anna Staszewska[1], BA, BSc; Pearl Zaki[1], BSc; Joon Lee[1], PhD

Health Data Science Lab, School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Joon Lee, PhD
Health Data Science Lab
School of Public Health and Health Systems
University of Waterloo
200 University Ave W
Waterloo, ON, N2L 3G1
Canada
Phone: 1 5198884567 ext 31567
Email: joon.lee@uwaterloo.ca

Abstract

Background: Shared decision making (SDM) is important in achieving patient-centered care. SDM tools such as decision aids are intended to inform the patient. When used to assist in decision making between treatments, decision aids have been shown to reduce decisional conflict, increase ease of decision making, and increase modification of previous decisions.

Objective: The purpose of this systematic review is to assess the impact of computerized decision aids on patient-centered outcomes related to SDM for seriously ill patients.

Methods: PubMed and Scopus databases were searched to identify randomized controlled trials (RCTs) that assessed the impact of computerized decision aids on patient-centered outcomes and SDM in serious illness. Six RCTs were identified and data were extracted on study population, design, and results. Risk of bias was assessed by a modified Cochrane Risk of Bias Tool for Quality Assessment of Randomized Controlled Trials.

Results: Six RCTs tested decision tools in varying serious illnesses. Three studies compared different computerized decision aids against each other and a control. All but one study demonstrated improvement in at least one patient-centered outcome. Computerized decision tools may reduce unnecessary treatment in patients with low disease severity in comparison with informational pamphlets. Additionally, electronic health record (EHR) portals may provide the opportunity to manage care from the home for individuals affected by illness. The quality of decision aids is of great importance. Furthermore, satisfaction with the use of tools is associated with increased patient satisfaction and reduced decisional conflict. Finally, patients may benefit from computerized decision tools without the need for increased physician involvement.

Conclusions: Most computerized decision aids improved at least one patient-centered outcome. All RCTs identified were at a High Risk of Bias or Unclear Risk of Bias. Effort should be made to improve the quality of RCTs testing SDM aids in serious illness.

(JMIR Med Inform 2017;5(4):e36) doi:10.2196/medinform.6405

KEYWORDS

decision making; decision aids; evidence-based medicine; user-computer interface; chronic disease

Introduction

Background and Significance

Shared decision making (SDM) is important in achieving patient-centered care, as it involves both the patient and the health care provider in medical decision making [1]. More than one reasonable treatment decision exists for the majority of medical decisions, and thus, patient involvement is of great value [2]. As patient involvement in treatment decisions increases, it is more likely that the treatment decision will be consistent with their preferences, lifestyles, and goals [3]. Competing values and perspectives between physicians and patients are often compounded by ineffective patient-provider communication regarding disease and goals of treatment [4]. Patients may choose treatment options based on erroneous

outcome expectations and misunderstanding of the disease. For example, in a study by Weeks et al [5], 69.0% (490/710) of patients with metastatic lung cancer and 80.9% (391/483) of patients with metastatic colorectal cancer did not understand that chemotherapy was not likely to be curative.

SDM tools such as decision aids are intended to inform the patients with regard to the risks, benefits, and trade-offs associated with a decision [6]. When used to assist in decision making between treatments, decision aids have been shown to reduce decisional conflict, increase ease of decision making, and increase modification of previous decisions [7,8]. Furthermore, it may be that patients who are informed about their disease because of the use of decision aids are less likely to choose nonbeneficial treatment [7].

Computerized decision aids can offer personalized evidence-based care, and if they are presented in an SDM capacity they can result in treatment decisions that respect the autonomy and preferences of the patient. Additionally, technological advances that use and process electronic health record (EHR) data may allow for the development of large-scale, low-cost assessments that can improve patient goals [9]. Computerized decision aids may provide additional benefits over traditional paper or video tools, as they have the potential for individualized content, a greater degree of interaction, and scalability [10].

A recent systematic review by Austin et al [11] synthesized the evidence for the use of decision aids in serious illness through the evaluation of randomized controlled trials (RCTs) and non-RCTs. However, the review by Austin et al [11], a relevant Cochrane systematic review on SDM tools by Stacey et al [8], and a review on the use of video decision aids in advanced care planning [12] do not focus on the ability of *computerized* decision tools to improve patient-centered outcomes. A focus on computerized decision aids is both timely and necessary because of the possibility of greater personalization of computerized decision aids, which is congruent with the goal toward individualized treatment plans. Additionally, computerized decision aids offer greater scalability over the traditional static decision aids [10]. Finally, a systemic move toward the digitization of health data allows for the natural progression of its use in decision support systems. Few other systematic reviews have focused on computerized decisions tools. Syrowatka et al [13] conducted a systematic review and meta-analyses to classify the features that have been integrated into computerized decision aids and assessed whether these features enable higher-quality decision making. Sheehan and Sherman [14] evaluated the effectiveness of various computerized decision aids in preference-sensitive health-related contexts such as treatments, screening, genetic testing, and risk-management decisions. Their study found that computerized decision aids were efficacious in improving decision-specific knowledge, reducing decisional conflict, and facilitating satisfaction with the decision-making process. Murray et al [15] examined the use of interaction health communication applications (IHCAs), a specific format of a computerized decision aid, for people with chronic disease. The findings have suggested that IHCAs are able to increase patients' knowledge and sense of support as well as improve clinical outcomes. These studies provide a foundation upon which to further assess computerized decision aids. Missing from the current literature is a review of the available computerized decision aids that specifically address shared decision making by seriously ill patients.

Objective

This systematic review builds on the work of Austin et al [11] and assesses the impact of computerized decision aids on patient-centered outcomes of seriously ill patients. Austin et al [11] defined serious illness to include "critical life-threatening illness, advance stages of major chronic diseases or multi-morbidity and frailty." The tools reviewed by Austin et al [11] included print, video, or Web-based formats. For the scope of this review, serious illness will refer to critical, life-threatening illness, chronic disease, multimorbidities, and frailty. This definition of serious illness is a modified version of the definition put forth by Austin et al [11]; the scope of the definition has been broadened to include all stages of chronic disease. Chronic disease is a growing burden and the most common and costly of all health problems; 86% of all health care spending in the United States in 2010 was for individuals with one or more chronic medical conditions [16]. Additionally, chronic diseases are generally long term, progressive in severity, rarely curable [17,18], and thus, may require many decisions to be made over a lifetime.

Methods

The preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist for systematic reviews was followed for this review. The study was not registered with the International Prospective Register of Systematic Reviews (PROSPERO), and therefore, registration information is not included.

Information Sources

PubMed and Scopus databases were searched in February 2016. The search was conducted without a limitation on the year of publication. The search strategy terms were based on the terms used in the systematic review of similar topic by Austin et al [11] and modified based on the specific technological interests of this paper.

The search terms utilized in the PubMed database were as follows:

> (Computer*[tw] OR electronic health records [MESH] OR internet[MESH] OR electronic medical record*[tw] or website[tw] or web site[tw]) AND (decision making[tw] OR decision support[tw] OR decision support techniques[MESH]) AND (shared[tw] OR patient[MESH] OR patient*[tw] OR patient*centered OR family[tw] OR physician patient relations[MESH] OR surrogate[tw] OR professional family relations[MESH] OR professional family relations[MESH]) AND (terminal[tw] OR chronic[tw] OR advanced[tw] OR severity[tw] OR severe[tw] OR failure*[tw] OR end stage[tw] OR endstage[tw] OR dying[tw] OR Intensive Care Units[MeSH] OR intensive care[tw] OR ICU[tw] OR hospice*[tw])

The above search was then modified for the Scopus database:

> (TITLE-ABS-KEY ("shared decision making" OR sdm OR "patient preferences") AND TITLE-ABS-KEY (illness OR disease OR "intensive care" OR serious) AND TITLE-ABS-KEY (web OR "web based" OR internet OR "computerized decision support" OR cdss OR "decision support" OR technology OR "electronic health record" OR "electronic medical record" OR ehr))

Study Selection

Papers extracted from the search results mentioned SDM tools/aids, communication tools/aids, or SDM in relation to an illness or disease in the title or abstract. The abstracts and/or full text were then reviewed; papers were included if the study design was determined to be an RCT. Papers that assessed the use of noncomputerized tools or aids such as videos or pamphlets were excluded as the purpose of this review was to consider computerized decision tools. Tool formats included were Web-based, EHR portals, or computerized decision support software. Included RCTs had to discuss the use of computerized decision aids in serious illness as defined in the introduction. Finally, the paper had to discuss the tool in relation to aspects of SDM such as reducing decisional conflict and increasing knowledge. Tools were included if they were for the use of patients and/or family of patients. The patient population considered included both adults and children living with serious illness.

The references in the selected papers were hand-searched for relevant papers. Data from the final papers were manually extracted. Only published papers and papers in English were included in the study. The selection of papers was completed by one investigator.

Quality Assessment

Papers were graded on quality using a Cochrane Risk of Bias Tool (Modified) for Quality Assessment of Randomized Controlled Trials. The quality assessment included the following study validity domains: selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias. Studies were assessed as either High, Low, or Unclear Risk of Bias. Identified problems in one domain would result in the study being labeled as "High Risk of Bias." Assessment of the quality of the selected papers was completed by 2 investigators. In case of a disagreement between the 2 investigators completing quality assessment, a third investigator was consulted.

Analysis

Study characteristics of all included RCTs were described according to PRISMA systematic review guidelines. All patient-centered outcomes in relation to SDM or communication were described, regardless of whether they differed significantly from the control. Patient-centered outcomes extracted from studies varied and included satisfaction with decision, decisional conflict, clinical outcomes, knowledge, preparation for decision making, emotional well-being, perceived involvement in medical decision making, patient expectations, satisfaction with physician discussion, parental activation, and number of school or work days missed. As *P* values for study outcomes were available for the majority of patient-centered outcomes measured in the RCTs, they were used to describe the efficacy of the interventions.

Results

A total of six papers describing RCTs of SDM tools for serious illness were selected and reviewed (Figure 1): three papers described the efficacy of Web-based tools [19-21]; one paper described a tool that operated through an EHR portal [22]; and two papers described interactive computer application tools [23,24]. The ensuing sections will describe each paper in more depth. The effects of the tools on patient-centered outcomes are also shown in Multimedia Appendix 1 [19-24].

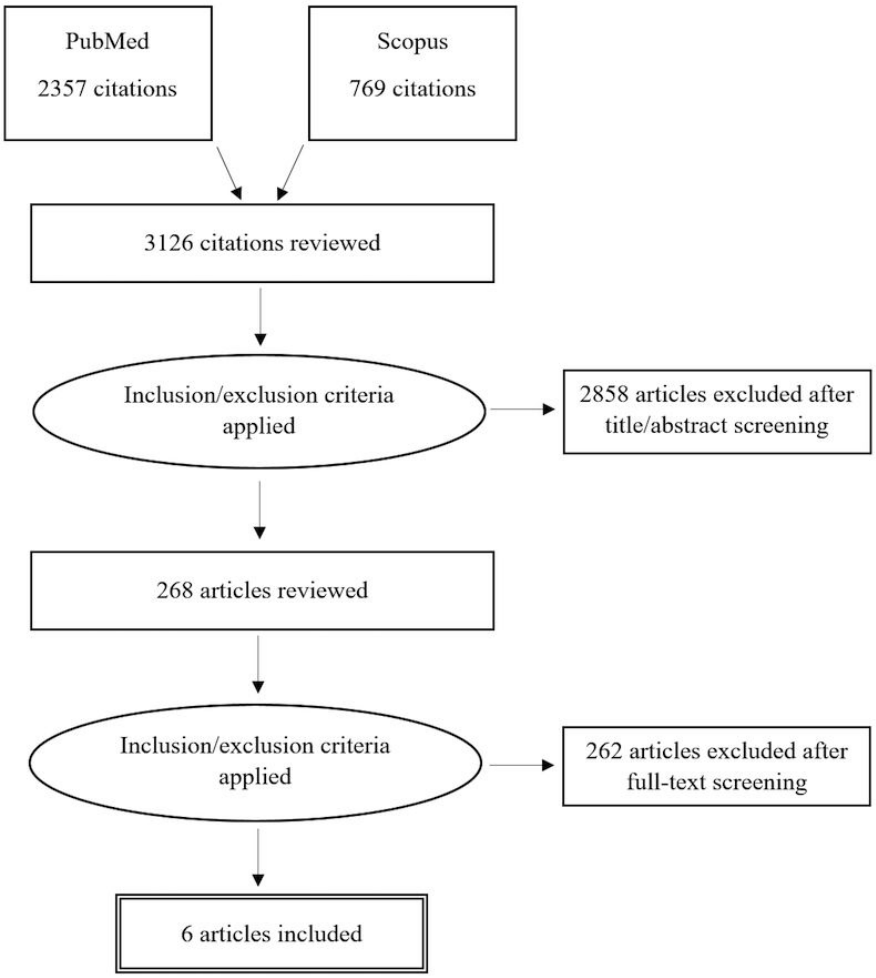
The results suggest that computerized decision aids may be used for various types of serious illnesses in a variety of different health care settings to assist both patients and clinicians in decision making. Generally, the selected RCTs demonstrated that computerized decision aids were able to reduce decisional conflict [19,21,23], improve satisfaction with decisions [19,23], and improve health outcomes [18,19,24]. Other factors that may have influenced the use and efficacy of the computerized aids included type and severity of illness [19,21,22,24], patients' age [23,24], and patients' education and computer literacy [23,24].

Each of the formats of the computerized decision aids included common features among them. The Web-based decision tools commonly used surveys or questionnaires to ascertain patient preferences, which were then used to guide patient-physician communication or to provide treatment options. The EHR portal decision tool featured the ability for patients to track relevant information and provided educational content, both of which were ultimately used to guide treatment plans. Estimates of treatment efficacy and prognosis were common in interactive computer applications. Of the six selected RCTs, Web-based decision tools were described by Meropol et al [19] for metastatic cancers, van der Krieke et al [20] for nonaffected psychosis, and Weymann et al [21] for type 2 diabetes (T2D) and chronic lower back pain (CLBP). An EHR portal decision tool was used for the management of asthma in the RCT by Fiks et al [22]. Hochlehnert et al [23] and Peele et al [24] described the use of interactive computer applications for the treatment and management of fibromyalgia and breast cancer, respectively.

Web-Based Decision Tools

In a single-blind RCT, Meropol et al [19] tested an interactive Web-based communication aid (CONNECT) for patients with solid metastatic tumors. Cancer patients were randomized into (1) control group, (2) CONNECT aid with communication skills training (CST) and summary report to the physician, and (3) CONNECT aid and CST without physician summary report. The control group was directed to the National Cancer Institute's website and received usual care.

Figure 1. Literature search and selection.



There were no statistically significant differences between the two different intervention arms on any of the satisfaction or decisional conflict responses; the summary report for the physician did not improve outcomes. Intervention arms were combined and analyzed against the control arm. Participants assigned to intervention groups had higher levels of satisfaction with discussions about the format of physician communications and quality of life issues but did not differ in satisfaction of discussion regarding diagnosis/prognosis, treatment options, or support community services. Those in the intervention arms found that CONNECT made it easier to reach treatment decisions and were more satisfied with their treatment choice. Participants in the intervention groups had decreased expectations of severe side effects with standard or experimental therapy. The CONNECT intervention was associated with increased satisfaction with overall communication in those with postsecondary education. Additionally, patients in the intervention arm reporting a lower baseline quality of life had greater satisfaction with overall communication.

The study was limited by a racially and ethnically homogenous sample population that was mostly gathered from large cancer centers. Furthermore, the eligibility criteria limited the study to include only those with personal Internet access or those who could arrive early to their appointments to access computers on-site. Additionally, patients in the control groups were directed to the National Cancer Institute website, where extensive searching by the patients may result in a reduced difference between groups. Furthermore, the merging of the intervention groups may place the study at risk of reporting bias. Using the modified Cochrane Risk of Bias Tool for Quality Assessment of Randomized Controlled Trials, the RCT by Meropol et al [19] was rated at High Risk of Bias.

An RCT by van der Krieke et al [20] examined the capability of a Web-based intervention to facilitate SDM for people with psychotic disorders. Patients in the intervention group were given usual care and access to a Web-based tool to support SDM. The control group was given usual care.

Perceived involvement in medical decision making did not differ from patients in the control condition. There were no differences in self-reported satisfaction with care between study arms. However, within the intervention group, those who received the allocated intervention reported lower satisfaction with care in comparison with those who did not receive the intervention.

The study demonstrated a low response rate (29.2%, 73/250) and a moderate participation rate. Furthermore, the study protocol was weakly implemented; not all participants in the intervention group were offered the possibility to use the decision aid, and treatment evaluation meetings where the SDM process would have been used to guide treatment plans did not always occur. The authors do not provide sufficient information regarding the blinding process, if any, that was implemented in

the study; therefore, there is unclear risk of selection, performance, and detection bias. The study by van der Krieke et al [20] was rated at Unclear Risk of Bias.

A Web-based, tailored, interactive health communication application for patients with T2D or CLBP was tested in an RCT by Weymann et al [21]. The intervention group received the Web-based tailored communication tool that provided basic information on T2D and CLBP, along with treatment options in an interactive dialogue format. The control group received an untailored Web-based communication tool that was not presented in a dialogue format.

Intention-to-treat analysis, which used the baseline data, found no statistically significant differences between the groups; however, there was a significant difference between T2D and CLBP users, indicating higher knowledge scores in the T2D group. Conversely, sensitivity analysis, which used data from the available cases, found that participants using the tailored system displayed more knowledge immediately after the first visit than those in the control group. Additionally, those in the intervention group had more emotional well-being as identified by a subscale of a patient empowerment scale at the 3-month follow-up. Sensitivity analysis did not result in significant differences between the intervention and control groups in decisional conflict and preparation for decision making.

The sample population was only limited to those with personal Internet access, which may not be representative of the general population. Additionally, the study did not assess outcome criteria at baseline or address potential confounders, both of which make it unclear whether any observable differences were a result of the intervention or other factors. The measure used to assess T2D/CLBP knowledge, a primary outcome of the trial, was also not validated. Moreover, despite blinding of the participants, the use of the dialogue format may have allowed participants to identify the intervention. The study by Weymann et al [21] is therefore at a risk of detection bias and was rated at High Risk of Bias.

EHR Portal Decision Tool

An RCT by Fiks et al [22] tested the impact of an EHR-linked patient portal with decision support directed at both families and clinicians on asthma outcomes in pediatric patients. The intervention consisted of an EHR-based Web portal, MyAsthma, which provided decision support to both families and clinicians. The families in the control arm did not have access to the portal, but their physicians had access to a clinician-focused decision support system.

The authors reported no statistically significant differences between the control and intervention groups' satisfaction with asthma care or medication receipt, but data were not made available in the study report. There was no effect on parental knowledge, skills, and confidence. Parents in the intervention group had a significant decrease in the number of days of work missed in comparison with the controls. Analysis indicated an improvement of the frequency of asthma flares in the intervention group compared with the control group. There were no differences in quality of life measurements between the two groups; however, compared with the control group, families of intervention group reported fewer emergency department visits and hospitalization over 6 months. Portal use was also found to be greater in parents of children with moderate to severe asthma than those whose children had mild persistent asthma.

As the participants were recruited based on referrals by physicians or EHR rosters, the sample is considered a convenience sample and its representativeness is unclear. Also, because of the small sample size, randomization did not result in a balance between intervention and control groups in terms of asthma severity. The inadequate randomization of participants places the study at a risk of selection bias, and therefore, the study by Fiks et al [22] was rated at High Risk of Bias.

Interactive Computer Applications

An RCT by Hochlehnert et al [23] examined the impact of a computerized information tool with and without physician communication training on SDM in patients with fibromyalgia. Patients were randomized into two study arms: (1) a shared decision group (SDM group) that was given a computer-based information tool and then an opportunity for consultation with a physician with communications training and (2) an information-only group (Info group) that was also given the computer-based information tool but was treated by doctors without communications training with no opportunity for feedback and discussion after viewing the tool.

There was no significant difference in satisfaction with decision or decisional conflict, as well as assessment of information tool between the two groups. The two groups were merged for analysis, and it was found that those who were satisfied with the information presented in the tool experienced more satisfaction with their decision and experienced less decisional conflict. Furthermore, those who perceived the tool to be useful in a general practitioner's office and were satisfied with introduction of the tool (ie, training) were more likely to be satisfied with their decision.

The authors do not provide sufficient information regarding the blinding process that was implemented in this trial; therefore, the risk of performance and detection bias is unclear. The study by Hochlehnert et al [23] was rated at an Unclear Risk of Bias.

An RCT by Peele et al [24] compared rates of breast cancer adjuvant therapy between an intervention group that received a patient-specific decision aid in the form of a computer program and a control group that received an informational pamphlet. Women with breast cancer, who completed their primary surgical treatment, were candidates for adjuvant therapy (chemotherapy, hormonal, or combination therapy) and were randomized into control or intervention groups. The computer program, Adjuvant!, produced prognostic estimates of survival with and without adjuvant therapy by using estimates of individual patient prognosis as well as estimates of the efficacy of adjuvant therapy options.

Women who received the decision aid were significantly less likely to choose adjuvant therapy than those in the control group; one-third fewer women in the intervention group received adjuvant therapy than their counterparts in the control group. The impact of the decision aid based on tumor severity found that the participants in the intervention group with low tumor

severity rejected adjuvant therapy significantly more often than the participants in the control group. Generally, women with higher tumor severity, younger women, and women with a university-based physician were more likely to choose adjuvant therapy.

Neither patients nor clinicians were blinded in this study, indicating risk for performance and detection bias. A higher proportion of university-based physicians were randomized into the intervention group, which places the study at a risk of selection bias as well. The study by Peele et al [24] was therefore categorized as at a High Risk of Bias.

Discussion

Principal Findings

Study results from the six RCTs discussed in the Results section demonstrate that computerized decision aids have the potential to improve patient-centered outcomes. Furthermore, decision aids have differing impacts on various patient-centered outcomes that can possibly be attributed to tool design, user characteristics, or type of disease. Coincidentally, in this review, each of the selected RCTs employed computerized decision aids in management of chronic illnesses, although this was not specified in the search strategy. Furthermore, the small number of studies that are included in this review also suggests that there is still much work to be done in this area. Of the six computerized decision aids discussed in this review, only the tool used by Hochlehnert et al [23] is available online in German (accessed here: www.fibronet.org).

Decisional conflict was addressed in four RCTs [19-21,23]. The CONNECT decision aid was the only decision tool that resulted in a significant reduction of decisional conflict in comparison with control groups [19]. This result is atypical of the high-quality evidence from the Cochrane Review that demonstrated the ability of decision aids to reduce decisional conflict [8]. The failure of the decision tools to reduce decisional conflict in Hochlehnert et al [23] and Weymann et al [21] may be due to the presence of computerized decision aids in both control and intervention groups rather than the control group receiving usual care. Therefore, these studies effectively compare the difference between different types of computerized decision aids and their effects on patient-centered outcomes. The addition of a control group without decision aid access in the studies by Hochlehnert et al [23] and Weymann et al [21] would have allowed for the evaluation of the SDM tools' effectiveness in comparison with usual care. Additional factors may have also affected efficacy of the computerized decision aids in reducing decisional conflict. Whereas Weymann et al [21] found no significant effects for decisional conflict, they did observe an impact on knowledge, suggesting that the tool used in the study may act more as an educational rather than a decisional tool.

EHR portals that function as a decision support system for both patients and physicians present a unique opportunity to manage care from the home. The MyAsthma portal for pediatric asthma did not have an effect on quality of life measures but did result in decreased days of work missed by parents of pediatric patients and a reduction of asthma flares [22]. This suggests that EHR portals can help patients or family members self-manage chronic illnesses. The use of EHR portals to facilitate SDM is fitting as electronic medical record utilization is considerable; approximately 75% of the Canadian and US physicians use electronic medical records [25,26].

The information presented in a decision tool is of importance to achieving meaningful patient-centered outcomes. Hochlehnert et al [23] demonstrated that satisfaction of tool information, tool usefulness, and tool introduction was significantly associated with satisfaction of treatment decision and decreased decisional conflict for fibromyalgia patients. Meropol et al [19] also reported an increase in patient satisfaction in the intervention group and also found this to be related to patients' education level and baseline quality of life scores. Patients with higher levels of education and poorer physical functioning were found to be more satisfied following tool use. Conversely, while van der Krieke et al [20] did not find any overall difference in patient satisfaction, it was found that those in the intervention group who had received the opportunity to use the tool reported lower satisfaction compared with those who did not. This finding may have been a result of poor implementation of the study protocol or may have been due to other factors such as the format of the computerized decision aid use, the setting in which the tool was used, and whether guidance on tool use is provided. It is, therefore, important to consider contextual factors that may influence the use and effectiveness of computerized decision aids. Evidence-based frameworks, such as the Ottawa Decision Support Framework, have been used to develop and evaluate patient decision aids [27]. Further research should focus on determining which formats or, more specifically, which features of computerized decision aids are most helpful for patients.

It is possible that patients may benefit from decision tools without the need for specialized communications training or extra involvement of physicians. The computerized information tool for fibromyalgia patients was tested with and without consultation of a physician specially trained in facilitating SDM. There were no statistical differences between groups on any patient-centered outcomes, including decisional conflict or satisfaction with decision [23]. Additionally, the intervention arms for the CONNECT Web-based communication aid with and without a summary report for the patient's physician did not differ in any patient-centered outcomes [19]. Time constraints are often cited as a barrier to the implementation of SDM [28]; therefore, reduced physician involvement may lead to greater acceptance of SDM tools.

A computerized decision aid, such as Adjuvant!, can present the risks and benefits of treatments to the patients and allow them to consider their preferences and values when making treatment decisions. This may result in a reduction in therapies that are not in line with patient preferences or disease severity and, consequently, can reduce treatment cost: Adjuvant! demonstrated a reduction in adjuvant therapy, such as chemotherapy, in breast cancer patients and was effective at decreasing adjuvant treatment in patients with low tumor severity [24].

Only one study compared computerized decision aids against a nontechnological decision aid. Adjuvant! resulted in decreased use of adjuvant therapy in comparison with control group participants who received informational pamphlets about adjuvant therapy [24]. This suggests that the computerized aid was more effective than a traditional pamphlet in communicating information on treatment options and expectations. Further research that compares traditional noncomputerized decision aids against computerized decision aids in serious illness would be useful; computerized decision aids may be more sophisticated in their ability to communicate health information to patients than traditional aids because of their greater degree of interactivity and personalization.

The tools discussed in this review are relatively simple from a technological perspective. There is potential for greater detail and personalization in SDM with the advent of more advanced decision support tools and the widespread of EHRs. For example, it has been suggested that dynamic clinical data mining can be used to provide real-time decision support. Search engine queries of a population database built on deidentified EHR would provide clinical data support using prior clinical cases, relevant statistics, scholarly resources, and protocols [29]. However, to properly facilitate SDM, a patient interface would need to be included. Additionally, integration of genomic data into EHRs can provide genomic risk scores and personalized risk information to the patient and help guide SDM [30]. Finally, a move toward more universal decision support with the ability to update based on new research findings, patient experience, and postdecision outcomes may be more cost-effective than separate and static decision aids for each disease and treatment options.

Limitations

A limitation of this study was the quality of the RCTs selected for review. The RCTs included in this review were either at High Risk of Bias or Unclear Risk of Bias. Risk of bias should be considered when assessing the strength of evidence provided by the RCTs in this review. The literature search was also only limited to published papers and is therefore subject to publication bias. Furthermore, the search was only limited to PubMed and Scopus databases. Although these databases consist of an extensive amount of literature on the topic, the results may not have been representative of the entirety of the literature. Additionally, secondary search strategies were not performed. The study was also not registered in PROSPERO, which limits the study in terms of adhering to current best practices for systematic reviews. Finally, although quality assessment of the selected papers was completed by 2 investigators, the study is limited because of the fact that the selection of papers was completed by only a single investigator.

Conclusions

Most computerized decision aids improved at least one patient-centered outcome. The RCTs differed in patient outcomes measured and the efficacy of decision aids in improving the aspects of SDM. All RCTs identified were at High Risk of Bias or Unclear Risk of Bias according to a modified version of the Cochrane Risk of Bias Tool for Quality Assessment of Randomized Controlled Trials. Efforts should be made to improve the quality of RCTs testing SDM aids in serious illness.

Acknowledgments

This work was supported by a Discovery Grant (RGPIN-2014-04743) from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Study description, effect of decision tools on patient-centered outcomes, and risk of bias.

[PDF File (Adobe PDF File), 67KB - medinform_v5i4e36_app1.pdf]

References

1. Lenert L, Dunlea R, Del Fiol G, Hall LK. A model to support shared decision making in electronic health records systems. Med Decis Making 2014 Nov;34(8):987-995 [FREE Full text] [doi: 10.1177/0272989X14550102] [Medline: 25224366]
2. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
3. Pollard S, Bansback N, Bryan S. Physician attitudes toward shared decision making: A systematic review. Patient Educ Couns 2015 Sep;98(9):1046-1057. [doi: 10.1016/j.pec.2015.05.004] [Medline: 26138158]
4. Khatcheressian J, Harrington SB, Lyckholm LJ, Smith TJ. 'Futile care': what to do when your patient insists on chemotherapy that likely won't help. Oncology (Williston Park) 2008 Jul;22(8):881-8; discussion 893, 896, 898 [FREE Full text] [Medline: 18709900]

5. Weeks JC, Catalano PJ, Cronin A, Finkelman MD, Mack JW, Keating NL, et al. Patients' expectations about effects of chemotherapy for advanced cancer. N Engl J Med 2012 Oct 25;367(17):1616-1625 [FREE Full text] [doi: 10.1056/NEJMoa1204410] [Medline: 23094723]
6. Jones JB, Bruce CA, Shah NR, Taylor WF, Stewart WF. Shared decision making: using health information technology to integrate patient choice into primary care. Transl Behav Med 2011 Mar;1(1):123-133 [FREE Full text] [doi: 10.1007/s13142-011-0023-5] [Medline: 24073039]
7. Tulsky JA. Decision aids in serious illness: moving what works into practice. JAMA Intern Med 2015 Jul;175(7):1221-1222. [doi: 10.1001/jamainternmed.2015.1702] [Medline: 25984701]
8. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Dec 12;4:CD001431. [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
9. Halpern SD. Toward evidence-based end-of-life care. N Engl J Med 2015 Nov 19;373(21):2001-2003. [doi: 10.1056/NEJMp1509664] [Medline: 26465826]
10. Clauser SB, Wagner EH, Aiello Bowles EJ, Tuzzio L, Greene SM. Improving modern cancer care through information technology. Am J Prev Med 2011 May;40(5 Suppl 2):S198-S207 [FREE Full text] [doi: 10.1016/j.amepre.2011.01.014] [Medline: 21521595]
11. Austin CA, Mohottige D, Sudore RL, Smith AK, Hanson LC. Tools to promote shared decision making in serious illness: a systematic review. JAMA Intern Med 2015 Jul;175(7):1213-1221 [FREE Full text] [doi: 10.1001/jamainternmed.2015.1679] [Medline: 25985438]
12. Jain A, Corriveau S, Quinn K, Gardhouse A, Vegas DB, You JJ. Video decision aids to assist with advance care planning: a systematic review and meta-analysis. BMJ Open 2015 Jun 24;5(6):e007491 [FREE Full text] [doi: 10.1136/bmjopen-2014-007491] [Medline: 26109115]
13. Syrowatka A, Krömker D, Meguerditchian AN, Tamblyn R. Features of computer-based decision aids: systematic review, thematic synthesis, and meta-analyses. J Med Internet Res 2016 Jan 26;18(1):e20 [FREE Full text] [doi: 10.2196/jmir.4982] [Medline: 26813512]
14. Sheehan J, Sherman KA. Computerised decision aids: a systematic review of their effectiveness in facilitating high-quality decision-making in various health-related contexts. Patient Educ Couns 2012 Jul;88(1):69-86. [doi: 10.1016/j.pec.2011.11.006] [Medline: 22185961]
15. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
16. Centers for Disease Control and Prevention. CDC. 2016. Chronic disease overview URL: https://www.cdc.gov/chronicdisease/overview/index.htm [WebCite Cache ID 6tKY9AomQ]
17. WHO. Noncommunicable diseases URL: http://www.who.int/topics/noncommunicable_diseases/en/ [WebCite Cache ID 6j87sEfaF]
18. health.gov. 2007. Preventing and managing chronic disease: Ontario's framework URL: http://www.health.gov.on.ca/en/pro/programs/cdpm/ [WebCite Cache ID 6j87z3UDD]
19. Meropol NJ, Egleston BL, Buzaglo JS, Balshem A, Benson 3rd AB, Cegala DJ, et al. A Web-based communication aid for patients with cancer: the CONNECT Study. Cancer 2013 Apr 01;119(7):1437-1445 [FREE Full text] [doi: 10.1002/cncr.27874] [Medline: 23335150]
20. van der Krieke L, Emerencia AC, Boonstra N, Wunderink L, de Jonge P, Sytema S. A web-based tool to support shared decision making for people with a psychotic disorder: randomized controlled trial and process evaluation. J Med Internet Res 2013;15(10):e216 [FREE Full text] [doi: 10.2196/jmir.2851] [Medline: 24100091]
21. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a Web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
22. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, Localio AR, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973. [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
23. Hochlehnert A, Richter A, Bludau H, Bieber C, Blumenstiel K, Mueller K, et al. A computer-based information-tool for chronic pain patients. Computerized information to support the process of shared decision-making. Patient Educ Couns 2006 Apr;61(1):92-98. [doi: 10.1016/j.pec.2005.02.014] [Medline: 16533681]
24. Peele PB, Siminoff LA, Xu Y, Ravdin PM. Decreased use of adjuvant breast cancer therapy in a randomized controlled trial of a decision aid with individualized risk information. Med Decis Making 2005;25(3):301-307. [doi: 10.1177/0272989X05276851] [Medline: 15951457]
25. Collier R. National physician survey: EMR use at 75%. CMAJ 2015 Jan 6;187(1):E17-E18. [doi: 10.1503/cmaj.109-4957] [Medline: 25487665]
26. Jamoom EW, Yang N, Hing E. Adoption of certified electronic health record systems and electronic information sharing in physician offices: United States, 2013 and 2014. NCHS Data Brief 2016 Jan(236):1-8 [FREE Full text] [Medline: 26828707]

27. Ottawa Hospital Research Institute. 2015. Ottawa decision support framework (ODSF) URL: https://decisionaid.ohri.ca/odsf.html [WebCite Cache ID 6tKYeOaPi]
28. Gravel K, Légaré F, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: a systematic review of health professionals' perceptions. Implement Sci 2006 Aug 09;1:16 [FREE Full text] [doi: 10.1186/1748-5908-1-16] [Medline: 16899124]
29. Celi LA, Zimolzak AJ, Stone DJ. Dynamic clinical data mining: search engine-based decision support. JMIR Med Inform 2014;2(1):e13 [FREE Full text] [doi: 10.2196/medinform.3110] [Medline: 25600664]
30. Larson EA, Wilke RA. Integration of genomics in primary care. Am J Med 2015 Nov;128(11):1251.e1-1251.e5. [doi: 10.1016/j.amjmed.2015.05.011] [Medline: 26031886]

Abbreviations

CLBP: chronic lower back pain
CST: communication skills training
EHR: electronic health record
IHCAs: interaction health communication applications
PRISMA: preferred reporting items for systematic reviews and meta-analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCTs: randomized controlled trials
SDM: shared decision making
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 25.07.16; peer-reviewed by CA Austin, M Härter, C Short; comments to author 20.11.16; revised version received 31.07.17; accepted 09.08.17; published 06.10.17.

Please cite as:
Staszewska A, Zaki P, Lee J
Computerized Decision Aids for Shared Decision Making in Serious Illness: Systematic Review
JMIR Med Inform 2017;5(4):e36
URL: https://medinform.jmir.org/2017/4/e36/
doi:10.2196/medinform.6405
PMID:28986341

Viewpoint

Examining Tensions That Affect the Evaluation of Technology in Health Care: Considerations for System Decision Makers From the Perspective of Industry and Evaluators

Laura Desveaux[1,2], PT, PhD; James Shaw[1,2], PT, PhD; Ross Wallace[1,3], MPA, MBA; Onil Bhattacharyya[1,2,4], MD, PhD; R Sacha Bhatia[1,2,5], MD, MBA; Trevor Jamieson[1,5,6], MD, MBI

[1]Institute for Health Systems Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada
[2]Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada
[3]Santis Health, Toronto, ON, Canada
[4]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[5]Department of Medicine, University of Toronto, Toronto, ON, Canada
[6]Division of General Internal Medicine, St. Michael's Hospital, Toronto, ON, Canada

Corresponding Author:
Laura Desveaux, PT, PhD
Institute for Health Systems Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 323 6400 ext 8356
Fax: 1 416 323 7308
Email: laura.desveaux@wchospital.ca

Abstract

Virtual technologies have the potential to mitigate a range of challenges for health care systems. Despite the widespread use of mobile devices in everyday life, they currently have a limited role in health service delivery and clinical care. Efforts to integrate the fast-paced consumer technology market with health care delivery exposes tensions among patients, providers, vendors, evaluators, and system decision makers. This paper explores the key tensions between the high bar for evidence prior to market approval that guides health care regulatory decisions and the “fail fast” reality of the technology industry. We examine three core tensions: balancing user needs versus system needs, rigor versus responsiveness, and the role of pre- versus postmarket evidence generation. We use these to elaborate on the structure and appropriateness of evaluation mechanisms for virtual care solutions. Virtual technologies provide a foundation for personalized, patient-centered medicine on the user side, coupled with a broader understanding of impact on the system side. However, mechanisms for stakeholder discussion are needed to clarify the nature of the health technology marketplace and the drivers of evaluation priorities.

(JMIR Med Inform 2017;5(4):e50) doi:10.2196/medinform.8207

KEYWORDS

technology; evaluation; policy; healthcare

Introduction

Providing patient-centered care is an ongoing challenging due to rising costs [1], poor access [2], increasing complexity of patient needs [3], and the provider-centered structure of health systems [4]. Virtual care technologies have the potential to mitigate these challenges by lowering costs, improving access, and managing complexity [5-7], while being tailored to the needs and wants of users. These technologies can also support population-level research and the application of scientific evidence at a system level by providing real-time access to data across a broad population [8]. Despite this, the uptake of both provider- and patient-facing technologies has been limited in health systems compared to many other industries [9,10]. However, the mobile devices needed to access virtual care technologies are already in the hands of most individuals [11], which raises the question “What is limiting their potential in health care”?

The traditional approach of evidence-based medicine is at odds with the "fail fast" mentality of the technology industry, where rapid iterative testing facilitates early feedback from users leading to course corrections and better solutions [12]. Potential health care disrupters are confronted with a web of regulations, contractual obligations, provider interests, and interlocking financial incentives [10]. Some of the most challenging roadblocks are related to safety concerns and risk management; just because people like and use certain health-related apps does not mean they are safe and achieve positive health-related outcomes. To address questions of safety and effectiveness, pharmaceutical and medical device industries have established evaluation paradigms [13,14]. However, these approaches may not be appropriate for virtual care solutions because their high cost, long timelines, and rigid protocols do not account for the dynamic nature of software and the speed of the technology marketplace [15,16].

Efforts to integrate the high-paced consumer technology market with health care delivery exposes tensions at the intersection of users (including patients and health care providers), vendors, third-party evaluators (including scientific researchers), and system decision makers. The objective of this paper is to explore the key tensions between the high bar for evidence prior to market approval that guides health care regulatory decisions and the "fail fast" reality of the technology industry. We then elaborate on the implications of these tensions on the structure and appropriateness of evaluation mechanisms for virtual care solutions. Our goal is to carefully examine three core tensions: (1) balancing user needs versus system needs, (2) rigor versus responsiveness, and (3) the role of pre- versus postmarket evidence generation—the latter exploring the extent to which evidence of effectiveness and/or safety should be (and can be accurately) demonstrated before a product is used in real life. These observations come from our experiences with virtual care implementation and evaluation, including large randomized controlled trials (RCTs) [17-19], consultation with technology start-ups through the Canadian Government's Industrial Research Assistance Program [20], and dialogue with policy stakeholders [21].

The integration of virtual care "is hampered because different stakeholders hold different assumptions, values and world views, 'talk past' each other, and compete for recognition and resources" [22]. Though stakeholder engagement has long been proposed in system design, it is not routinely done in health care. Mechanisms for more effective stakeholder dialogue are needed in order to establish a common vision, including consensus on what constitutes value, how it is determined, and who should be the primary beneficiaries. The presentation of the following tensions is intended to both illuminate these problems and facilitate this dialogue.

Should Technologies and Evaluations Prioritize User Needs Versus System Needs?

Policy making for virtual care technologies must balance several priorities, such as economic and health care objectives [23]. Economic policy defines success as the creation of jobs, with the assumption that virtual care innovation "paves the road to economic development, solving societal challenges along the way" [23]. This favors economic interest over health system objectives, creating a tension between a system designed to facilitate innovation and a system that benefits its users. Health technology assessment (HTA) emerged as an approach to provide information about the efficacy, safety, and cost-effectiveness of health technologies for the purposes of decision making. Although its application aims to initiate processes that support the institutionalization of virtual care solutions, few examples of HTA demonstrate a commitment to understanding the needs, realities, and practices of end users [24]. HTA has standardized the value of different outcomes but does not address how a given solution fits with the end user's reality. In short, HTA is primarily devoted to the needs of policy makers, as opposed to the needs of end users.

The consumer-oriented nature of virtual care allows for customization to the specific needs of the user, resulting in a wide range of vendors, user interface options, mobile apps, and wearables. These solutions can be rapidly developed and distributed, and they are easy to modify based on ongoing feedback from users, generating products that are tailored to user priorities, thus increasing the likelihood of adoption and meaningful use. However, focusing primarily on local user priorities can lead to solutions that are developed without considering system-level priorities, such as interoperability, change management, system-level cost-effectiveness, and population-level outcomes. As a consequence, many pilots of virtual care technologies have shown local effectiveness across a variety of clinical areas but fail to be used widely in practice [25,26]. In contrast, the development of universal solutions that meet industry standards and are widely implemented often occurs at the expense of responsiveness to diverse user needs across different contexts.

We suggest that a stronger consideration of system needs (ie, how virtual care technologies will function within the context of the larger system) while incorporating user needs, priorities, and values, will facilitate widespread adoption. This could be done by (1) identifying how a virtual solution fits within a larger system strategy, (2) developing a targeted outcome assessment that reflects both user and system needs, and (3) creating a change management plan that considers contextual factors to support rapid scale-up of successful interventions. It is worth noting that "scaling" in this context does not necessarily imply replication, but rather the facilitated diffusion (and evaluation) of solutions from one setting to another with the flexibility to allow for tailoring and modification.

Despite the range of perspectives, it is important to acknowledge that when it comes to the implementation of new technologies into organizations and systems, context and culture tend to drive changes in the form and use of technology, rather than the other way around [29]. This underscores the critical role of evaluation to help understand the local realities that may explain observed effects, unanticipated harms, or lead to broad rejection of the technology altogether. For example, a systematic review of 37 interventions revealed that suboptimal implementation was explained by the lack of attention to (1) specifying the purposes

and benefits of virtual solutions and establishing their value to users, (2) effects on roles and responsibilities, (3) risk management, and (4) using user knowledge to modify implementation processes [30]. Technology may fail to become adopted when users do not perceive that the organization has a culture that is supportive of change, the solution does not align with perceived organizational priorities, and the impact of using the technology on individual accountability and liability is not understood [31].

Do Evaluations Prioritize Rigor Versus Responsiveness?

Technology is dynamic and easy to modify on an ongoing basis. Approaching evaluation as if virtual care technologies are a static intervention may lead to the perception of more rigor, but the results may ultimately be rendered obsolete in light of updates to the technology itself. The rigor required for market entry in health care is a key element of regulatory systems, highlighting the need for system decision makers to consider the estimated opportunity costs, financial costs, and potential harms of a more open or closed market. Unlike pharmaceuticals, virtual care technologies often exist simultaneously in both the health care and consumer device marketplaces. These intertwined marketplaces must necessarily influence regulations and requirements that govern entry into the health care marketplace. The risk of rigorous evaluations with lengthy timelines is that health care technologies become "fixed" relative to the consumer marketplace, resulting in a confusing mismatch for the end user between technologies that do similar things. The risk of a marketplace without constraints to entry is the proliferation of technologies of uncertain value. In the best case, they may be inexpensive and outcome neutral; in the worst case, they may be costly and have an adverse effect on health care quality.

Many virtual care solutions can collect and transmit data, so it is possible to continuously monitor safety and respond quickly. In a study among patients with heart failure, remote monitoring ensured the timely transmission of data to the health care team, resulting in early intervention as needed and a 3% mortality rate in the intervention group (compared to 8.2% among controls) [32]. Extracting user data from devices automatically as a condition of market entry can alleviate issues of access and data reconciliation that plague health outcomes research [33,34]. This supports evaluations that move beyond the traditional RCT model towards a more adaptive model [35,36], better positioning them to balance both rigor and responsiveness. Unlike traditional RCTs, adaptive trial design allows for modifications to the trial after its initiation, including aspects such as target population, intervention design, dose, duration, and statistical procedures [37]. Adaptive trials maintain validity and integrity by ensuring modifications occur in response to observations made during the trial and prior to the unblinded analysis of trial data.

Pre-market evaluations employing rapid, iterative cycles are well suited to support this model [38]. These types of solutions would parallel existing models in quality improvement, allowing for deployment, monitoring, incremental improvement, and local tailoring. Although this does not eliminate the need to define requirements for evaluation rigor, it provides some reassurance that pre-market rigor can be compromised without the threat of long-term negative consequences. It is also important to note that while harnessing the collection of simple, patient-generated data expedites the timeliness of ongoing evaluation and improvement cycles, it requires clarity around the nature of "ownership" and the appropriate mechanisms to ensure security of the data [39].

Is the Evaluation Structure Influenced by Pre- or Postmarket Status?

Pharmaceuticals and medical devices are subject to pre-market evaluations to establish their suitability for entry into the broader health system. Although this assumption has carried over into the domain of virtual care, the extent to which pre-market evaluation can truly establish safety and effectiveness for these dynamic solutions remains unclear. Virtual solutions have the added advantage of flexibility when compared to pharmaceuticals, as regular updates enable modifications to address safety concerns (as opposed to being taken off the market entirely).

Safety in virtual care extends far beyond the typical health care considerations of morbidity and mortality. Direct users of virtual care technologies may experience loss of privacy, poor data quality, and suboptimal clinical-decision support. The latter extends from mild and relatively inconsequential decisions (eg, being advised of a suboptimal exercise scheme) to harmful and irreversible clinical actions (eg, being advised to take a harmful dose of insulin) [40]. A recent Institute of Medicine report outlined such safety concerns but did not uncover evidence of significant issues [41]. The majority of evaluators held the opinion that strict regulation would greatly stifle innovation in this space [41].

Current evaluation standards stem from pharmaceutical and medical device industries, where health evidence is generated through rigorous, (ideally) randomized and controlled evaluations that demonstrate both safety and effectiveness prior to use in a real-world setting [13,14]. Such controlled studies are generalizable and can apply across a diverse array of settings because the intervention (eg, medication or a medical device) does not interact substantially with an array of external, contextual factors. However, due to the complex adaptive nature of sociotechnical systems (in which virtual care technologies must fit) [42], safety and effectiveness cannot be determined in a pre-market vacuum. They depend on a range of factors, including interaction between the technology, users, organizations, and environmental conditions that vary across sites [43]. Simply put, the "intervention" in any local environment is the intersection of all of these components. "Idealized" and controlled evaluations therefore *limit* generalizability with respect to virtual care technologies, arguably wasting valuable time and resources, whereas real-life interplay truly determines whether a solution works.

The overarching theme throughout this dialogue is *complexity* in a system of intersecting priorities, within which virtual care technologies can act as a disruptive catalyst. System tolerance

for risk in health care has historically been quite low, resulting in strict requirements for market entry; however, the presence of technology is slowly increasing individual risk tolerance for readily available solutions. This reality may signal an inevitable shift in focus from dimensions of safety and effectiveness to the balanced evaluation of "value-add" for end users. If establishing effectiveness is the priority, the outcome of interest will dictate requirements for recruitment and adequate exposure time, in order to reliably detect a meaningful difference. Ultimately, the nature of the market and the degree of regulation will inform the extent to which safety, effectiveness, or user preference drives the uptake of virtual care solutions and the role that evaluation plays within it.

Conclusion

A recurring theme across our scientific and industry-based engagements has been the challenge faced by technology vendors when attempting to define their customer (defined here as the individual who will purchase their technology). Owing to the relatively open nature of the consumer marketplace, many vendors simultaneously market their virtual care solution to patients, clinicians, clinics, and the government. This reality highlights the failure of the system to clearly and explicitly define the health care marketplace, including the regulations and evaluation parameters required for entry. This is compounded by the fact that, in both single- and multipayer health care systems, the health care "payer" is rarely the end user.

It is important to note that we have simplified these considerations by presenting them as distinct tensions, whereas the interplay across these categories in practice makes them difficult to disentangle. Systematic consideration of the questions arising from these tensions will help define the type of evaluation that meets the needs of both the system and its end users (see Table 1). At a higher level, it can support transformation by challenging the current structure of health care delivery that limits change [44]. Productive dialogue relies on collaborative relationships and requires that those involved acknowledge the range of priorities and accountabilities that operate across the system; but more importantly, they are essential in order to curtail fragmentation and define the appropriate yet comprehensive parameters of both the market and the role of evaluation within it. Stakeholder discussions are needed to clarify the nature of the health technology marketplace, for whom the marketplace primarily aims to generate value, and hence, the drivers of evaluation priorities. These discussions will inform how virtual care solutions are developed, evaluated, and incorporated into health care delivery.

Simply put, regulations that heavily prioritize the system risk rejection by end users, the development of workarounds, or suboptimal outcomes resulting from a failure to consider local context. In contrast, regulations that prioritize end users risk a degree of technological customization that exacerbates fragmentation within the system and is unlikely to improve overall health outcomes or costs. Similarly, strict requirements for pre-market assessment are likely to lead to overly general evaluation results that provide false reassurance, while sparse regulation may lead to the introduction of unsafe and ineffective tools. However, the technologies themselves provide a promising return in exchange for navigating these complex tensions. The plethora of data they generate provides a foundation for personalized, patient-centered medicine on the user side, coupled with a broader understanding of impact on the system side.

Table 1. Tensions and underlying key questions.

Tension	Questions
Prioritizing user needs versus system needs	How does a local strategy fit within the larger system?
	What are the relevant outcomes that reflect user and system needs?
	How can system infrastructure support the scale of successful solutions?
Prioritizing rigor versus responsiveness	How does the consumer virtual care marketplace influence the health care marketplace?
	What outcomes require a rigorous approach?
	What infrastructure is needed to support real-time consolidation and analysis of data?
Pre- or postmarket status influence on evaluation structure	What is the minimum requirement for system entry?
	What are the appropriate pathways for solutions to enter the health care system?
	How can we embed ongoing monitoring and evaluation alongside the use of virtual care solutions?

Acknowledgments

The authors would like to acknowledge our colleagues and partners at the Women's College Hospital Institute for Health Systems Solutions and Virtual Care for their participation in ongoing discussions surrounding these tensions.

Conflicts of Interest

None declared.

References

1. Organisation for Economic Co-operation and Development. Fiscal Sustainability of Health Systems: Bridging Health and Finance Perspectives. Paris, France: OECD Publishing; Sep 25, 2015:1-264.
2. Canadian Institute for Health Information. How Canada Compares: Results From The Commonwealth Fund 2015 International Health Policy Survey of Older Adults. Ottawa, ON: CIHI; 2016 Jan. URL: https://www.cihi.ca/sites/default/files/document/commonwealth_fund_2015_pdf_en.pdf [accessed 2017-12-01] [WebCite Cache ID 6vODAJozN]
3. Osborn R, Moulds D, Schneider EC, Doty MM, Squires D, Sarnak DO. Primary Care Physicians In Ten Countries Report Challenges Caring For Patients With Complex Health Needs. Health Aff (Millwood) 2015 Dec;34(12):2104-2112. [doi: 10.1377/hlthaff.2015.1018] [Medline: 26643631]
4. van der Eijk M, Nijhuis FAP, Faber MJ, Bloem BR. Moving from physician-centered care towards patient-centered care for Parkinson's disease patients. Parkinsonism Relat Disord 2013 Nov;19(11):923-927. [doi: 10.1016/j.parkreldis.2013.04.022] [Medline: 23742970]
5. Lyerly MJ, Wu T, Mullen MT, Albright KC, Wolff C, Boehme AK, et al. The effects of telemedicine on racial and ethnic disparities in access to acute stroke care. J Telemed Telecare 2016 Mar;22(2):114-120 [FREE Full text] [doi: 10.1177/1357633X15589534] [Medline: 26116854]
6. Elbert NJ, van Os-Mendendorp H, van Renselaar W, Ekeland AG, Hakkaart-van RL, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
7. 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]
8. Webb D. Mobile technology gives home-care managers access to real-time data. 2017 Mar 30. URL: http://www.canhealth.com/blog/mobile-technology-gives-home-care-managers-access-to-real-time-data/ [accessed 2017-12-01] [WebCite Cache ID 6vODbGaB2]
9. Keown OP, Parston G, Patel H, Rennie F, Saoud F, Al KH, et al. Lessons from eight countries on diffusing innovation in health care. Health Aff (Millwood) 2014 Sep;33(9):1516-1522. [doi: 10.1377/hlthaff.2014.0382] [Medline: 25201655]
10. Detsky AS, Garber AM. Uber's Message for Health Care. N Engl J Med 2016 Mar 03;374(9):806-809. [doi: 10.1056/NEJMp1512206] [Medline: 26962900]
11. World Bank. Mobile Cellular Subscriptions. 2016. URL: https://data.worldbank.org/indicator/IT.CEL.SETS.P2 [accessed 2017-11-29] [WebCite Cache ID 6vL1jKoe8]
12. Bhattacharyya O, Bhatia RS. Learning about Failure from Successful Ecosystems. Healthc Pap 2015;15(2):32-36. [Medline: 26853609]
13. Ferguson M. Medical devices are different to pharmaceuticals in the Health Technology Assessment process. J Comp Eff Res 2014 May;3(3):217-219. [doi: 10.2217/cer.14.17] [Medline: 24969145]
14. National Institute for Health and Care Excellence. Medical technologies evaluation programme. London: NICE; 2013. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-medical-technologies/Medical-technologies-evaluation-programme-process-guide.pdf [accessed 2017-12-01] [WebCite Cache ID 6vOEFyZvz]
15. Moehr JR. Evaluation: salvation or nemesis of medical informatics? Comput Biol Med 2002 May;32(3):113-125. [Medline: 11922929]
16. Steinberg D, Horwitz G, Zohar D. Building a business model in digital medicine. Nat Biotechnol 2015 Sep;33(9):910-920. [doi: 10.1038/nbt.3339] [Medline: 26348957]
17. Desveaux L, Agarwal P, Shaw J, Hensel JM, Mukerji G, Onabajo N, et al. A randomized wait-list control trial to evaluate the impact of a mobile application to improve self-management of individuals with type 2 diabetes: a study protocol. BMC Med Inform Decis Mak 2016 Nov 15;16(1):144 [FREE Full text] [doi: 10.1186/s12911-016-0381-5] [Medline: 27842539]
18. Hensel J, Shaw J, Jeffs L, Ivers NM, Desveaux L, Cohen A, et al. A pragmatic randomized control trial and realist evaluation on the implementation and effectiveness of an internet application to support self-management among individuals seeking specialized mental health care: a study protocol. BMC Psychiatry 2016 Oct 18;16(1):350 [FREE Full text] [doi: 10.1186/s12888-016-1057-5] [Medline: 27756281]
19. Jeffs L, Jain AK, Man RH, Onabajo N, Desveaux L, Shaw J, et al. Exploring the utility and scalability of a telehomecare intervention for patients with chronic kidney disease undergoing peritoneal dialysis-a study protocol. BMC Nephrol 2017 May 10;18(1):155 [FREE Full text] [doi: 10.1186/s12882-017-0557-y] [Medline: 28486991]
20. National Research Council Canada. Industrial Research Assistance Program (IRAP). 2017. URL: https://www.nrc-cnrc.gc.ca/eng/irap/index.html [accessed 2017-11-29] [WebCite Cache ID 6vL2mxas4]
21. Shaw J, Jamieson T, Agarwal P, Griffin B, Wong I, Bhatia R. Virtual care policy recommendations for patient-centred primary care: findings of a consensus policy dialogue using a nominal group technique. J Telemed Telecare 2017 Jan 01. [doi: 10.1177/1357633X17730444] [Medline: 28945161]
22. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012;2(4) [FREE Full text] [doi: 10.1136/bmjopen-2012-001574] [Medline: 22815469]

23. Lehoux P, Daudelin G, Denis J, Miller F. A Concurrent Analysis of Three Institutions that Transform Health Technology-Based Ventures: Economic Policy, Capital Investment, and Market Approval. Review of Policy Research 2017 Apr 29;34(5):636-659. [doi: 10.1111/ropr.12246]
24. Lehoux P, Denis J, Tailliez S, Hivon M. Dissemination of health technology assessments: identifying the visions guiding an evolving policy innovation in Canada. J Health Polit Policy Law 2005 Aug;30(4):603-641. [Medline: 16318164]
25. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. Eval Health Prof 2006 Mar;29(1):126-153. [doi: 10.1177/0163278705284445] [Medline: 16510882]
26. Damschroder L, Aron D, Keith R, Kirsh S, Alexander J, Lowery J. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
27. Perera G, Holbrook A, Thabane L, Foster G, Willison DJ. Views on health information sharing and privacy from primary care practices using electronic medical records. Int J Med Inform 2011 Feb;80(2):94-101. [doi: 10.1016/j.ijmedinf.2010.11.005] [Medline: 21167771]
28. Greiver M. Do electronic medical records improve quality of care? No. Can Fam Physician 2015 Oct;61(10):847-849,-847-8852 [FREE Full text] [Medline: 26472787]
29. Leonardi P, Barley S. What's Under Construction Here? Social Action, Materiality, and Power in Constructivist Studies of Technology and Organizing. The Academy of Management Annals 2010 Jan;4(1):1-51. [doi: 10.1080/19416521003654160]
30. Mair F, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 01;90(5):357-364 [FREE Full text] [doi: 10.2471/BLT.11.099424] [Medline: 22589569]
31. Finch TL, Mair FS, O'Donnell C, Murray E, May CR. From theory to 'measurement' in complex interventions: methodological lessons from the development of an e-health normalisation instrument. BMC Med Res Methodol 2012 May 17;12:69 [FREE Full text] [doi: 10.1186/1471-2288-12-69] [Medline: 22594537]
32. Hindricks G, Taborsky M, Glikson M, Heinrich U, Schumacher B, Katz A, IN-TIME study group*. Implant-based multiparameter telemonitoring of patients with heart failure (IN-TIME): a randomised controlled trial. Lancet 2014 Aug 16;384(9943):583-590. [doi: 10.1016/S0140-6736(14)61176-4] [Medline: 25131977]
33. Doshi J, Hendrick F, Graff J, Stuart B. Data, Data Everywhere, but Access Remains a Big Issue for Researchers: A Review of Access Policies for Publicly-Funded Patient-Level Health Care Data in the United States. EGEMS (Wash DC) 2016;4(2):1204 [FREE Full text] [doi: 10.13063/2327-9214.1204] [Medline: 27141517]
34. Burgun A, Bodenreider O. Accessing and integrating data and knowledge for biomedical research. Yearb Med Inform 2008:91-101 [FREE Full text] [Medline: 18660883]
35. Lai T, Lavori P, Tsang K. Adaptive design of confirmatory trials: Advances and challenges. Contemp Clin Trials 2015 Nov;45(Pt A):93-102 [FREE Full text] [doi: 10.1016/j.cct.2015.06.007] [Medline: 26079372]
36. Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. Stat Med 2016 Feb 10;35(3):325-347. [doi: 10.1002/sim.6472] [Medline: 25778935]
37. Mahajan R, Gupta K. Adaptive design clinical trials: Methodology, challenges and prospect. Indian J Pharmacol 2010 Aug;42(4):201-207 [FREE Full text] [doi: 10.4103/0253-7613.68417] [Medline: 20927243]
38. Gobble M. Design thinking. Research Technology Management 2014;57(3):59-62.
39. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who Owns the Data? Open Data for Healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
40. Lewis T, Wyatt J. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
41. Cotra Technology, Institute of Medicine. Health IT and Patient Safety: Building Safer Systems for Better Care. Washington, DC: National Academies Press; 2012.
42. Sturmberg JP, Martin CM. The dynamics of health care reform--learning from a complex adaptive systems theoretical perspective. Nonlinear Dynamics Psychol Life Sci 2010 Oct;14(4):525-540. [Medline: 20887693]
43. Sturmberg J, Lanham H. Understanding health care delivery as a complex system: achieving best possible health outcomes for individuals and communities by focusing on interdependencies. J Eval Clin Pract 2014 Dec;20(6):1005-1009. [doi: 10.1111/jep.12142] [Medline: 24797788]
44. Phillips AB, Merrill JA. Innovative use of the integrative review to evaluate evidence of technology transformation in healthcare. J Biomed Inform 2015 Dec;58:114-121 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.014] [Medline: 26429591]

Abbreviations

HTA: health technology assessment
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 16.06.17; peer-reviewed by T Copeland, H Van Os-Medendorp, M Abdelhamid; comments to author 17.09.17; revised version received 09.10.17; accepted 30.10.17; published 08.12.17.

Please cite as:
Desveaux L, Shaw J, Wallace R, Bhattacharyya O, Bhatia RS, Jamieson T
Examining Tensions That Affect the Evaluation of Technology in Health Care: Considerations for System Decision Makers From the Perspective of Industry and Evaluators
JMIR Med Inform 2017;5(4):e50
URL: http://medinform.jmir.org/2017/4/e50/
doi:10.2196/medinform.8207
PMID:29222075

Original Paper

Characteristics of Innovators Adopting a National Personal Health Record in Portugal: Cross-Sectional Study

Liliana Laranjo[1,2], MPH, MD, PhD; Inês Rodolfo[3], BDes; Ana Marta Pereira[4], MSc; Armando Brito de Sá[5], MD, PhD

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[2]Public Health Research Center (CISP/UNL), Portuguese School of Public Health, Universidade Nova de Lisboa, Lisboa, Portugal

[3]NOVA-LINCS - Faculdade de Cie^ncias e Tecnologia, Faculty of Science and Technology, Universidade Nova de Lisboa, Lisboa, Portugal

[4]Faculty of Human and Social Sciences, Universidade Nova de Lisboa, Lisboa, Portugal

[5]Institute of Preventive Medicine, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

Corresponding Author:
Liliana Laranjo, MPH, MD, PhD
Centre for Health Informatics
Australian Institute of Health Innovation
Macquarie University
Level 6, 75 Talavera road
Sydney, 2113
Australia
Phone: 61 98502426
Fax: 61 98502499
Email: liliana.laranjo@gmail.com

Abstract

Background: Personal health records (PHRs) are increasingly being deployed worldwide, but their rates of adoption by patients vary widely across countries and health systems. Five main categories of adopters are usually considered when evaluating the diffusion of innovations: innovators, early adopters, early majority, late majority, and laggards.

Objective: We aimed to evaluate adoption of the Portuguese PHR 3 months after its release, as well as characterize the individuals who registered and used the system during that period (the *innovators*).

Methods: We conducted a cross-sectional study. *Users* and *nonusers* were defined based on their input, or not, of health-related information into the PHR. Users of the PHR were compared with nonusers regarding demographic and clinical variables. Users were further characterized according to their intensity of information input: *single input* (one single piece of health-related information recorded) and *multiple inputs*. Multivariate logistic regression was used to model the probability of being in the *multiple inputs* group. ArcGis (ESRI, Redlands, CA, USA) was used to create maps of the proportion of PHR registrations by region and district.

Results: The number of registered individuals was 109,619 (66,408/109,619, 60.58% women; mean age: 44.7 years, standard deviation [SD] 18.1 years). The highest proportion of registrations was observed for those aged between 30 and 39 years (25,810/109,619, 23.55%). Furthermore, 16.88% (18,504/109,619) of registered individuals were considered *users* and 83.12% (91,115/109,619) *nonusers*. Among PHR users, 32.18% (5955/18,504) engaged in *single input* and 67.82% (12,549/18,504) in *multiple inputs*. Younger individuals and male users had higher odds of engaging in multiple inputs (odds ratio for male individuals 1.32, CI 1.19-1.48). Geographic analysis revealed higher proportions of PHR adoption in urban centers when compared with rural noncoastal districts.

Conclusions: Approximately 1% of the country's population registered during the first 3 months of the Portuguese PHR. Registered individuals were more frequently female aged between 30 and 39 years. There is evidence of a geographic gap in the adoption of the Portuguese PHR, with higher proportions of adopters in urban centers than in rural noncoastal districts.

(JMIR Med Inform 2017;5(4):e37) doi:10.2196/medinform.7887

KEYWORDS

personal health records; diffusion of innovation; digital divide; patient participation; geographic information systems

Introduction

Person-centered care is a pivotal element in facilitating quality improvement in health care [1]. The growing shift from paternalism to shared decision making has unveiled the importance of people's access to their medical records and management of their health information [2,3]. Indeed, putting patients in control of their health information has been advocated as one of the solutions to the current fragmentation of health information [4,5].

Personal health records (PHRs) aim to fill the gap in personal health information management, empowering people to participate more actively in their own care. PHRs are electronic applications that enable individuals to access, manage, and share their health information in a private and secure environment [6-10]. PHRs may be classified according to their integration with the electronic health record (EHR) of a health care organization, going from *tethered* (ie, EHR-based patient portals), to stand-alone, or *untethered* PHRs (an example being Google Health, which was discontinued in 2012 for lack of widespread adoption) [2,6,9,11]. In the middle of this spectrum lie several types of hybrid PHRs, where patient-controlled functionalities are available (eg, patient-generated data entry), as well as some level of integration with EHRs [6]. Examples of hybrid PHRs include the three national government-funded PHRs that were developed in the United Kingdom, Portugal, and Australia.

Two different pathways to developing and implementing a national PHR were followed by the United Kingdom, Portugal, and Australia. In the first pathway, implemented in the United Kingdom (later discontinued) and in Portugal, the PHR is connected to a national shared record integrating EHR data from multiple providers from the National Health Service (NHS) (eg, primary health care centers and hospitals) [12-15]. In both countries, the PHRs were implemented in an opt-in model (ie, people had to actively sign up if they wanted to have an account), and the national shared records were created in an opt-out model, which means that there was implied consent for the creation of a record for each person [16]. The second pathway, followed by Australia, involves an opt-in PHR that is able to collect several summary documents from different providers without integrating them (ie, a national shared record is absent) [17].

Evidence regarding the effectiveness of PHRs for improving the quality of health care is increasing [18-21]. Published literature suggests that PHRs may lead to improvements in communication with health care providers[22-24], medication safety [24-26], medication adherence [27-30], satisfaction with care [7,22], and also in several processes of care [30-32], among other benefits. Furthermore, PHRs are increasingly being used in chronic disease management [33,34].

Nevertheless, despite the increasing deployment of PHRs by health care institutions and governments worldwide, their adoption by patients has remained slower than expected [2,12,14,35-39]. Several individual and sociotechnical factors are known to affect PHR adoption [40], such as the digital divide (ie, the gap that exists between individuals, groups, or communities in availability and use of technology) [41-43]. On the other hand, technology-related factors also play a role, such as the PHR's design, perceived usefulness, and perceived ease of use [44,45]. Indeed, better understanding of the factors that impact PHR adoption is a crucial step in the PHR research agenda [39,46].

One theoretical framework that has been previously applied to the adoption and use of PHRs is the *diffusion of innovations* model [11]. Rogers identified five main categories of adopters with respect to their time of adoption of an innovation: innovators, early adopters, early majority, late majority, and laggards [47]. The innovators are the first group to adopt an innovation and generally correspond to 2.5% of the population in a social system [47]. Previous research has defined PHR adoption innovators as the individuals signing up for the PHR in the initial 3-month period after deployment [11]. Identifying and characterizing the innovators in PHR adoption may be an important link in delineating a strategy for the *diffusion* of this technology.

In May 2013, a Web-based national PHR was officially launched in Portugal, provided freely by the Ministry of Health. This study aimed to evaluate the adoption of the Portuguese PHR for the first 3 months after its release. A further aim was to assess registered individuals in terms of their demographic characteristics, number of health problems and medications, and frequency of PHR use to input personal health information.

Methods

Setting

Portugal has a National Health Service (NHS) following the principle of universal coverage. Within the NHS, access to secondary and tertiary care is mostly done through general practice referrals. The majority of primary care practices in the country use the same EHR software; the opposite is true in secondary and tertiary care. There are huge variations in terms of hardware and bandwidth speed across primary care practices [48].

Internet access and use in Portugal is lower than in several other European countries: 57% of the population has Internet access at home, and 47.1% of the population never used the Internet [49]. There is a wide gap between younger and older generations: 90.6% of individuals in the age group of 15 to 24 years use the Internet, compared with only 5% of individuals above 65 years; also, Internet use varies with educational level, with lower education being associated with lower rates of Internet use [49]. Disparities in Internet use between districts in Portugal have sharply increased between 2008 and 2014, with urban districts showing much higher rates of Internet use when compared with rural noncoastal districts [50].

Geographical disparities in Portugal are also observed in health status. There is a direct association between population health status and the coastal location and urbanization of municipalities: those with a higher score of health status are located in the western coastal line of Portugal; lowest scores are observed in rural areas [51]. Furthermore, geographical location of health care facilities unequally affects the ease of

access of different groups of consumers, with poor accessibility to health services being particularly concerning for elderly individuals in less urbanized and rural areas of the country [52].

Patient involvement in health care is still relatively weak in Portugal [53], and patients' requests to access their medical records are rare, generally involving a lengthy administrative process [54]. Less than half of the population uses the Internet for health information retrieval [55].

Personal Health Record

The Portuguese PHR is a Web-based platform provided freely by the Ministry of Health and was officially released in May 2013 (after a 1-year period of beta-testing). At the time of data collection (end of July 2013), the PHR allowed patients to input health information (eg, health problems, chronic medication, and biometric measurements) and book primary care consultations. Integration between the PHR and the national shared record was negligible at the time of study, and very few patients had access to summary care records (SCRs), which were then being rolled out nationally (SCRs include aggregated EHR information, eg, list of health problems, current and past medications, and allergies). Furthermore, access by patients to their SCR was only possible via an e-card reader, a device rarely owned by the general public in Portugal. Planned features for future versions of the PHR included communication with health care providers, sharing data from the PHR with clinicians, prescription refills, and widespread access to SCRs.

One important factor enabling health data aggregation in Portugal is that patients registered with the NHS have a unique patient identifier (NHS number), which allows the correct integration of individual health data originating from different sources. Implementation of the PHR followed an opt-in model, which means that people had to actively register on the Web-based platform if they wanted to have an account. Access to the PHR was done through authentication with the individual NHS number and password after online registration occurred.

At the time of data collection, advertisement for the PHR was negligible, and there was no strategy in place to promote adoption. For this study, we were unable to determine the degree of public awareness regarding the existence of the PHR or to estimate the number of people reached by any communications about the PHR through mass media (eg, newspapers and radio).

Study Design

We conducted a cross-sectional study analyzing individual-level data from registered users of the Portuguese PHR. Data were collected by the information technology (IT) services of the Ministry of Health. The dataset provided to the research team was deidentified and corresponded to the first 3 months after release (May to July 2013). The study was approved by the Ethics Committee of Lisbon Medical School.

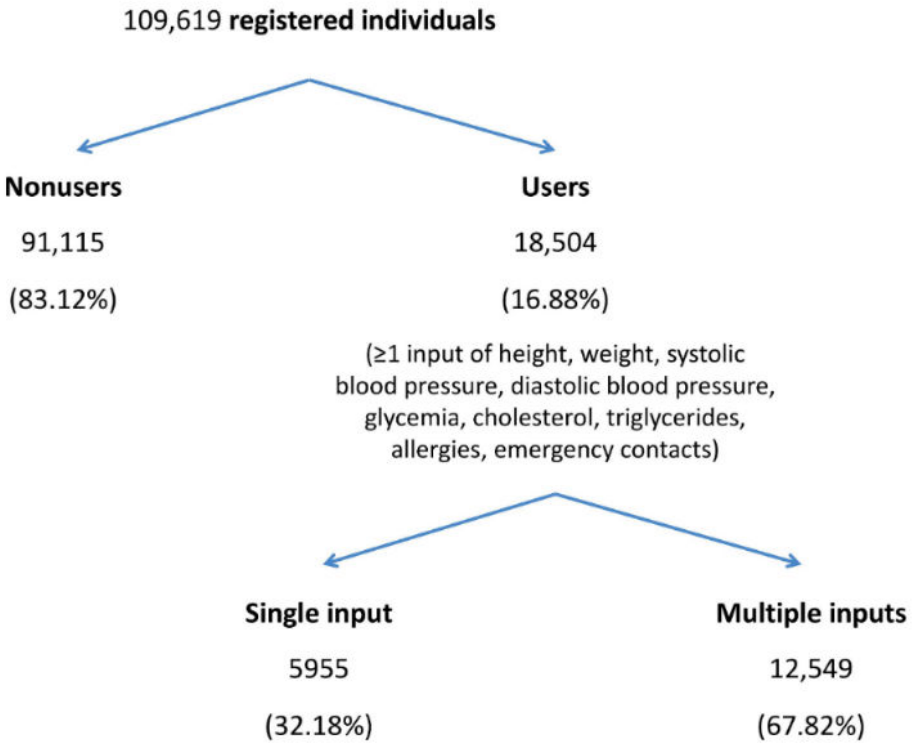
Individual-level data from patients registered in the PHR were collected regarding age, gender, region and district of residence, chronic conditions, chronic medication, and number of times information had been entered in specific PHR fields (emergency contacts, allergies, height, weight, systolic blood pressure, diastolic blood pressure, glycemia, cholesterol, and triglycerides). Data on age, gender, and residence were automatically populated in the PHR for each patient upon registration (these administrative data are associated with each NHS number). Remaining variables were self-reported (information entered by patients in the PHR).

As in other studies, we used heuristic definitions to characterize subgroups of adopters according to their use of the PHR [11,56-59]. We did not use log-ins as a proxy for PHR use and focused instead on the actual input of personal health information by individuals registered in the PHR. Consequently, for the purpose of this study, a classification was created to aid the characterization of PHR adopters (Figure 1). *Registered individuals* were defined as having an account created in the PHR, independent of their actual use of the platform to input information. *Users* were defined as having entered information in at least one of the following fields: allergies, emergency contacts, height, weight, systolic blood pressure, diastolic blood pressure, glycemia, cholesterol, or triglycerides levels. *Nonusers* were defined as individuals who had signed up for an account in the PHR, but who had not entered any of those data at the time of the study. Users were further divided into *single input* and *multiple inputs*, corresponding to the recording of either one or more than one piece of information regarding any of those data fields.

Statistical Analysis

SAS statistical software (version 9.2; SAS Institute, Inc) was used for all analyses. The distribution of continuous variables was checked for normality, and means and standard deviations (SDs) were calculated; proportions and counts were calculated for categorical variables. Univariate logistic regression models of the odds of being in the *multiple inputs* group, as a function of each individual predictor, were used to calculate crude odds ratios (ORs). Multivariate logistic regression was used to model the probability of being in the *multiple inputs* group, as a function of age category, gender, region of residence, number of health problems (categorical variable), and number of medications (categorical variable). The ArcMap functionality of ArcGis version 10 (ESRI, Redlands, CA, USA) was used to create maps of the proportion of PHR registrations by region and district.

Figure 1. Diagram representing the classification and distribution of registered individuals into "users" and "nonusers," as well as the classification of "users" into the "single input" group and the "multiple inputs" group.



Results

We identified 109,619 individuals registered in the PHR (60.58% women; mean age: 44.7 years, SD 18.1 years), which corresponded to approximately 1% of the Portuguese population (Table 1). The highest proportion of registrations was observed in the age category from 30 to 39 years (25,810/109,619; 23.55%). Geographic analysis revealed higher proportions of PHR adoption in urban centers when compared with the rural regions of the country (ie, noncoastal areas of Portugal) (Figure 2). The districts with the highest number of registered individuals were Lisbon and Oporto (Figure 2).

Among the 109,619 registered individuals, 91,115 (83.12%) had not entered any information in the PHR regarding emergency contacts, allergies, height, weight, systolic blood pressure, diastolic blood pressure, glycemia, cholesterol, or triglycerides. This group was classified as *nonusers* (Figure 1). The remaining 18,504 individuals were classified as PHR *users*, corresponding to 16.88% of registered individuals. Users provided a total of 45,039 entries in the above-specified fields, of which, the most common were height, weight, allergies, and emergency contacts (data not shown). Users tended to be male, younger, and Lisbon residents, when compared with nonusers (Table 1).

A total of 9543 health problems and 10,913 medications were self-reported by users and nonusers (Table 1). The most commonly reported health problems were high blood pressure, diabetes, and asthma (data not shown).

PHR users were further characterized as engaging in *single input* (5955/18,504, 32.18%) or *multiple inputs* (12,549/18,504, 67.82%) (Figure 1). The differences between them, as well as the crude and adjusted ORs are illustrated in Table 2. Younger individuals had higher odds of engaging in multiple inputs, as well as male users (adjusted OR for male individuals 1.32, CI 1.19, 1.48).

Table 1. Characteristics of the 109,619 individuals registered in the Portuguese personal health record (PHR) according to their use of the system to input information (*nonusers* vs *users*).

Characteristic		Nonusers n (%)	Users n (%)	Total n (%)
Age category (years)				
	<30	18,130 (19.90)	4189 (22.63)	22,319 (20.36)
	30-40	20,157 (22.12)	5653 (30.55)	25,810 (23.55)
	40-50	17,111 (18.78)	3601 (19.46)	20,712 (18.89)
	50- 65	20,603 (22.61)	3198 (17.28)	23,801 (21.71)
	≥65	15,114 (16.59)	1863 (10.07)	16,977 (15.49)
	Total	91,115 (83.12)	18,504 (16.88)	109,619 (100.00)
Gender				
	Female	56,585 (62.10)	9823 (53.09)	66,408 (60.58)
	Male	34,530 (37.90)	8681 (46.91)	43,211 (39.42)
Region				
	Lisbon and Tagus Valley	39,925 (43.82)	8414 (45.47)	48,339 (44.10)
	North	34,486 (37.85)	6698 (36.20)	41,184 (37.57)
	Other	16,704 (18.33)	3392 (18.33)	20,096 (18.33)
	Total	91,115 (83.12)	18,504 (16.88)	109,619 (100.00)
Health problems				
	None	230 (21.18)	1238 (14.64)	1468 (15.38)
	1	589 (54.24)	5058 (59.81)	5647 (59.17)
	2	149 (13.72)	1017 (12.03)	1166 (12.22)
	≥3	118 (10.87)	1144 (13.53)	1262 (13.22)
	Total	1086 (11.38)	8457 (88.62)	9543 (100.00)
Medications				
	0	255 (15.74)	1679 (18.07)	1934 (17.72)
	1	658 (40.62)	3793 (40.82)	4451 (40.79)
	≥2	707 (43.64)	3821 (41.12)	4528 (41.49)
	Total	1620 (14.84)	9293 (85.16)	10,913 (100.00)

Table 2. Characteristics of users (n=18,504) according to their input with crude and adjusted odds ratios.

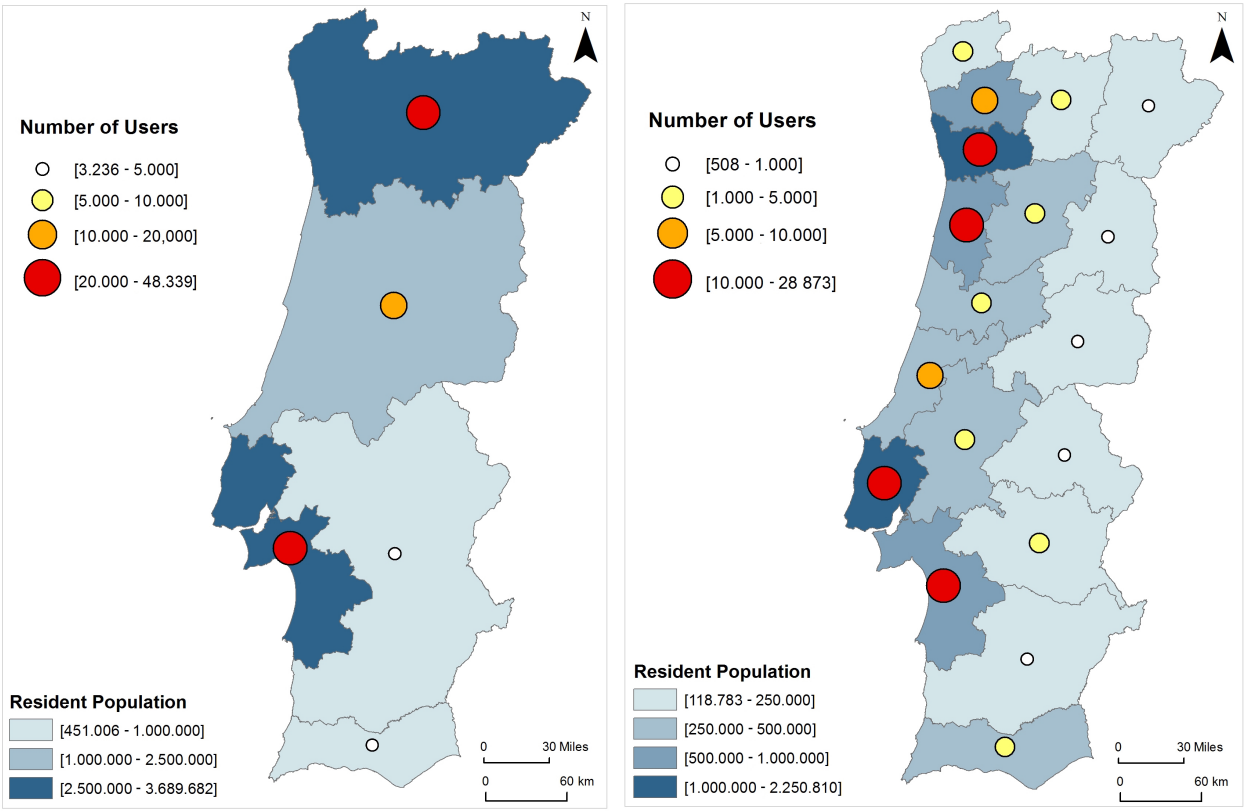
Characteristic[a]	Single input, n (%)	Multiple inputs, n (%)	Crude odds ratio[b] (95% CI)	Adjusted odds ratio[c] (95% CI)
Age category (years)				
<30	1106 (18.57)	3083 (24.57)	1.46 (1.32-1.60)	1.52 (1.29-1.80)
30-40	1694 (28.45)	3959 (31.55)	1.22 (1.12-1.33)	1.46 (1.25-1.7)
40-50	1235 (20.74)	2366 (18.85)	(Reference)	(Reference)
50-65	1126 (18.91)	2072 (16.51)	0.96 (0.87-1.06)	0.84 (0.71-1.0)
≥65	794 (13.33)	1069 (8.52)	0.7 (0.63-0.79)	0.60 (0.49-0.73)
Total	5955 (32.18)	12,549 (67.82)		
Gender				
Female	3342 (56.12)	6481 (51.65)	(Reference)	(Reference)
Male	2613 (43.88)	6068 (48.35)	1.20 (1.13-1.27)	1.32 (1.19-1.48)
Region				
Lisbon and Tagus Valley	2706 (45.44)	5708 (45.49)	(Reference)	(Reference)
North	2189 (36.76)	4509 (35.93)	0.98 (0.91-1.05)	0.95 (0.84-1.06)
Other	1060 (17.80)	2332 (18.58)	1.04 (0.96-1.14)	1.12 (0.96-1.30)
Total	5955 (32.18)	12,549 (67.82)		

[a]Some percentages do not total 100% due to rounding.

[b]Crude odds ratios calculated from univariate logistic regression where the probability of "multiple inputs" was modeled.

[c]Logistic regression model with predictors: age category, gender, region of residence, number of health problems, and medications.

Figure 2. Number of patients registered in the Portuguese Personal Health Record (PHR), by region (left image) and district (right image). The right-side image (district-level data) shows higher proportions of PHR adoption (largest circles, red) in urban areas (coastal districts on the left) than in rural areas (smallest circles, white, in the noncoastal districts on the right).



Discussion

Principal Findings

To the best of our knowledge, this is the first study analyzing the adoption of a PHR in Portugal. The number of registered individuals in the Portuguese PHR 3 months after its official release was 109,619 (approximately 1% of the Portuguese population; 60.6% women; mean age: 44.7 years, SD 18.1 years), from which 16.88% (18,504/109,619) were considered *users* and 83.12% (91,115/109,619) were *nonusers*. PHR users were also characterized as engaging in *single input* (5955/18,504, 32.18%) or *multiple inputs* (12,549/18,504, 67.82%) of information related to allergies, emergency contacts, height, weight, systolic blood pressure, diastolic blood pressure, glycemia, cholesterol, or triglycerides levels. Younger individuals had higher odds of engaging in multiple inputs, as well as male users. There was evidence of a geographic gap in the adoption of the Portuguese PHR, with higher proportions of adopters in urban centers than in rural noncoastal districts.

Comparison With Published Literature

The number of registered individuals in the Portuguese PHR 3 months after its release was 109,619, which is comparable to the adoption of other PHRs. For instance, Kaiser Permanente reported slightly higher adoption rates, with 58,734 members registering to use the site each month, on average [60]. In 2012, 4 million individuals (roughly 63% of the total number of members) had registered to use Kaiser Permanente's PHR, making it one of the most successful and actively used PHRs in the world [61].

On the other hand, national PHRs such as the United Kingdom's *HealthSpace* or Australia's personally controlled EHR have not had the same success in adoption. HealthSpace, which was introduced in the English NHS in 2007, had only 172,950 registrations by the end of October 2010 and ended up being discontinued in 2012 [14]. HealthSpace allowed individuals to input health information, communicate with clinicians, and access a summary care record, but research showed that the PHR was perceived by patients as neither useful nor easy to use [14]. Similarly, Australia's PHR had very low adoption rates since it was launched in 2012 [62]. After 3 years, with only 2 million people registered, the Australian government announced a "rescue package," including a change of its name to "My Health Record" and the move to an opt-out model, scheduled to begin in 2018 [63,64]. As of July 2017, less than 20% (4.78 million people) of Australia's population is registered for the My Health Record.

Our study contributes to the sparse literature on the development and implementation of national government-funded PHRs. Considering the turbulent paths of two well-known PHRs, developed in the United Kingdom and in Australia, future studies should evaluate how adoption of the Portuguese PHR is unfolding (particularly as new features become available), as well as analyze the sustainability of its use and the perceptions of patients and clinicians.

Characteristics of PHR adopters in our study are in line with findings from previous research: PHR registration and use was more frequent in women [42,65-67] and younger individuals [11], with lower registration rates and use being seen in people above 65 years [59,65,67-69]. The lower adoption by elderly patients should be further studied, as it may be associated with several different factors such as access and use of computers and the Internet, literacy, numeracy, and socioeconomic status [37]. Interestingly, studies have shown that, once enrolled, older patients were more likely to use the portal than their younger counterparts [65].

Geographic analysis revealed higher proportions of PHR adoption in urban centers than in the rural districts of the country (ie, noncoastal areas of Portugal). This divide is particularly apparent when geographical data are analyzed by district: the right-side image in Figure 2 shows that the largest circles (red), corresponding to higher proportions of PHR adoption, are located in urban areas, whereas the smallest circles (white), which indicate the lowest proportions of adoption, are located in rural areas.

The existence of an urban-rural gap raises concerns regarding the possible widening of disparities due to the digital divide [37,41,70]. In Portugal, rural areas have a higher proportion of elderly people and a less diversified network of health care services [52], showing lower scores of health status, when compared with coastal urbanized districts [51]. Additionally, Internet access and use is lower in the elderly, less educated, and those living in rural districts of the country. These groups are less likely to become adopters of a Web-based PHR, even when diffusion of this innovation spreads to early adopters and the early and late majority. Consequently, specific strategies may be needed to lessen the effects of the digital divide on existing health inequalities.

Disparities in PHR adoption have also been previously shown to be associated with race and ethnicity [42,65,67,68,70], as well as socioeconomic status, educational level [42,67,68,71], and health literacy [70], thereby raising concerns that access to this type of technology may be limited to a more socially advantaged population. For our study of PHR innovators, we were unable to access data on these types of variables. Nevertheless, if one considers the area of residence as a proxy for socioeconomic status [72], our findings reveal important disparities in PHR adoption. Given that innovators have previously been found to be similar to subsequent adopters in most sociodemographic characteristics [11], there is a concern that the urban-rural gap may be maintained as adoption of this PHR continues. Furthermore, innovators may act as opinion leaders or change agents in their communities, which could further contribute to perpetuate this gap. Research is needed to study current adoption of the Portuguese PHR, as well as to investigate its potential impact in widening health disparities.

Implications for Clinical Practice, Health Policy, and Research

Dissemination of PHRs will facilitate change into a more patient-centered model of care. This will require a significant cultural change in Portugal, where patients' access to their medical records and control of health information are still highly uncommon [54]. In addition, given that PHR adoption by patients is influenced by their providers' endorsement [73],

clinician involvement in the design, development, and implementation of PHRs seems crucial for their success. Despite concerns with the impact on workload, studies show that clinicians generally see several advantages in PHRs [74-76].

Continued adoption of the Portuguese PHR will depend on the availability of features that have been shown to be valued by patients, such as communication with providers, access to medical records, and administrative functionalities such as prescription refills [20,60,77-81]. For instance, registrations for Kaiser Permanente's PHR tripled when features such as online test results and emailing a doctor's office became available [60].

Previous studies of the Portuguese PHR found numerous usability problems, particularly in terms of readability and information architecture [80,82]. Additional studies evaluating ease of use of the current version of the Portuguese PHR are needed, as this is known to be a crucial aspect in the adoption of PHRs [11]. User-centered design strategies should help guide the development of PHR features and characteristics desired by citizens, with the aim of increasing PHR adoption [83,84].

Finally, attention should be given to the possible unanticipated consequences of the dissemination of this technology, such as the widening of inequalities and propagation of the inverse care law [85]. Therefore, ensuring universal Internet and computer access seems paramount, now that health care is increasingly reliant on IT [70]. At the same time, it is equally important to accommodate the needs of those not using or adopting this technology and make sure that they have access to the same quality of care as PHR adopters [86].

Considering the high costs associated with the development and implementation of a national PHR, independent evaluations of the implementation process should be conducted, as well an assessment of the potential value derived from this technology. Furthermore, the general public should have access to updated statistics on registrations and use of the Portuguese PHR, as well as to any evaluations that are carried out involving the system.

Future research should assess the evolution of the adoption curve, sustainability of use, perceptions of patients and clinicians, and the impact of the Portuguese PHR on process and outcome measures of health care. Furthermore, studies should evaluate the current users of the Portuguese PHR, as well as investigate possible signs of disparities between adopters and nonadopters.

Strengths and Limitations

This study has several strengths. We studied adoption both in terms of number of registrations and actual use of the PHR to input health information, providing a comprehensive perspective on the uptake of the PHR by citizens. We were able to collect and analyze individual-level data regarding region and district of residence, which allowed for the use of geographic information systems to study the geographic distribution of PHR adoption in the country. Also, the large sample size of our study provides robustness to the results.

Some limitations should also be recognized. Data analyzed in this study correspond to the period between May and July 2013. We were only able to obtain data corresponding to 3 months after the official PHR release, which is a relatively short period of time in the adoption curve. Consequently, these results might not be generalizable to the whole population of initial adopters. Furthermore, the number of features available in the PHR at the time of release was considerably less than what is provided nowadays, a fact that should be taken into consideration when interpreting these results.

We were unable to determine the degree of public awareness regarding the existence of the PHR, which would have been important to evaluate the context in which adoption occurred.

In light of the study design, selection bias and unmeasured confounding cannot be ruled out. Potentially important variables could not be evaluated, namely socioeconomic status, access to computers, and educational level. Furthermore, collection of ethnicity data in the EHR is not permitted in Portugal, hampering a comprehensive analysis of disparities in the adoption and use of PHRs by ethnic minority groups. Data regarding health problems was available for less than 9% (n=9543) of the total sample (N=109,619). Our definitions of *users*, *nonusers*, *multiple inputs*, and *single input* were conditioned by the particularities of this specific PHR and the data that we were able to collect. We included both dynamic (eg, blood pressure) and more static (eg, allergies) types of data to define PHR use and to characterize frequency of use. The use of these different types of data in the characterization of adopters should be further studied. Data regarding the booking of primary care consultations through the PHR or the access to summary care records were not available to researchers at the time of the study. Finally, this study was limited to a specific country, so caution should be exercised when trying to generalize the results to other populations and health care systems.

Conclusions

During the first 3 months after introducing the Portuguese PHR, 1% of the country's population registered to use it. Registered individuals were more frequently female and aged between 30 and 39 years. There is evidence of a geographic gap in the adoption of the Portuguese PHR, with higher proportions of adopters in urban centers than in rural noncoastal districts. Future research should assess the evolution of the adoption curve and investigate possible signs of disparities between adopters and nonadopters.

Acknowledgments

The authors would like to thank Dr David Bates, Dr Kenneth Mandl, Dr Joaquim Ferreira, Dr Albino Oliveira-Maia, and Dr Steffen Petersen for their insights during the design of this study. We would also like to thank the team at Serviços Partilhados do Ministério da Saúde, who facilitated data collection for this study, namely Henrique Martins, Diogo Reis, Álvaro Rebuge, Paulo Jorge Sá, and Daniel Cibrão. We thank Annie Lau for her feedback and for reviewing a draft of this paper.

This work was supported by a Junior Clinical Research award from the Harvard Medical School—Portugal program (HMSP-ICJ/0005/2010; Fundação para a Ciência e Tecnologia) attributed to the first author.

Authors' Contributions

All the authors contributed substantially to the design of the study, as well as the acquisition, analysis, and/or interpretation of data. The paper was drafted by the first author and critically reviewed by all the remaining authors. All the authors approved the final version to be published.

Conflicts of Interest

None declared.

References

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
2. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
3. Walker J, Darer JD, Elmore JG, Delbanco T. The road toward fully transparent medical records. N Engl J Med 2014 Jan 02;370(1):6-8. [doi: 10.1056/NEJMp1310132] [Medline: 24304001]
4. Steinbrook R. Personally controlled online health data--the next big thing in medical care? N Engl J Med 2008 Apr 17;358(16):1653-1656. [doi: 10.1056/NEJMp0801736] [Medline: 18420496]
5. Mandl KD, Kohane IS. Tectonic shifts in the health information economy. N Engl J Med 2008 Apr 17;358(16):1732-1737. [doi: 10.1056/NEJMsb0800220] [Medline: 18420506]
6. 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 Mar;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
7. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014 Jul;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
8. Mandl KD, Simons WW, Crawford WC, Abbett JM. Indivo: a personally controlled health record for health information exchange and communication. BMC Med Inform Decis Mak 2007 Sep 12;7:25 [FREE Full text] [doi: 10.1186/1472-6947-7-25] [Medline: 17850667]
9. Roehrs A, da Costa CA, Righi RD, de Oliveira KS. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
10. Bates DW, Wells S. Personal health records and health care utilization. J Am Med Assoc 2012 Nov 21;308(19):2034-2036. [doi: 10.1001/jama.2012.68169] [Medline: 23168828]
11. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012 Nov;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]
12. Greenhalgh T, Morris L, Wyatt JC, Thomas G, Gunning K. Introducing a nationally shared electronic patient record: case study comparison of Scotland, England, Wales and Northern Ireland. Int J Med Inform 2013 May;82(5):e125-e138. [doi: 10.1016/j.ijmedinf.2013.01.002] [Medline: 23434362]
13. Bratan T, Stramer K, Greenhalgh T. 'Never heard of it'- understanding the public's lack of awareness of a new electronic patient record. Health Expect 2010 Dec;13(4):379-391 [FREE Full text] [doi: 10.1111/j.1369-7625.2010.00608.x] [Medline: 20579117]
14. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. Br Med J 2010 Nov;341:c5814. [doi: 10.1136/bmj.c5814]
15. Greenhalgh T, Stramer K, Bratan T, Byrne E, Mohammad Y, Russell J. Introduction of shared electronic records: multi-site case study using diffusion of innovation theory. Br Med J 2008 Oct 23;337:a1786 [FREE Full text] [Medline: 18948344]
16. Greenhalgh T, Wood GW, Bratan T, Stramer K, Hinder S. Patients' attitudes to the summary care record and HealthSpace: qualitative study. Br Med J 2008 Jun 07;336(7656):1290-1295 [FREE Full text] [doi: 10.1136/bmj.a114] [Medline: 18511764]
17. Andrews L, Gajanayake R, Sahama T. The Australian general public's perceptions of having a personally controlled electronic health record (PCEHR). Int J Med Inform 2014 Dec;83(12):889-900. [doi: 10.1016/j.ijmedinf.2014.08.002] [Medline: 25200198]
18. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]

19. Heyworth L, Paquin AM, Clark J, Kamenker V, Stewart M, Martin T, et al. Engaging patients in medication reconciliation via a patient portal following hospital discharge. J Am Med Inform Assoc 2014 Feb;21(e1):e157-e162 [FREE Full text] [doi: 10.1136/amiajnl-2013-001995] [Medline: 24036155]
20. 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 06;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048]
21. Gustafson DH, Hawkins R, Boberg E, Pingree S, Serlin RE, Graziano F, et al. Impact of a patient-centered, computer-based health information/support system. Am J Prev Med 1999 Jan;16(1):1-9. [Medline: 9894548]
22. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005 Aug 05;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
23. Weingart SN, Carbo A, Tess A, Chiappetta L, Tutkus S, Morway L, et al. Using a patient internet portal to prevent adverse drug events: a randomized, controlled trial. J Patient Saf 2013 Sep;9(3):169-175. [doi: 10.1097/PTS.0b013e31829e4b95] [Medline: 23965840]
24. 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]
25. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012 Sep;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]
26. Chrischilles EA, Hourcade JP, Doucette W, Eichmann D, Gryzlak B, Lorentzen R, et al. Personal health records: a randomized trial of effects on elder medication safety. J Am Med Inform Assoc 2014 Jul;21(4):679-686 [FREE Full text] [doi: 10.1136/amiajnl-2013-002284] [Medline: 24326536]
27. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
28. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin C. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
29. Keith McInnes D, Shimada SL, Rao SR, Quill A, Duggal M, Gifford AL, et al. Personal health record use and its association with antiretroviral adherence: survey and medical record data from 1871 US veterans infected with HIV. AIDS Behav 2013 Jan 20;17(9):3091-3100. [doi: 10.1007/s10461-012-0399-3] [Medline: 23334359]
30. Lau AY, Sintchenko V, Crimmins J, Magrabi F, Gallego B, Coiera E. Impact of a web-based personally controlled health management system on influenza vaccination and health services utilization rates: a randomized controlled trial. J Am Med Inform Assoc 2012 Sep;19(5):719-727 [FREE Full text] [doi: 10.1136/amiajnl-2011-000433] [Medline: 22582203]
31. Horvath M, Levy J, L'Engle P, Carlson B, Ahmad A, Ferranti J. Impact of health portal enrollment with email reminders on adherence to clinic appointments: a pilot study. J Med Internet Res 2011 May;13(2):e41 [FREE Full text] [doi: 10.2196/jmir.1702] [Medline: 21616784]
32. Palen TE, Ross C, Powers JD, Xu S. Association of online patient access to clinicians and medical records with use of clinical services. J Am Med Assoc 2012 Nov 21;308(19):2012-2019. [doi: 10.1001/jama.2012.14126] [Medline: 23168824]
33. 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 08;168(16):1776-1782 [FREE Full text] [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]
34. Wagner PJ, Dias J, Howard S, Kintziger KW, Hudson MF, Seol Y, et al. Personal health records and hypertension control: a randomized trial. J Am Med Inform Assoc 2012 Jul;19(4):626-634 [FREE Full text] [doi: 10.1136/amiajnl-2011-000349] [Medline: 22234404]
35. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
36. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
37. Kim EH, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009 Oct;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
38. North F, Hanna BK, Crane SJ, Smith SA, Tulledge-Scheitel SM, Stroebel RJ. Patient portal doldrums: does an exam room promotional video during an office visit increase patient portal registrations and portal use? J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i24-i27 [FREE Full text] [doi: 10.1136/amiajnl-2011-000381] [Medline: 21976028]
39. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008 Nov;15(6):729-736 [FREE Full text] [doi: 10.1197/jamia.M2547] [Medline: 18756002]

40. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013 Feb;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
41. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
42. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009 Sep;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]
43. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017;24(e1):e47-e54. [doi: 10.1093/jamia/ocw098] [Medline: 27402138]
44. Venkatesh V, Morris M, Davis G, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
45. Lazard AJ, Watkins I, Mackert MS, Xie B, Stephens KK, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. J Am Med Inform Assoc 2016 Apr;23(e1):e157-e161 [FREE Full text] [doi: 10.1093/jamia/ocv174] [Medline: 26635314]
46. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
47. Rogers EM. Diffusion of innovations. New York City, NY: Free Press; 2003.
48. Granja M, Ponte C, Cavadas LF. What keeps family physicians busy in Portugal? A multicentre observational study of work other than direct patient contacts. Br Med J Open 2014 Jun 15;4(6):e005026 [FREE Full text] [doi: 10.1136/bmjopen-2014-005026] [Medline: 24934208]
49. Obercom. 2012. A sociedade em rede: A Internet em Portugal [Portuguese] URL: https://obercom.pt/wp-content/uploads/2016/06/A-Sociedade-em-Rede-A-Internet-em-Portugal-2012.pdf [accessed 2017-10-02] [WebCite Cache ID 6tvale8rT]
50. Autoridade Nacional de Comunicações (ANACOM). Anacom. 2015. Evolucao dos acessos a Internet em Portugal URL: https://www.anacom.pt/streaming/estudoSpeedtest2015velocidadesBL.pdf?contentId=1374020&field=ATTACHED_FILE
51. Santana P, Vaz A, Fachada M. O estado de saúde dos Portugueses - uma perspectiva espacial. Rev Estud Demograficos 2004;36(1):5-28.
52. Santana P. Ageing in Portugal: regional iniquities in health and health care. Soc Sci Med 2000 Apr;50(7-8):1025-1036. [doi: 10.1016/S0277-9536(99)00352-4]
53. OECD. Portugal 2015: raising standards. In: OECD Reviews of Health Care Quality. Paris: OECD Publishing; 2015.
54. Laranjo L, Neves A, Villanueva T, Cruz J, Brito de Sá A, Sakellarides C. [Patients' access to their medical records]. Acta Med Port 2013;26(3):265-270 [FREE Full text] [Medline: 23815842]
55. Kummervold PE, Wynn R. Health information accessed on the Internet: the development in 5 European countries. Int J Telemed Appl 2012:1-3. [doi: 10.1155/2012/297416]
56. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient portals: who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501 [FREE Full text] [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
57. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN. Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 2014 Oct 24;16(10):e241 [FREE Full text] [doi: 10.2196/jmir.3371] [Medline: 25344621]
58. Jones JB, Weiner JP, Shah NR, Stewart WF. The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res 2015 Feb 20;17(2):e42 [FREE Full text] [doi: 10.2196/jmir.3157] [Medline: 25707036]
59. Jung SY, Lee K, Hwang H, Yoo S, Baek HY, Kim J. Support for sustainable use of personal health records: understanding the needs of users as a first step towards patient-driven mobile health. JMIR Mhealth Uhealth 2017 Feb 23;5(2):e19 [FREE Full text] [doi: 10.2196/mhealth.6021] [Medline: 28232300]
60. Silvestre AL, Sue VM, Allen JY. If you build it, will they come? The Kaiser Permanente model of online health care. Health Aff (Millwood) 2009 Mar;28(2):334-344 [FREE Full text] [doi: 10.1377/hlthaff.28.2.334] [Medline: 19275988]
61. Healthcareitnews. Kaiser PHR sees 4 million sign on, most active portal to date URL: http://www.healthcareitnews.com/news/kaiser-phr-sees-4-million-sign-most-active-portal-date [accessed 2017-07-20] [WebCite Cache ID 6s605UhrB]
62. Xu J, Gao X, Sorwar G, Croll P. Implementation of e-health record systems in Australia. Int Technol Manag Rev 2013;3:92-104.
63. Glance D. Theconversation. 2015. New name and opt-out policy won't save the personal health record URL: https://theconversation.com/new-name-and-opt-out-policy-wont-save-the-personal-health-record-41601 [accessed 2017-09-24] [WebCite Cache ID 6ti0PpjmU]

64. Abc.net. Budget 2015: new 'opt out' e-health system to see all Australians given electronic record URL: http://www.abc.net.au/news/2015-05-10/government-to-fund-ehealth-for-all-australians/6457940 [accessed 2017-07-20] [WebCite Cache ID 6s62G1f9k]
65. 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]
66. Ralston JD, Rutter CM, Carrell D, Hecht J, Rubanowice D, Simon GE. Patient use of secure electronic messaging within a shared medical record: a cross-sectional study. J Gen Intern Med 2009 Mar;24(3):349-355 [FREE Full text] [doi: 10.1007/s11606-008-0899-z] [Medline: 19137379]
67. Ancker JS, Barrón Y, Rockoff M, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
68. Weingart SN, Rind D, Tofias Z, Sands DZ. Who uses the patient internet portal? The PatientSite experience. J Am Med Inform Assoc 2006 Jan;13(1):91-95 [FREE Full text] [doi: 10.1197/jamia.M1833] [Medline: 16221943]
69. 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]
70. 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 Aug;15(Suppl 2):183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
71. Ancker JS, Silver M, Kaushal R. Rapid growth in use of personal health records in New York, 2012-2013. J Gen Intern Med 2014 Jun;29(6):850-854 [FREE Full text] [doi: 10.1007/s11606-014-2792-2] [Medline: 24519102]
72. Santana P. Poverty, social exclusion and health in Portugal. Soc Sci Med 2002 Jul;55(1):33-45. [doi: 10.1016/S0277-9536(01)00218-0]
73. Lyles CR, Sarkar U, Ralston JD, Adler N, Schillinger D, Moffet HH, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: the Diabetes and Aging Study. J Am Med Inform Assoc 2013 Nov;20(6):1128-1131 [FREE Full text] [doi: 10.1136/amiajnl-2012-001567] [Medline: 23676243]
74. Ross SE, Todd J, Moore LA, Beaty BL, Wittevrongel L, Lin C. Expectations of patients and physicians regarding patient-accessible medical records. J Med Internet Res 2005 May 24;7(2):e13 [FREE Full text] [doi: 10.2196/jmir.7.2.e13] [Medline: 15914460]
75. Miller DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
76. Vydra TP, Cuaresma E, Kretovics M, Bose-Brill S. Diffusion and use of tethered personal health records in primary care. Perspect Heal Inf Manag 2015;12(1):1-25. [Medline: PMC4696089]
77. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to Electronic Health Records: results from a survey of veterans. Med Care 2013 Mar;51(3):S52-S56. [doi: 10.1097/MLR.0b013e31827808db]
78. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente electronic health record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009 Mar 01;28(2):323-333. [doi: 10.1377/hlthaff.28.2.323]
79. Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care 2007 Jul;13(7):418-424 [FREE Full text] [Medline: 17620037]
80. Rodolfo I, Laranjo L, Correia N, Duarte C. Design strategy for a national integrated personal health record. 2014 Presented at: Nordic Conference on Human-Computer Interaction - NordiCHI; October 26-30; Helsinki, Finland. [doi: 10.1145/2639189.2641205]
81. Lau AY, Proudfoot J, Andrews A, Liaw S, Crimmins J, Arguel A, et al. Which bundles of features in a web-based personally controlled health management system are associated with consumer help-seeking behaviors for physical and emotional well-being? J Med Internet Res 2013 May;15(5):e79 [FREE Full text] [doi: 10.2196/jmir.2414] [Medline: 23649790]
82. Rodolfo I, Laranjo L, Correia N, Duarte C. The importance of mental models in the design of integrated PHRs. 2014 Presented at: American Medical Informatics Association 2014 Annual Symposium; 2014; Washington DC.
83. Arsand E, Demiris G. User-centered methods for designing patient-centric self-help tools. Inform Health Soc Care 2008 Sep;33(3):158-169. [doi: 10.1080/17538150802457562] [Medline: 18850399]
84. Oudshoorn N, Pinch T. How Users Matter: The Co-Construction of Users and Technologies. Cambridge: MIT Press; 2003.
85. Hart JT. The inverse care law. Lancet 1971 Feb 27;1(7696):405-412. [Medline: 4100731]
86. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. The digital divide and patient portals: internet access explained differences in patient portal use for secure messaging by age, race, and income. Med Care 2016;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]

Abbreviations

EHR: electronic health record
IT: information technology
NHS: National Health Service
PHR: personal health record
SCR: summary care record

Edited by T Kool; submitted 19.04.17; peer-reviewed by T Vreugdenhil, C Pearce, N Archer; comments to author 21.05.17; revised version received 21.07.17; accepted 03.09.17; published 11.10.17.

Please cite as:
Laranjo L, Rodolfo I, Pereira AM, de Sá AB
Characteristics of Innovators Adopting a National Personal Health Record in Portugal: Cross-Sectional Study
JMIR Med Inform 2017;5(4):e37
URL: http://medinform.jmir.org/2017/4/e37/
doi:10.2196/medinform.7887
PMID:29021125

Original Paper

Patient Portal Use and Experience Among Older Adults: Systematic Review

Dawn K Sakaguchi-Tang[1*], MS; Alyssa L Bosold[2*], BA; Yong K Choi[3*], MPH; Anne M Turner[2,4*], MLIS, MPH, MD

[1]Department of Human Centered Design and Engineering, College of Engineering, University of Washington, Seattle, WA, United States

[2]Department of Health Services, School of Public Health, University of Washington, Seattle, WA, United States

[3]Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States

[4]Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

[*]all authors contributed equally

Corresponding Author:
Dawn K Sakaguchi-Tang, MS
Department of Human Centered Design and Engineering
College of Engineering
University of Washington
3960 Benton Lane NE
428 Sieg Hall
Seattle, WA, 98195
United States
Phone: 1 206 543 2567
Fax: 1 206 543 8858
Email: dawnsaka@uw.edu

Abstract

Background: The older adult population (65 years or older) in the United States is growing, and it is important for communities to consider ways to support the aging population. Patient portals and electronic personal health records (ePHRs) are technologies that could better serve populations with the highest health care needs, such as older adults.

Objective: The aim of this study was to assess the existing research landscape related to patient portal and ePHR use and experience among older adults and to understand the benefits and barriers to older adults' use and adoption of patient portals and ePHRs.

Methods: We searched six pertinent bibliographic databases for papers, published from 2006 to 2016 and written in English, that focused on adults 60 years or older and their use of or experience with patient portals or ePHRs. We adapted preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines to review papers based on exclusion and inclusion criteria. We then applied thematic analysis to identify key themes around use, experience, and adoption.

Results: We retrieved 199 papers after an initial screening and removal of duplicate papers. Then we applied an inclusion and exclusion criteria, resulting in a final set of 17 papers that focused on 15 separate projects. The majority of papers described studies involving qualitative research, including interviews and focus groups. They looked at the experience and use of ePHRs and patient portals. Overall, we found 2 main barriers to use: (1) privacy and security and (2) access to and ability to use technology and the Internet. We found 2 facilitators: (1) technical assistance and (2) family and provider advice. We also reported on older adults' experience, including satisfaction with the system and improvement of the quality of their health care. Several studies captured features that older adults wanted from these systems such as further assistance managing health-related tasks and contextual health advice and tips.

Conclusions: More research is needed to better understand the patient portal experience among older adults from initial use to adoption. There are also opportunities to explore the role of design in addressing barriers and supporting facilitators to patient portal and ePHR use. Finally, the future use of these systems by older adults should be anticipated and considered in the design process.

(JMIR Med Inform 2017;5(4):e38) doi:10.2196/medinform.8092

KEYWORDS

aged; patient portals; personal health records; utilization; usability; user experience

Introduction

Background

In 2014, the adult population aged 65 years or older in the United States was 46.2 million, and this number is projected to increase to 98 million by 2060 [1]. With this expected growth in the older adult population, it is essential for communities to consider ways to support their aging population. To this end, there has been a growing interest in the design of technologies for older adults, including technologies that can support older adults through health maintenance and health information management. Such technologies have the potential to support older adults by allowing them to age in their own homes, maintain their health, and provide a sense of autonomy. Although there have been gains in the adoption of technology by adults 65 years or older, older adults have consistently trailed the general American population, especially in adopting digital health technologies [2,3].

Interest in electronic health records (EHRs), patient portals, and electronic personal health records (ePHRs) has increased in recent years [4-6]. Ancker et al (2016) [6] conducted a survey of New York State residents to understand the rate of patient portal and personal health records (PHRs) adoption over time. They found that use of PHRs by New Yorkers increased from 11% in 2012 to 27.1% in 2015. Ford et al (2016) [5] forecasted the adoption of PHRs based on the 2008, 2011, and 2013 Health Information National Trends Surveys. They anticipated that PHR adoption will grow beyond 75% by 2020. These studies show that the use of patient portals and PHRs will likely continue to grow.

The level of research on this topic raises awareness about how these technologies are being used and has implications for improved and innovative design. Research on digital health technology adoption by older adults also signals a focus on how technology could better serve populations with the highest needs, who often manage complex health conditions and multiple chronic illnesses. The incidence of multiple chronic conditions increases with age [7], and the prevalence of some chronic conditions such as hypertension, asthma, cancer, and diabetes has also increased among older adults [8]. Hospitals, clinics, and organizations have started to offer patients a way to stay connected to their health information and manage their wellness and health care needs through patient portals and ePHRs.

There are several definitions of ePHRs and patient portals within the literature, and patient portals are sometimes described as a type of ePHR. For the purposes of this paper, *patient portals* are defined as systems for health information management that are linked, or tethered, to a patient's EHR [9,10]. For example, the US Department of Veterans Affairs offers patients access to My Health *e* Vet [11], and several hospitals in the United States use Epic's MyChart portal [12]. Both portals give patients access to their health information and include features such as the ability to schedule appointments, view test results, request prescription renewals, and send messages to health care providers. In addition to tethered patient portals, there are *ePHRs* that are not connected to EHRs, such as Microsoft HealthVault and the Health app on Apple devices. In these systems, the individual is responsible for entering their own health information. ePHR systems often include features such as health tracking or medication lists. Other features of these systems include the ability to share health information with others and track fitness and personal health goals. The major distinction between ePHRs and patient portals is that patient portals are tethered (to EHRs) and ePHRs are not. Both offer a centralized location for storing and organizing electronic health information.

Objectives

Although much has been written about the use of patient portals and ePHRs in general, there is less material focused on the use of patient portals by older adults. Technologies such as patient portals and ePHRs have the potential to help older adults by strengthening their ability to manage, understand, and control their health information. However, it is a leap to assume that patient portals and ePHRs, as they are currently designed and used, will effectively address the health information needs of the older adult population. It is important to first understand the facilitators of and barriers to older adult use and adoption of health-related technology. It is also important to understand their experiences with ePHRs and patient portals and how these experiences have influenced or changed their personal health information management. Understanding the facilitators and barriers will provide insights to why older adults decide to use or adopt patient portals and ePHRs. Similarly, learning about older adults' experiences with these systems and their impact on health information management can provide guidance on how to improve their design and ensure their effective use and adoption. Finally, it is important to understand what design recommendations have been proposed, and what is important to older adults. Considering these objectives, the goal of this systematic review was to investigate the existing research landscape with a focus on answering the following questions:

1. In the literature, what barriers and facilitators to older adults' use and adoption of patient portals and ePHRs have been described? What is the evidence that these barriers and facilitators exist?
2. How do older adults describe their experience using patient portals and ePHRs?
3. What design recommendations have been proposed to help overcome barriers and enhance facilitators of older adults' experience, use, and adoption of patient portals or ePHRs?

Methods

Revised PRISMA protocol

We adapted the preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2009 checklist to guide our systematic review of the use of patient portals and ePHRs among older adults [13]. As PRISMA is positioned toward standardized study designs, such as clinical trials that aim to support universal

interpretation of results, we modified the PRISMA protocol to accommodate the study methodologies in this review more common to information sciences, specifically qualitative and mixed-method studies. Thus, we reviewed the methods and metrics used in the studies rather than the standardized outcome variables one would typically see in traditional systematic reviews of controlled trials. Our protocol included a systematic search, a study selection, and a qualitative review of the findings.

Literature Search

We conducted our search in six databases that spanned the medical, nursing, and engineering literature. These databases were PubMed, EMBASE, CINAHL Complete, Compendex (includes ACM digital library and IEEE XPlore), and Inspec. We consulted with librarians in the University of Washington Health Sciences and Engineering libraries on the selection of databases and the mechanics of using them (eg, controlled vocabulary, using filters, and syntax). We also received assistance narrowing down keywords to use. We searched all databases with the keywords "older adult," "seniors," or "elders," and "patient portal," "electronic medical record," or "personal health record" (see Table 1). We did a general search in Google Scholar to find potential papers that did not result from our searches in the other databases. In PubMed and EMBASE, we used additional keywords such as usage, utilization, adoption, and patient satisfaction. We did not use the additional keywords in CINAHL Complete, Compendex, and Inspec because it narrowed rather than broadened our search results. We limited our search to papers published within a 10-year period (January 2006-November 2016). Although we recognize that a 10-year period is a broad timeline given the fast pace of advancement in technology, we selected this time range to get an expanded view about needs and experiences of older adults related to health information technology, and included commentary on changes in technology and findings over time.

Table 1. Searches used in each database.

Database	Description	Citations
PubMed	older adult OR seniors OR elderly OR aged AND patient portal OR electronic health record OR personal medical record OR personal health record AND usage OR using OR utilization OR utilize OR adopt OR adoption OR preferences OR patient access to records OR patient satisfaction AND english NOT letter OR editorial AND last 10 years	885
EMBASE	older adult OR older adults OR seniors OR elderly OR aged OR aged AND patient portal OR electronic medical record OR personal medical record OR personal health record AND usage OR utilization OR utilize OR adopt OR adoption OR preference OR patient access to records OR patient attitude AND english AND [embase]/lim NOT [medline]/lim AND (2006-2016)/py	409
CINAHL	patient portal OR electronic health record OR personal health record older adult OR senior OR elder Limiters: published date: 2006-01-01-2016-12-31; English language	129
Compendex and Inspec	older adult OR senior OR elder AND electronic health record OR personal health record AND 2006-2016 AND english	484

Textbox 1. Inclusion criteria.

Inclusion criteria

- Include participation of adults who were 60 years or older. These older adults could be the sole focus of the study or be a group of adults who were part of a larger study. Typically, older adults are characterized as 65 years and older; however, we decided to use a wider age range to include a broader set of papers.
- Focus on patient portals or ePHRs
- Discuss use, adoption, or experience with patient portals and ePHRs or features of those systems (eg, studies that evaluated patient experiences using secure messaging with providers, or having electronic access to medical records)
- Examine features of patient portals and ePHRs to inform design
- Published from 2006 to 2016
- Written in English

Inclusion and Exclusion Criteria

Papers were selected based on the inclusion criteria in (Textbox 1) and exclusion criteria provided here. We excluded studies that were not focused on older adults' use, experience, or adoption of patient portals, ePHRs, or features of those systems. Although studies do not consistently report clear definitions of use and adoption, we chose to differentiate between these two terms for this review. Specifically, we refer to *use* as short-term activity within a patient portal for a period of less than 1 year, whereas we define *adoption* as a commitment to continued use of systems beyond 1 year. We defined *experience* as a person's perceptions of their interactions with patient portals or ePHRs. We also included *formative* studies that were focused on information gathering for design, including user testing of new systems and assessments to inform development of systems or test the acceptability of particular systems. Formative studies were not focused on adoption or use or factors influencing the

initial use of a particular developed system. The types of papers that were excluded were studies focused on patient online communities or the provider experience using patient portals or ePHRs. Papers that solely recorded log-in data and demographics were also not considered to be focused on use and were excluded from this review. In addition, we excluded nonempirical studies such as commentaries, letters to the editors, notes, books, reviews, and conceptual papers.

One researcher (DST) conducted an initial screening of the paper titles and abstracts, removing records that were irrelevant such as those focused on provider experience, implementation of EHRs, and using EHRs to recruit participants. Then 3 researchers applied the inclusion and exclusion criteria to the abstract of each paper using the Covidence (Melbourne, Victoria) software [14]. Each paper was reviewed by at least 2 of the 3 researchers (DST, AB, and YC), and any disagreements were discussed. In cases where a resolution could not be reached, a third researcher made the final decision. After excluding an initial set of papers, the same 3 researchers applied the inclusion and exclusion criteria to the full text of the papers using the same process described above to resolve disagreements.

After applying the inclusion and exclusion criteria, we conducted a thematic analysis of the papers. Two researchers (DST and AB) created codes using an inductive process. They summarized each of the papers and collectively came up with a list of key points from the summaries and from the papers themselves. These key points were then grouped into codes. The groups of codes were then further refined into themes, and the final list of themes was informed by the project's research questions and decided collectively in a meeting with the team researchers.

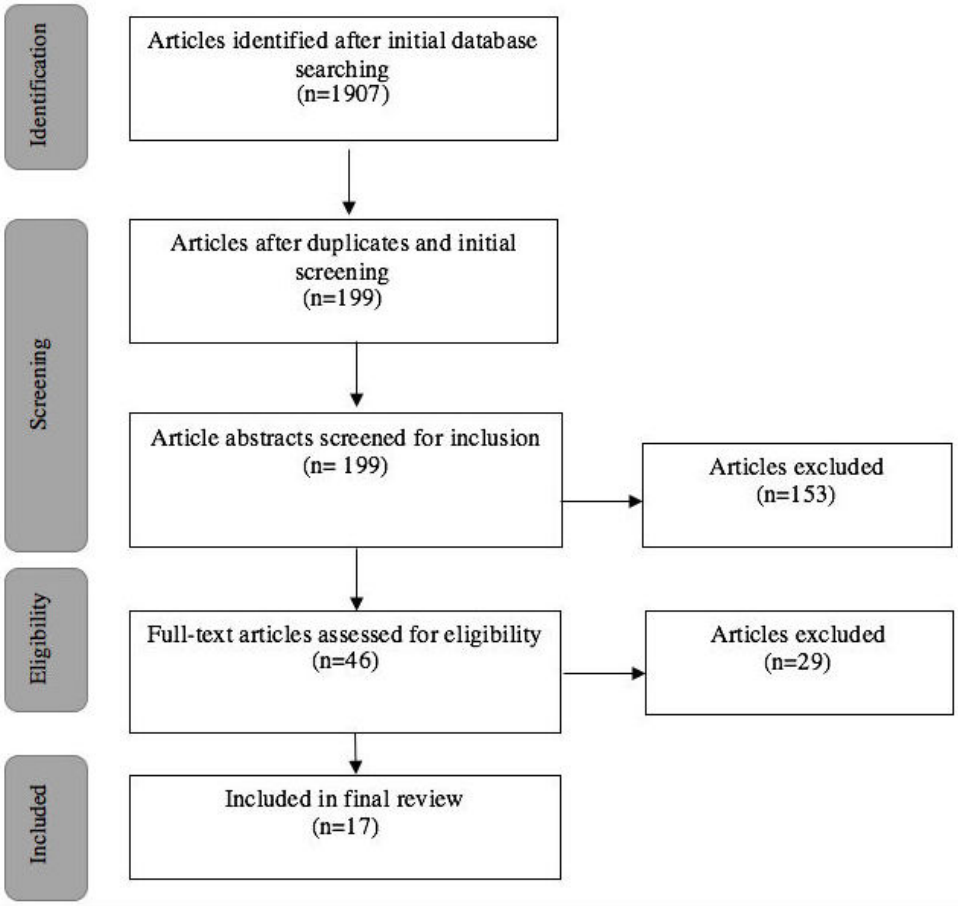
Quality Review

We reviewed the papers using the top two guidelines from the mini Statement on the Reporting of Evaluation studies in Health Informatics (STARE-HI), ranked as essential by professionals in health informatics for reporting studies [15]. They were "Interpret the data and an answer to the study question" and "Description of the outcome measure or evaluation criteria." We added two additional guidelines because they provided key information related to our study questions: "Provides a description of system and its functionalities" [16] and "Provides clear description of how results impact design recommendations." We gave the papers a score for each of the four guidelines outlined above. The score ranged between 0 (does not meet the criteria) and 2 (fully meets the criteria), for a total score of 8.

Results

The search returned 1907 papers in total after removing duplicates. An initial screening of paper titles and abstracts resulted in 199 papers. Abstract review, described above, resulted in 46 papers for full-text review. The full-text review resulted in a final set of 17 papers (see Figure 1).

Figure 1. Systematic review process.



Description of Papers

The final set of 17 papers focused on 15 separate projects (see Multimedia Appendix 1). Papers spanned the 10-year period from 2006 to 2016. All papers published before 2014 examined ePHRs, whereas those papers published from 2014 to 2016, with the exception of one [17], looked at patient portals. Of the 17 papers, 7 (41%) were conference proceedings. All conference proceedings were peer-reviewed. Authors used a range of research methods in the final set of papers: 10 of 17 (59%) were interviews, observations, focus groups, design sessions, and user studies; 9 of 17 (53%) were surveys or questionnaires; and 4 of 17 (24%) were mixed-methods studies. The sample size of the papers ranged from 16 participants in a user study to 231,082 participants in a survey. Six papers focused on patient portals [9,10,18-21], 8 papers focused on ePHRs [22-29], and 2 papers looked at other similar systems, specifically a personal health application and the Swedish medication registry [30,31]. Half (8/17) of the papers evaluated patient portals or ePHR systems overall [9,18-20,23,24,26,29]; others focused on specific features such as messaging systems [25] or medication management tools [22,23,30,31].

Seven papers focused on short-term use or factors influencing the initial use of a system. Nine papers were primarily formative, collecting information related to system design, development, or usability. Formative papers collected information to inform design of systems generally [9,10,20,21,28,30] or focused on developing specific systems [18,22,23]. Only 3 papers compared short-term use and long-term adoption [17,24,25]. In 2 cross-sectional papers, Lam et al (2013) [25] and Zettel-Watson and Tsukerman (2016) [17], participants most commonly reported using systems anywhere from 1 month to 1 year and reported an average period of use of over 3 years, respectively. In the Kim et al (2009) [24] paper that looked at patterns of use longitudinally, 51% of the participants only used the system once during the first year of the study period.

Qualitative and cross-sectional papers provided insight into both specific systems and general experience. In 3 of the 17 papers (18%), participants used a system and were given a survey or questionnaire to gain feedback on their experience [19,25,31]. There were 2 papers (12%) that evaluated a system in a lab setting [9,18] and 3 papers, focused on two projects, (18%) [24,26] where participants used a system in a community setting such as a retirement or housing facility [22,24,26]. Four papers (24%,) did not focus on a specific system but instead asked participants to reflect on their experiences with patient portals or ePHRs in general [10,17,21,27]. Two papers (12%, 2/17) focused on developing a personal health application with participants [23,30]. Another approach that 3 papers (18%, 3/17) took was to gather information needs from participants through qualitative methods such as interviews, design sessions, and a diary method to inform design of a system [21,22,28] (see Multimedia Appendix 2).

Participant Characteristics

Demographic details about participants are provided in Multimedia Appendix 2. All papers had participants who were 65 years or older [9,10,17-31]. Two papers [19,31] analyzed differences between age categories within the older adult group. In all other papers, the older adults were reported as one group. Of the 13 studies that reported gender, 11 had more female than male participants [9,10,17,19-21,24,26,27,30,31].

Quality Review

All of the papers met the criteria "Interpret the data and an answer to the study question," and almost all (14 of 17) met the criteria "Description of the outcome measure or evaluation criteria." The last two criteria were more varied. There were 7 papers that did not provide enough detail about a system and its functionalities [17-19,21,22,29,30]. For example, Sack et al (2011) [29] evaluated PHR technologies using Web and mobile-based Google Health. In discussing the technologies, they did not provide details about the features or functions of the system beyond it being Web or mobile-based. Papers were given full points if they provided a definition including functionality for a patient portal or an ePHR. Descriptions of the system provide a context for results and recommendations. It also provides a status of the technology at that time.

The other criterion that papers were varied on was "Provide clear description of how results impact design recommendations." Although a majority of papers did not aim to provide design recommendations, one of our research questions was to learn about design recommendations that have been proposed to address the barriers and facilitators to use, adoption, and experience. We did find that 15 of 17 papers connected their findings to design considerations or suggestions for improving use of system [9,10,17-23,25,27-31], for example, training to increase adoption [27]. Papers were given a partial score if their recommendations were brief and vague. Papers received full points if authors offered clear considerations for design and gave detailed recommendations. For detailed ratings, see Multimedia Appendix 3.

Barriers

We found commonalities among all papers concerning barriers and facilitators to the use and adoption of patient portals or ePHRs by older adults. We identified two main barriers across studies: (1) privacy and security and (2) access and ability to use technology and the Internet.

Privacy and Security

In 7 papers, older adults expressed a concern about the privacy and security of their information when using patient portals, ePHRs, or Web-based health management tools [10,17,20-22,26,28]. Privacy and security concerns were linked to the storage and use of data collected in patient portals. Hourcade et al (2011) [22] reported that participants were worried about pharmaceutical or drug companies accessing and misusing their data. Despite reassurance that the research was confidential and for academic purposes, participants expressed worry that researchers might not fully disclose partnerships with government institutions or drug companies. In the Kerai et al (2014) [20] paper, 63% of participants were concerned about security. Participants in the Latulipe et al (2015) [21] paper were concerned that the government or insurance companies would access their records without their permission. In the Lober et al (2006) [26] paper, participants were living in a government housing authority and had to be able to live independently to

stay there. They were protective of their health information because they did not want to be evicted if their physical health limited their ability to remain independent.

Access and Abilities

Lack of access to technology and the Internet was mentioned as a barrier in 5 papers. However, the results from these papers are based on small sample sizes, and two of them were focused on lower income communities. In the papers we reviewed, disparities in age, race, and ability to pay for the Internet were mentioned. Turner et al (2015) [10] reported that some of their participants had difficulty accessing the Internet because of its cost. Logue and Effken (2012) [27] identified that older and younger seniors had similar access to computers but differed in Internet access. Seniors over the median age of 78 years had less access to and familiarity with the Internet than seniors aged under 78 years [27]. Of the 38 participants in the Lober et al (2006) [26] study, 27 (71%) did not own computers. Latulipe et al (2015) [21] reported that older adults were aware of Internet access in their communities, and over half had a digital device such as a computer, laptop, or tablet. However, some participants did not have access to the Internet at home, suggesting that the devices were not being used [21]. Two papers noted gendered differences in Internet access, but results were mixed [20,27].

Seven papers defined computer and Internet skills as a barrier, and papers focused on both actual and perceived abilities. Lober et al (2006) [26] reported that major barriers to use of their portal system were computer literacy and computer anxiety. They described computer literacy as instances where participants were unable to do tasks on their own, such as turning on the computer or using a mouse or keyboard. Computer anxiety was a refusal to complete tasks on the computer, despite having the cognitive or physical abilities to accomplish the tasks. Turner et al (2015) [10] also identified that confidence in the ability to use computers and computer anxiety impacted the use of patient portals. Turner et al (2015) [10] found that of the 59 participants who were nonusers of patient portals, 19% (11/59) had never learned how to use a computer [10].

Disparities in age and race were also mentioned. Logue and Effken [23] found that older seniors were less confident than younger seniors in their ability to use an Internet-based PHR. Older seniors (older than 78 years) were also less likely to know how to find health resources on the Internet and less interested in using PHRs [23]. Gordon and Hornbrook (2016) reported that 10.09% (260/2602) of seniors surveyed received help from someone to go on the Web or had someone go on the Web for them. They also found Chinese, non-Hispanic whites, and younger seniors (aged 65-69 years) were more likely to use the Internet for email and health-related tasks than black, Latino, and Filipino seniors and those who were aged 75 years and older [19].

Some studies also mentioned disparities based on physical and cognitive ability [19,26]. Lober et al (2006) [26] found that 13 of 38 participants had cognitive issues that impacted their use of a computer, presenting problems specifically when remembering the URL of the system, usernames, and passwords. Older adults with vision, hearing, and physical limitations leading to decreased mobility had difficulty using the system on their own [26]. Gordon and Hornbrook (2016) [19] also reported that physical issues inhibit use of a computer or the Internet. They noted that this posed more of a problem to seniors in the oldest age group (75-79 years) [19].

Facilitators

We identified two major factors that facilitated older adults' use and adoption of patient portals and ePHRs: (1) technical assistance and (2) the advice of family and providers.

Technical Assistance

Three papers mentioned the role of technical assistance in initially facilitating portal use [19,22,24]. Hourcade et al (2011) [22] described a video to help present the ePHR that they were testing among older adults. They also explained that they saw a benefit in working with older adults over several weeks, which allowed them to introduce older adults to the ePHR concept, assist with system navigation, and ultimately gather more meaningful feedback from a group that was informed about the ePHR tool [22]. In their paper, Gordon and Hornbrook (2016) [19] found that participants wanted technical assistance with using a portal and preferred help from a person rather than a Web video [19]. Kim et al (2009) [24] had graduate nursing students available to assist participants with using a patient health information management system (PHIMS) portal. They noted that the most frequent use of PHIMS coincided with the days when the nursing students were onsite [24].

Family and Provider Advice

Other papers noted family and provider advice as facilitators to portal use. Lam et al [25] found that participants were significantly more likely to be introduced to a portal messaging system by their providers than were nonusers [25]. Similarly, Zettel-Watson and Tsukerman (2016) [17] reported that patients cited their doctor's recommendation as being important when initially using the portal but not for adoption or continued use. Logue and Effken (2012) [27] found that Hispanic women, in particular, were likely to be influenced to use a PHR based on a family member's recommendation. Forty-six percent of Hispanic women stated that this was the case. They also reported that older adults who felt they were a part of a team with their health care provider were more motivated to try a PHR, to believe that an Internet-based PHR would give them their desired health outcomes, and to select a particular practice because PHRs were a part of care [27].

User Experience

The papers that were reviewed spanned a 10-year period. This is considerable, as technology tends to rapidly change over time. It is likely that experiences with newer technologies are different from older technologies. In the papers that we reviewed, we found 10 papers from 2006 to 2013 that focused on ePHRs, whereas 6 papers from 2014 to 2016 focused on patient portals.

There were several papers that evaluated participants' use of patient portals, ePHRs, or Web-based health management tools [9,10,17-21,24,26,27,29,31]. Participants reported an overall satisfaction with the system they used [25,24]. In addition to participants' satisfaction with the system, they reported that the system was useful and it improved the quality of the health care

they received [24,26]. Sack et al (2011) [29] conducted focus groups to evaluate mobile PHRs versus Web-based PHRs. They used a cost (negative comments) versus benefit (positive comments) analysis as a strategy to interpret their findings. They found that overall there were more benefit comments than cost comments for Web-based PHRs [29].

Function and Usability

Although some studies reported an ease of use in setting up and accessing their accounts, transferring information, and navigating the system [10,17], there were also some studies that raised usability issues such as difficulty in logging in and navigating a complex system [10,21]. These issues can negatively impact the user's experience and may interfere with a user's ability to complete tasks. Lam et al (2013) [25] found that some of the participants (19.0%) (31/163) who logged into a patient-physician messaging system wanted added features and functionality, 11.0% (18/163) wanted more providers in the system, and 4.3% (7/163) wanted faster response to messages [25]. Khan et al (2010) [23] found participants appreciated pictorial representations on the Colorado Care Tablet interface but had difficulty understanding what they represented. They suggested adding text to describe the pictures [23]. Two papers found that participants did not like entering text information and preferred the system to do more data input [22,23].

Features

In several papers, participants reported features of systems that they frequently used and liked. They appreciated the health information management tasks such as checking lab results, learning about health conditions [17], preparing for appointments through medication list management [23,31], and record management [17]. Participants also appreciated the ability to communicate directly with providers through secure messaging [10,19].

Six papers identified features that participants wanted from a patient portal or an ePHR system. Two mentioned that participants wanted to share health information, such as medication lists, with others or share different views of their health information depending on the person or situation [28,30]. Participants in the Sack et al (2011) [29] paper suggested that medical personnel should have a security password for record access in emergency situations.

Several papers indicated participants' desire for systems with further health management capacity and those that offered more contextual health information. Two papers reported that participants wanted the system to provide reminders for upcoming appointments, remind them when to refill medications, and help them manage their bills and health status over time [17,28]. In 3 papers, participants wanted the system to provide lifestyle advice and tips or a dictionary of medical terms [17,22,29]. Participants in 2 papers wanted the system to provide diagnosis and prognosis [28,29].

Other participants requested features specific to medication such as warnings about medication interactions and the ability to make changes to their medication lists [30,29]. Hourcade et al (2011) [22] suggested that medication information and warnings should be layered from basic to advanced information [22]. Other desired features included ability to print information, access to complete medical records, having good technical support, and ability to take voice commands [21,29].

Changes in Health Information Management and Provider Communication

Five papers described the impact of patient portals on health information management, focusing on increased access to records and improved storage of health information [10,17,20,21,24]. Zettel-Wattson and Tsukerman (2016) [17] explained that 90.6% of portal users (56/62) thought a portal helped them better manage health, and 89.7% (55/62) reported that health management tools allowed them to keep all of their records in one place. Additionally, 80.4% (50/62) explained that health information tools gave them a sense of control over their health.

In one paper, findings regarding older adult views on record access and management were mixed: 86% of participants (69/80) wanted access to their records in one place but did not necessarily want to be responsible for managing records, and 84% of participants (67/80) preferred that their records continue to be managed by primary care providers [20].

Papers also described changes in patient-provider communication. In one paper, participants expressed that having access to patient portals made them feel more prepared for emergencies and made visits with providers more efficient [24]. However, physicians thought that giving patients access to records may increase their worry [20], and some patients were concerned about a loss of face time with providers [21].

Areas to Explore

Health literacy, defined as the ability to collect, interpret, and process basic health information [32], was mentioned in 4 papers as a barrier. However, these papers measured and defined health literacy differently [9,26,27], making it difficult to categorize health literacy as a barrier in this review but highlighting it as an area for future research. Of those papers that mentioned health literacy, one defined and measured health literacy by looking at participant questions related to the content of patient portals, particular diseases, and interpreting medical terminology [26]. This paper found that health literacy was a barrier for 29% of participants (11/38) who had questions about these issues [26]. Logue and Effken (2012) [27] defined and measured health literacy using the eHealth Literacy Scale (eHEALS) and criteria that looks specifically at the ability to identify, evaluate, and synthesize health information delivered electronically. They found that all three eHealth literacy indicators from the eHEALS were positively correlated with confidence in communicating with others on the Internet, ability to express oneself in writing, and using an Internet-based PHR. Taha et al (2014) [9] measured health numeracy or the ability to interpret health information reported as numbers. They found that 52.9% of their participants (27/51) correctly answered only 5 or fewer objective numeracy questions on an 11-question measure. However, on a Subjective Numeracy Scale, which measures perceived health numeracy, several participants gave themselves a high rating, indicating that many had overestimated their health numeracy skills [9].

Design Suggestions

Several papers provided guidance about features and functions of patient portals and ePHRs [9,23,28]. At a basic level, these systems should provide health information, including medical history, test results, and medication information [28]. Information should be provided in a way that does not overwhelm the user [23,28]. Tools and aids were suggested to help users gain an understanding of health information and complete health management tasks [9]. Price et al (2013) [28] suggested that an ePHR should provide memory support to patients. For example, it should store a patient's health history and help them remember daily tasks [28]. Khan et al (2010) [23] mentioned a need for clear communication between experts, designers, and patients regarding their understanding of personal health information. This would guard against the bias of one group impacting system design [23].

Discussion

Overview

With this review, we set out to identify and assess the evidence of barriers and facilitators to the use and adoption of patient portals and ePHRs by older adults. We also wanted to gain an understanding of older adults' experiences with these systems and learn about the design recommendations resulting from study findings. Through our systematic review, we identified 2 barriers (privacy and security, and access and abilities) and 2 facilitators (technical assistance, and family and provider advice) to the use and adoption of patient portals and ePHRs. We also gained an understanding of older adults' experiences with these systems, specifically perceived benefits, satisfaction, and desired features. Some of the papers did not present specific design recommendations, making it difficult to translate findings to improve the design of patient portals and ePHRs. We also found that some papers lacked a detailed description of patient portals or ePHRs; this is an issue because systems are not static and likely changed over time. Having a detailed description of the system would provide context to study results.

Overall, even though we were able to identify barriers and facilitators, the evidence lacked strength. There were several reasons for this, including the fact that many of the studies had a small sample size and were a convenience sample. In addition, our search results included a diversity of studies, making it difficult to draw firm conclusions related to our research questions.

It should also be noted that, throughout our analysis, we reported themes by grouping papers on ePHRs and patient portals together. It could be argued that the type of technology used (ePHR vs patient portal) would influence results related to user experience, barriers, and facilitators. When it came to barriers and facilitators, we noticed no clear trends in terms of concerns about privacy and security but found that the barrier of access to the Internet was more often mentioned in papers about patient portals [10,20,21], whereas facilitators were mostly mentioned in papers that focused on ePHRs [22,24,25,27]. However, this could be because there were very few papers focused on facilitators in general, and the majority of those that mentioned facilitators were also looking at initial use [17,19,24,25,27]. In contrast, the papers on patient portals that mentioned barriers were all formative in nature [10,20,21]. This difference in paper topic (formative vs initial use) may have accounted more for the patterns in results related to barriers and facilitators than the technology itself. In terms of user experience, there were no overall trends demonstrating differences between ePHRs and patient portals. However, log-in issues were reported only from formative papers involving patient portals [10,21], and suggestions for added features (discussed in detail under experience and design) came mainly from papers involving ePHRs [22,23,25,28,29].

On the basis of our review, we identified a need for more longitudinal evaluation of patient experience and use, more nuanced understanding of older adult subgroups, and further discussion of barriers and facilitators to inform design recommendations. There were 2 papers [19,24] that looked at older adult portal use through a cohort study design, examining log-in data and uses of the portal over the course of a year [19] and almost 3 years [24]. However, other papers examined average length of use of the portals. One paper and found that several participants used the portal for longer than a year [25] and in another paper several participants used the portal for an average of 3 years [17]. Further research with longitudinal studies could help to show how use evolves into adoption and why. It could also help to better identify barriers and facilitators to adoption of patient portals or ePHRs.

Papers used different approaches to evaluate patient portals or ePHRs. Although common themes emerged across papers, the variety of approaches made drawing conclusions difficult. It would be helpful to have more research on specific and widely used systems to produce results that are comparable and generalizable.

Principal Findings

Barriers and Facilitators

Overall, it was more common for papers to describe barriers than facilitators to patient portal use. Concerns about privacy and security and lack of access or ability to use computers and technology were all commonly identified as barriers. These barriers are consistent with what has been identified in related literature. Some barriers were explained in more detail than others, and very few papers offered concrete solutions for addressing barriers, particularly among older adult populations.

Papers consistently described privacy and security issues. However, there were not many specific suggestions for making older adults feel secure, and there were no design suggestions from older adults about what would make them trust the security of a system.

Other papers more specific to privacy and security concerns found that although unauthorized access to records was an issue for older adults, it was also a concern for the general population [33]. In fact, older adults were significantly more willing than the general population to share health information with a provider [33]. Privacy and security concerns about patient portals are warranted, especially in today's climate where breaches to data are often in current news. For example, in 2016, Molina Healthcare shut down its patient portal because of a

security flaw that allowed patients to access other patient's claims without authentication [34]. In 2017, there was a breach of UC Davis Health patient health records when an employee responded to a phishing email that allowed the hacker to access the employee's emails and personal health information of patients. Fifteen thousand patients were impacted by this incident [35].

Authors of security-specific literature offered design suggestions to alleviate privacy concerns such as allowing patients to restrict access and sharing within a portal, and providing patients with an access log and list of any changes to medical information [33]. More research should be done to determine whether these and other design suggestions can work to mitigate security concerns, while still providing a positive user experience. Addressing security concerns could affect usability of a system. For example, users required to go through a 2-step log-in may perceive it as being cumbersome [36].

In patient portal research in general, there is recognition of systematic gaps in technology access and portal use [37-40]. Similar gaps in access have been identified in this literature review. Gordon and Hornbrook's (2016) [19] paper was an exemplary publication with a large sample size that identified differences in portal use and technology access within subpopulations of older adults based on age and race and asked critical questions about physical ability. However, among the papers reviewed, there was not enough evidence to understand whether there are inequities in access to technology that in turn influence older adults' portal use, skill, and quality of health care at a broader level. As noted by Kneale and Demiris (2017) [41], evaluations of patient portals often lack diversity or fail to report differences based on race, ethnicity, and gender. Generally, evaluations that report demographics conduct evaluations primarily with younger, white, non-Hispanic males who are highly educated [41]. Further evaluation of socioeconomic, racial, and gender disparities is necessary. Only a few papers drew explicit connections between access and its impact on perceived computer and Internet skills [10,26]. These papers generally did not examine the reasons behind computer anxiety or lack of confidence. Understanding and overcoming perceived barriers may be key to encouraging use and adoption of portals, but more research is necessary to identify why these perceived barriers exist.

A more in-depth discussion of facilitators, particularly among different cultural, social, and economic groups of older adults, may also be an important step toward creating a supportive system for older adults. Mention of facilitators in the literature is mainly limited to providing technical assistance [17,21,24,26]. Only Gordon and Hornbrook [19] offer suggestions for large-scale assistance programs, including user handbooks, a hotline, and workshops. Another facilitator described in the literature was provider advice. Although provider perspective was not the focus of our review, other studies suggest that provider EHR use has an impact on whether patients adopt portal technology [42,43]. Overall, additional research should focus on what facilitators are important to older adults and how these facilitators can be incorporated into the patient portal experience and implementation.

Health Literacy

Low health literacy and technology have been identified as barriers for adoption of patient portals among underserved adult populations [37,44], and privacy and technological concerns are common barriers to older adults adopting technology in general [45]. In this review, the papers varied in the way they defined and measured health literacy. One looked at health literacy by focusing on numeracy [9], another used eHEALS [27], and another measured health literacy by the number of questions that were asked about the content in the patient portal [26]. More research is needed to measure this barrier using a uniform method to identify how it affects portal use for older adults and to find design or implementation solutions that can be used to support health literacy among different subgroups of the older adult population.

Experience and Design

The papers in this review have used exploratory and evaluative methods to understand the factors that impact the use and experience of patient portals and ePHRs. However, there are opportunities to apply a design framework to developing patient portals and ePHRs. Nath and Sharp (2015) [46] proposed building on existing research methods, such as those that identify patient needs and preferences, using approaches such as user-centered design. Doing so will bridge the gap between needs and preferences and the design of a system. User-centered design is a process that aims to create usable systems that improve productivity, enhance user acceptance, reduce errors, and offer training and support. Human-centered design is based on the principle of actively involving users who have contextual knowledge of the tasks the system will be used for and the environment that the system will be used in. Human-centered design principles also include gaining an understanding of the tasks that the system will do, gaining early feedback from users through prototypes, and involving a multidisciplinary team [47].

Many of the papers reviewed identified barriers and facilitators to use and adoption. There were some that also gathered requirements for and input on system development. These findings can be used in the user-centered design process. There could be additional exploratory research done to gain an understanding of the user in context and the tasks they aim to complete. Including the user at the beginning of the process ensures that their needs are a part of the design process. Participatory design approaches have been used in this framework to engage and empower older adults in designing technology such as smart homes [48,49]. Using inclusive approaches can lead to unexpected discoveries of functions and features that are important to older adults.

Although studies in this review captured overall user experience, there is room for more exploration to better understand older adults' experience with and use of patient portals and ePHRs. Research could focus on usability by learning about participants' expectations and navigation of systems. This information could then provide designers with necessary feedback to make iterative improvements to particular systems. To understand what older adults need from patient portals and ePHRs, designers should consider including older adults in the design process.

This review looked across the user experience, examining both patient portals and ePHRs. However, these technologies do offer different experiences. The primary difference is that, as patient portals are tethered to the patient's health record, patients do not need to manually enter their information, whereas ePHRs, which are not typically tethered to a patient's health record, require patients to manually enter information. This distinction has impacted the user experience and resulted in some of the feedback about desired functionality of ePHRs that is solved by patient portals, such as limiting the amount of text entry, providing access to lab results, and the ability to contact providers. However, there were still some desired features that could be further investigated, such as reminders for appointments and medication refills, lifestyle tips, help managing claims, and voice commands. The differences between patient portals and ePHRs can perhaps also be seen as an impact of technology developing over time.

Considering that people are increasingly incorporating technology into their daily lives, desired features that provide contextual advice are a reasonable expectation. However, further research with older adults is needed to understand how patient portals or ePHRs could be integrated into older adults' health management. In addition, researchers should consider relating their findings to the design of patient portals and ePHR systems. The recommendations could provide actionable changes and lead to opportunities to explore for potential features and functionality of the systems. For example, one desired feature mentioned in a paper was voice activation; patient portals could be paired with an intelligent personal assistant, such as the Amazon Echo, to increase convenience and access to health information.

Another consideration is for researchers and designers to think about the long-term adoption of these systems. Friedman and Nathan [49] proposed an approach called multi-lifespan information system design to challenge the short life cycle of a technology, which is usually 5 years. It asks researchers to think about the future of the technology, including its impact and how its use might change over time [49]. The method may be fitting for the design of patient portals and ePHRs because they are systems available for a wide range of people and may be used over lifetimes and generations.

Limitations

The search terms for this systematic review were carefully chosen and aimed to draw a wide search. However, patient portals and ePHRs can be described differently, and some papers may have been missed. Our wide search also resulted in a diverse set of papers that presented challenges to drawing specific conclusions related to our research questions. Due to our focus on older adults, we eliminated papers that focused on provider perspectives as well as papers that focused on the health implications of patient portal implementation. We also excluded papers that were not in English, and so, we may have missed papers that were pertinent to our topic but in a different language. In addition, our key themes were determined based on a small number of papers. Even though our review included papers that analyzed patient portal and ePHR use among age groups other than older adults, we did not do a comparison between older adults and those other age groups. In addition, because of the large range of ages, 60 years and older, we did not distinguish the impact of age on the exposure to technology. Finally, our search criteria spanned over a 10-year period; it is important to recognize the constantly changing technology environment and the advances that have been made to patient portals and ePHRs over the 10-year span of time. These advances likely impacted the use and experience of participants across the studies that were reviewed.

Conclusions

This review focused on understanding the barriers and facilitators to older adults' use and adoption of patient portals and ePHRs. Across the studies there were 2 main barriers: (1) concerns about privacy and security and (2) access and ability to use technology and the Internet. The 2 main facilitators were receiving technical assistance with a patient portal or ePHR and receiving advice to use patient portals from family and providers.

In terms of older adults' experience using patient portals and ePHRs, some papers indicated that patient portals and ePHRs helped older adults to better manage their health information. Older adults liked having a single place that they could access and archive their information. In some cases, older adults felt their communication with providers had improved because of their use of patient portals. Older adults also suggested improving patient portals and ePHRs to help them manage their health beyond record storage, for example, by providing diagnosis and prognosis.

Overall, this review demonstrated that there are a range of studies and methods to understand patient portal and ePHR use and experience among older adults. However, more research is needed to better understand and address barriers to patient portal and ePHR use and adoption by older adults. As many health care systems offer their patients a portal to their health information, there are opportunities for it to be an integral part in keeping patients informed about their health information and encouraging them to take an active role in their health care. This opportunity is especially great for the older adult population as it is expected to grow rapidly. In addition, evaluation of patient portal and ePHR systems should be continually done after they are launched to learn about the areas that are working and areas that could be improved. This is in line with the user-centered design process and communicates to users the organization's commitment to deliver a positive user experience. Finally, the changing technology landscape should be considered in the design process to design a system that is flexible and would ease future transitions from legacy systems.

Acknowledgments

This review was supported by the UW SOARING project (University of Washington Studying Older Adults and Researching their Information Needs and Goals), which is funded by the Agency for Healthcare Research and Quality (AHRQ) grant #R01 HS022106. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the AHRQ. This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1256082. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would also like to thank Julie Kientz, PhD and Jean O Taylor, PhD for their detailed review of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Final set of 17 papers.

[PDF File (Adobe PDF File), 113KB - medinform_v5i4e38_app1.pdf]

Multimedia Appendix 2

Paper summaries.

[PDF File (Adobe PDF File), 269KB - medinform_v5i4e38_app2.pdf]

Multimedia Appendix 3

Outcome of quality review.

[PDF File (Adobe PDF File), 17KB - medinform_v5i4e38_app3.pdf]

References

1. Administration on Aging, Administration for Community Living, U.S. Department of Health and Human Services. ACL. A profile of older Americans URL: https://www.acl.gov/sites/default/files/Aging%20and%20Disability%20in%20America/2015-Profile.pdf [accessed 2017-05-05] [WebCite Cache ID 6qGvnPtup]
2. Levine DM, Lipsitz SR, Linder JA. Trends in seniors' use of digital health technology in the United States, 2011-2014. JAMA 2016 Aug 2;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]
3. Pew Research Center. Tech adoption climbs among older adults URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2017/05/16170850/PI_2017.05.17_Older-Americans-Tech_FINAL.pdf [accessed 2017-08-01] [WebCite Cache ID 6sQZs1byp]
4. Zhang Y, Fleischmann KR, Gao J, Xie B. A systematic review of the literature on consumers' use of patient portals: preliminary results. 2016 Feb 24 Presented at: Association for Information Science and Technology (ASIS&T); November 6-10, 2015; St. Louis, Missouri p. 1-4. [doi: 10.1002/pra2.2015.1450520100142]
5. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
6. Ancker JS, Hafeez B, Kaushal R. Socioeconomic disparities in adoption of personal health records over time. Am J Manag Care 2016 Aug;22(8):539-540 [FREE Full text] [Medline: 27541700]
7. U.S. Department of Health and Human Services. HHS. Washington, DC; 2010 Dec. Multiple chronic conditions- a strategic framework: optimum health quality of life for individuals with multiple chronic conditions URL: https://www.hhs.gov/sites/default/files/ash/initiatives/mcc/mcc_framework.pdf [WebCite Cache ID 6qGzWtrur]
8. Federal Interagency Forum on Aging Related Statistics. Agingstats. 2016. Older Americans: key indicators of well-being URL: https://agingstats.gov/docs/LatestReport/Older-Americans-2016-Key-Indicators-of-WellBeing.pdf [WebCite Cache ID 6qGzpQn2I]
9. Taha J, Sharit J, Czaja SJ. The impact of numeracy ability and technology skills on older adults' performance of health management tasks using a patient portal. J Appl Gerontol 2014 Jun;33(4):416-436. [doi: 10.1177/0733464812447283] [Medline: 24781964]
10. Turner AM, Osterhage K, Hartzler A, Joe J, Lin L, Kanagat N, et al. Use of patient portals for personal health information management: the older adult perspective. AMIA Annu Symp Proc 2015;2015:1234-1241 [FREE Full text] [Medline: 26958263]
11. US Department of Veterans Affairs. MyHealtheVet. MyHealtheVet URL: https://www.myhealth.va.gov/ [accessed 2017-05-08] [WebCite Cache ID 6qKAKMFDA]

12. Epic. 2016. Patient engagement URL: https://www.epic.com/software [accessed 2017-05-07] [WebCite Cache ID 6qHokdSYo]
13. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: 19622511]
14. Covidence. Covidence. 2017. URL: https://www.covidence.org/ [WebCite Cache ID 6qHpiMYql]
15. de Keizer NF, Talmon J, Ammenwerth E, Brender J, Nykanen P, Rigby M. Mini Stare-HI: guidelines for reporting health informatics evaluations in conference papers. Stud Health Technol Inform 2010;160(Pt 2):1206-1210. [Medline: 20841875]
16. Talmon J, Ammenwerth E, Brender J, de KN, Nykänen P, Rigby M. STARE-HI--Statement on reporting of evaluation studies in Health Informatics. Int J Med Inform 2009 Jan;78(1):1-9. [doi: 10.1016/j.ijmedinf.2008.09.002] [Medline: 18930696]
17. Zettel-Watson L, Tsukerman D. Adoption of online health management tools among healthy older adults: an exploratory study. Health Informatics J 2016 Jun;22(2):171-183. [doi: 10.1177/1460458214544047] [Medline: 25149210]
18. Barron J, Bedra M, Wood J, Finkelstein J. Exploring three perspectives on feasibility of a patient portal for older adults. Stud Health Technol Inform 2014;202:181-184. [Medline: 25000046]
19. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
20. Kerai P, Wood P, Martin M. A pilot study on the views of elderly regional Australians of personally controlled electronic health records. Int J Med Inform 2014 Mar;83(3):201-209. [doi: 10.1016/j.ijmedinf.2013.12.001] [Medline: 24382474]
21. Latulipe C, Gatto A, Nguyen HT, Miller DP, Quandt SA, Bertoni AG, et al. Design considerations for patient portal adoption by low-income, older adults. Proc SIGCHI Conf Hum Factor Comput Syst 2015 Apr;2015:3859-3868 [FREE Full text] [doi: 10.1145/2702123.2702392] [Medline: 27077140]
22. Hourcade JP, Chrischilles EA, Gryzlak BM, Hanson BM, Dunbar DE, Eichmann DA, et al. Design lessons for older adult personal health records software from older adults. In: Stephanidis C, editor. Universal Access in Human-Computer Interaction. Users Diversity. Berlin, Heidelberg: Springer; 2011.
23. Khan DU, Siek KA, Meyers J, Haverhals LM, Cali S, Ross SE. Designing a personal health application for older adults to manage medications. In: Proceedings of the 1st ACM International Health Informatics Symposium. New York: ACM; 2010 Presented at: IHI'10; November 11-12, 2010; Arlington, Virginia p. 849-858 URL: https://doi.org/10.1145/1882992.1883124
24. Kim E, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
25. Lam R, Lin VS, Senelick WS, Tran H, Moore AA, Koretz B. Older adult consumers' attitudes and preferences on electronic patient-physician messaging. Am J Manag Care 2013 Nov;19(10 Spec No):eSP7-eS11 [FREE Full text] [Medline: 24511886]
26. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, Kim EH, et al. Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc 2006:514-518 [FREE Full text] [Medline: 17238394]
27. Logue MD, Effken JA. An exploratory study of the personal health records adoption model in the older adult with chronic illness. Inform Prim Care 2012;20(3):151-169 [FREE Full text] [Medline: 23710840]
28. Price MM, Pak R, Müller H, Stronge A. Older adults' perceptions of usefulness of personal health records. Univ Access Inf Soc 2012 Mar 8;12(2):191-204. [doi: 10.1007/s10209-012-0275-y]
29. Sack O, Pak R, Ziefle M. Older adults' perception of costs and benefits of web-based and mobile PHR technologies: a focus group approach. In: Holzinger A, Simonic KM, editors. Information Quality in e-Health. Berlin, Heidelberg: Springer; 2011:707-710.
30. Haverhals LM, Lee CA, Siek KA, Darr CA, Linnebur SA, Ruscin JM, et al. Older adults with multi-morbidity: medication management processes and design implications for personal health applications. J Med Internet Res 2011;13(2):e44 [FREE Full text] [doi: 10.2196/jmir.1813] [Medline: 21715286]
31. Montelius E, Astrand B, Hovstadius B, Petersson G. Individuals appreciate having their medication record on the web: a survey of attitudes to a national pharmacy register. J Med Internet Res 2008 Nov 11;10(4):e35 [FREE Full text] [doi: 10.2196/jmir.1022] [Medline: 19000978]
32. Centers for Disease Control Prevention. CDC. 2016. What is health literacy? URL: https://www.cdc.gov/healthliteracy/learn/ [WebCite Cache ID 6qI0ZoSGs]
33. Le T, Thompson H, Demiris G. An examination of electronic health information privacy in older adults. Stud Health Technol Inform 2013;192:709-713. [Medline: 23920649]
34. Davis J. Healthcareitnews. 2017. Molina healthcare breached, exposed patient data for over a month URL: http://www.healthcareitnews.com/news/molina-healthcare-breached-exposed-patient-data-over-month [accessed 2017-08-03] [WebCite Cache ID 6sRxXaFdY]
35. Davis J. Healthcareitnews. 2017. Phishing attack on UC Davis health breaches data on 15,000 patients URL: http://www.healthcareitnews.com/news/phishing-attack-uc-davis-health-breaches-data-15000-patients [accessed 2017-08-03] [WebCite Cache ID 6sRxk0ioC]

36. Tjora A, Tran T, Faxvaag A. Privacy vs usability: a qualitative exploration of patients' experiences with secure internet communication with their general practitioner. J Med Internet Res 2005;7(2):e15 [FREE Full text] [doi: 10.2196/jmir.7.2.e15] [Medline: 15998606]
37. Czaja SJ, Zarcadoolas C, Vaughon WL, Lee CC, Rockoff ML, Levy J. The usability of electronic personal health record systems for an underserved adult population. Hum Factors 2015 May;57(3):491-506. [doi: 10.1177/0018720814549238] [Medline: 25875437]
38. Weppner WG, Ralston JD, Koepsell TD, Grothaus LC, Reid RJ, Jordan L, et al. Use of a shared medical record with secure messaging by older patients with diabetes. Diabetes Care 2010 Nov;33(11):2314-2319 [FREE Full text] [doi: 10.2337/dc10-1124] [Medline: 20739686]
39. Wallace LS, Angier H, Huguet N, Gaudino JA, Krist A, Dearing M, et al. Patterns of electronic portal use among vulnerable patients in a nationwide practice-based research network: from the OCHIN practice-based research network (PBRN). J Am Board Fam Med 2016 Oct;29(5):592-603 [FREE Full text] [doi: 10.3122/jabfm.2016.05.160046] [Medline: 27613792]
40. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. The digital divide and patient portals: internet access explained differences in patient portal use for secure messaging by age, race, and income. Med Care 2016 Aug;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
41. Kneale L, Demiris G. Lack of diversity in personal health record evaluations with older adult participants: a systematic review of literature. J Innov Health Inform 2017 Jan 15;23(4):789-798 [FREE Full text] [Medline: 28346131]
42. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
43. Irani JS, Middleton JL, Marfatia R, Omana ET, D'Amico F. The use of electronic health records in the exam room and patient satisfaction: a systematic review. J Am Board Fam Med 2009;22(5):553-562 [FREE Full text] [doi: 10.3122/jabfm.2009.05.080259] [Medline: 19734402]
44. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
45. Young R, Willis E, Cameron G, Geana M. "Willing but unwilling": attitudinal barriers to adoption of home-based health information technology among older adults. Health Informatics J 2014 Jun;20(2):127-135. [doi: 10.1177/1460458213486906] [Medline: 24056750]
46. Nath PA, Sharp CD. A user-centered design approach to information sharing for older patients and their families. JAMA Intern Med 2015 Sep;175(9):1498-1499. [doi: 10.1001/jamainternmed.2015.2907] [Medline: 26147128]
47. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
48. Demiris G, Oliver DP, Dickey G, Skubic M, Rantz M. Findings from a participatory evaluation of a smart home application for older adults. Technol Health Care 2008;16(2):111-118. [Medline: 18487857]
49. Friedman B, Nathan L. Multi-lifespan information system design: a research initiative for the HCI community. Presented at: CHI '10; April 10-15, 2010; Atlanta, Georgia. [doi: 10.1145/1753326.1753665]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
eHEALS: eHealth Literacy Scale
EHR: electronic health record
ePHR: electronic personal health record
PHIMS: patient health information management system
PHR: personal health record
PRISMA: preferred reporting items for systematic reviews and meta-analyses
STARE-HI: Statement on the Reporting of Evaluation studies in Health Informatics

Edited by L Buis; submitted 24.05.17; peer-reviewed by B Crotty, M Whetstone, J Jones; comments to author 19.06.17; revised version received 13.08.17; accepted 30.08.17; published 16.10.17.

Please cite as:
Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM
Patient Portal Use and Experience Among Older Adults: Systematic Review
JMIR Med Inform 2017;5(4):e38
URL: http://medinform.jmir.org/2017/4/e38/
doi:10.2196/medinform.8092
PMID:29038093

Original Paper

Predicting Consumer Effort in Finding and Paying for Health Care: Expert Interviews and Claims Data Analysis

Sandra Long[1], PhD; Karen A Monsen[1], PhD, RN, FAAN; David Pieczkiewicz[1], PhD; Julian Wolfson[1], PhD; Saif Khairat[2], PhD

[1]University of Minnesota, Minneapolis, MN, United States

[2]University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Saif Khairat, PhD
University of North Carolina-Chapel Hill
428 Carrington Hall
CB #7460
Chapel Hill, NC, 27514
United States
Phone: 1 9198435413
Email: saif@unc.edu

Abstract

Background: For consumers to accept and use a health care information system, it must be easy to use, and the consumer must perceive it as being free from effort. Finding health care providers and paying for care are tasks that must be done to access treatment. These tasks require effort on the part of the consumer and can be frustrating when the goal of the consumer is primarily to receive treatments for better health.

Objective: The aim of this study was to determine the factors that result in consumer effort when finding accessible health care. Having an understanding of these factors will help define requirements when designing health information systems.

Methods: A panel of 12 subject matter experts was consulted and the data from 60 million medical claims were used to determine the factors contributing to effort.

Results: Approximately 60 million claims were processed by the health care insurance organization in a 12-month duration with the population defined. Over 292 million diagnoses from claims were used to validate the panel input. The results of the study showed that the number of people in the consumer's household, number of visits to providers outside the consumer's insurance network, number of adjusted and denied medical claims, and number of consumer inquiries are a proxy for the level of effort in finding and paying for care. The effort level, so measured and weighted per expert panel recommendations, differed by diagnosis.

Conclusions: This study provides an understanding of how consumers must put forth effort when engaging with a health care system to access care. For higher satisfaction and acceptance results, health care payers ideally will design and develop systems that facilitate an understanding of how to avoid denied claims, educate on the payment of claims to avoid adjustments, and quickly find providers of affordable care.

(JMIR Med Inform 2017;5(4):e39) doi:10.2196/medinform.7892

KEYWORDS

consumer health information; user effort; patient acceptance of health care; health expenditures; health services accessibility

Introduction

Background

Technology is used by over 70% of the US population to seek health information [1,2]. For consumers to successfully improve health outcomes, they must be engaged and satisfied that the information they receive is accurate in meeting their needs [3,4]. This includes engagement in finding health care providers, deciding on appropriate treatments, and paying for care [3,5,6]. Therefore, to be fully engaged, the consumer must satisfactorily accept the design of the system or process they follow to access information [7].

When reviewing the literature related to the acceptance and engagement of health information systems, some of the most frequently occurring barriers include the failure of the system to meet consumer needs [3,8]. As the solutions created using proven methodologies are not fully meeting the needs of the consumer, it is likely that not all requirements were correctly identified [9]. The literature reviewed did not address the relationship of finding affordable health care to engagement with health care systems or how the monetary cost of to the consumer can have an impact on acceptance. By engaging with a well-designed health care system early on and planning their care, consumers can prevent the need to resolve ongoing payment or access issues that arise.

Usability acceptance and design methodologies have been created to make sure information delivery systems meet user requirements [7]. Some of these include human-centered design, Agile, Design for Six Sigma, and technology acceptance model [7,10-12]. All of these involve steps where the consumer requirements are documented and solutions are then created to meet them. These requirements ideally include aspects that make sure the consumer accepts using the system. There are two primary aspects that can be considered to lead to acceptance. The first is how much the consumer perceives the system to be useful or "the degree to which a person believes that using a particular system will enhance his or her performance or outcome" [13]. This means that in health care, the consumer trusts that the information they are receiving will lead to better health by using the system. The other is the perceived ease of use, or "the degree to which a person believes that using a particular system would be free from effort" [13]. This means that the least amount of effort required for the consumer to use the system leads to the most accepted design [7,13]. In this paper, effort is defined as the work done on the part of the consumer to find and pay for health care services. Health insurance payers and providers ideally consider this level of effort when developing information systems.

Purpose and Aims

The purpose of this study was to explore how to design an accepted information system that assists consumers in accessing health care based on their diagnosis and the ability to easily find care. It focused on the ease-of-use or lack-of-effort aspect of acceptance. We aimed to define the types of users who put forth the most effort in accessing health care, resulting in improved consumer requirements for designing the health care system. The health care system of focus to this study is a call center that individuals, referred to as consumers, use to find a provider, understand payment for procedures, and get assistance with treatment decision support. It provides service to approximately 8 million consumers and receives 350,000 contacts from consumers per month. The consumer's health insurance provider gives the consumer a telephone number and secure electronic mail address that can be used to contact the call center. The system was created with a primary focus of operational efficiency and to quickly answer only the question the consumer is asking or to transfer the call to someone else who can answer. The person who answers the questions used desktop systems and databases to answer the questions. There are times when the consumer is not aware of additional aspects of health care access, so they may not ask all questions related to how finding care and payment may work. This can be frustrating when they must talk to several people, are later surprised to find out that a provider is not accepting patients, or that their care was not covered by their insurance plan after they received treatment. Few of those using the system actually utilize the treatment decision support and clinical aspects because they are frustrated with the amount of effort required with payment and access.

The treatment and amount of care required depends on a diagnosis. We hypothesized that consumers with certain diagnoses utilize the system more often for administrative issue resolution, regardless of the clinical complexity of the diagnosis. Therefore, in this study, we explored how diagnosis contributes to a consumer's understanding of how to access treatment.

Methods

Procedure

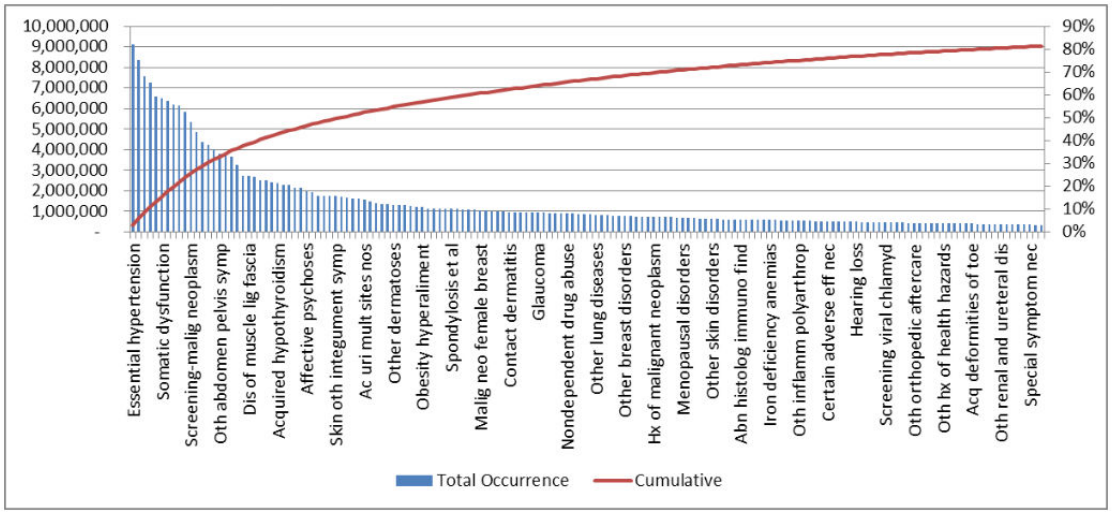
This study utilized expert knowledge and descriptive statistics to understand which variables predict effort and how they relate to diagnosis. A panel of experts within the organization was consulted to define which factors show whether a consumer is putting forth effort when accessing care. The panel consisted of 12 subject matter experts; 3 administrative call agents, 3 registered nurses who work directly with consumers to find care and assist with treatment decision support, 3 medical claims adjustors, and 3 data analysts familiar with medical claim and contact center data. They were recommended as being experts by other employees within the organization because of their health care education and experience credentials and were recruited through a conversation to determine their availability in assisting with the study. These experts had over 70 years of total experience working in various health care organizations throughout the United States, such as large hospital systems, health payer organizations, and clinical data analytic firms. Interviews were conducted individually with each panel member and consisted of two primary questions. The panel was first asked to define activities that contribute to a consumer's effort in paying for health care. Second, the panel was asked to state what defines effort when trying to find providers and treatments. When the majority of the panel's qualitative responses identified the same type of activity, that activity was determined to be an important factor. Validation of these factors also occurred through qualitative comparison with freeform responses from over 1000 consumers surveyed after their use of the system. Once the panel defined the factors, they were instructed to consider how the factors compare with each other and weigh the importance of the factors, giving higher weightage to those causing more effort.

Given that it is hypothesized that these factors occur more frequently for certain diagnoses, data needed to be analyzed to assign a value for each factor to each diagnosis. By giving a numerical value for each factor, it can be understood which diagnoses require more effort than others. The data related to payment and accessing care can be found in historical medical claims and records of consumer interactions with the health care organization.

Data Analysis

In this study, 12 months of health insurance claims data were used for analysis. It consisted of all data for the year of 2014, and all analysis was completed within the health care organization's secure system to maintain privacy of personal information. The data included the name and address of providers, International Classification of Diseases (ICD) codes, payment processing information, and consumer identification information that could be matched to their demographics and utilization of the system being studied. To narrow the scope, the top 79.89% (233M/292M) of occurring diagnoses were used. This allowed for the majority of consumers to be considered in the study, but it did eliminate those consumers with only rare disease diagnoses. This was a negligible number of consumers as most with a rare diagnosis have a comorbidity diagnosis that was common. The justification for using 1 years' worth of data was because this is the standard period for consumers to have the same health insurance plan. They typically select a new plan on a yearly basis, and even if they remain with the same one, the payer (government, employer, or administrator) often changes the benefits and coverage offered each year [14]. Provider association with the plan also changes. These yearly changes mean that the consumer must seek how to access and pay for care on a yearly basis, even if they are undergoing treatments similar to the prior year [14]. Although it is true that the treatment needed by a diagnosis may span more than 1 calendar year, the large quantity of data available eliminated concerns related to consumers being diagnosed at different points through the year. When grouping the diagnoses by ICD code, 117 codes made up 80% of all diagnoses occurring. Figure 1 shows the diagnoses used in the study and the volume of their overall contribution to the population. The mean and median were then calculated for the number of times a consumer experienced each of the effort factors defined by the expert panel.

Figure 1. Diagnoses sorted by occurrence where left axis is the number of diagnoses occurring in the dataset and right axis is the percentage of the dataset made up by the diagnoses.



Results

Approximately 60 million claims were processed by the health care insurance organization in a 12-month duration with the population defined. About 292 million diagnoses appeared on the claims. When looking at the effort required to pay for care, the expert panel agreed that the primary factors are the number of denied claims, number of claims adjustments (when a claim was processed incorrectly the first time and then needed to be reworked), and when a consumer visits a medical provider who is out of network for their insurance coverage. Their reasoning for this was that a denied claim means that the burden of payment is then placed on the consumer who may have expected their insurance plan to cover the full or part of the cost. It is usually the consumer who must notice and initiate action for correction when the medical claims are processed incorrectly and need adjustment. Adjustments result in delayed payments as the medical claims are being reworked, causing the provider to bill the consumer until payment is received. Visits to network providers who are out of their insurance coverage result in high cost to consumers because the costs of care are not negotiated between the provider and the payer.

When the expert panel was asked what defines effort as far as accessing care, frequency of phone calls to the health care organization, number of Web portal visits, and number of mobile app usages were determined to be the main factors. These instances show that a consumer was not sure on how to access care, so they needed to contact the health care organization with inquiries. The types of inquiries from consumers consist of finding providers, determining the best treatments, and understanding the different types of care available within their health insurance plans. Other questions related to how to use their health insurance plan to pay for care also comes in through these communication channels, which ties to the effort required in covering the cost of treatment.

During the questioning, the experts also repeatedly brought up that the amount of effort often depends on the number of people with health care claims in a family or household. Their reasoning for this was that multiple members in the household may need care, and the majority of the administrative burden falls to a single person acting as the caretaker. Due to this, the number of people needing care in a household was also included as a factor. This also helped to define a consumer as all the members

in a household or all the members needing care on a single health insurance plan subscribed to by the caretaker. The factors contributing to effort and weighting of importance assigned by the expert panel can be found in Table 1. The total effort put forth is considered to be 100%. The weightings were assigned based on how much each factor contributes to the total effort put forth by the consumer. Therefore, if half of the total effort came from a given factor, the weighting assigned would be 50%. The information provided by the panel aligned to the qualitative feedback consumers gave in surveys after using the system. The consumers were often frustrated when they received a bill from a provider for an unexpected amount of money and when they needed to contact the organization many times for resolution.

Table 1. Factors determined to represent consumer effort in accessing care and weighting of importance; summing the contribution of the factors equals total effort.

Factor of consumer effort	Weight of contribution to total (%)
Number of calls made by consumer	40
Number of claims where payment was denied by payer	25
Number of Web or mobile app visits	15
Visits to providers outside health insurance network	8
Number of adjustments required on claims	7
Number of people with claims in a household	5

Using existing data in the organization, the average number of times a consumer experienced each of the factors was determined. Therefore, each diagnosis had a value assigned for the average number of denied claims per consumer, average out-of-network usage per consumer, average number of claim adjustments per consumer, and average inquiries into the health care organization per consumer (phone, Web, or mobile app). The average number of people in a household was also calculated per consumer diagnosis.

The average of each effort factor was compared with the diagnosis. When the diagnoses were sorted based on the average of each individual factor, the list of those showing up at the top varied. For example, chronic kidney disease was the diagnosis with the second highest out-of-network provider visits per consumer but fell about halfway down the list when looking at how frequently the average consumer calls the health care organization. Correlations between factors were calculated to see how finding care and paying for it may relate. Although some variables are moderately correlated, each one is independent as it relates to the type of effort required. They are clinically significant as one action may lead to another. For instance, a consumer may have a denied claim, which may then qualify for an adjustment. Thus, they call or visit the portal to resolve the issue. The number of people in a household moderately correlates with all factors, with the exception of claim adjustments. An explanation for why there are not high correlations may be that one household member may have a very complex circumstance versus multiple household members having situations that are more easily resolved. The expert panel was consulted again to understand whether there is causation between variables. They concluded that although a person with a certain diagnosis may have trouble finding and paying for care, the effort required for one factor does not mean effort will be required for another factor. Given it is known that some factors require more effort than others based on the expert panel, the weightings defined by them were used to assign a score of effort to each diagnosis. This helped to define which diagnosis requires more effort than others.

The weightings were used similar to coefficients in a mathematical equation. The weighting of each factor was multiplied by the average number of occurrences for the factor. The results were then summed together to provide a total effort risk score for each diagnosis. See Table 2 for example of calculation.

Table 2. Calculation of effort risk score for dislocation of the knee.

Variable	Weight (%)	Average	Score
Number of people with claims in a household	5	2.14	0.11
Visits to providers outside health insurance network	8	4.48	0.36
Number of adjustments required on claims	7	0.57	0.04
Number of claims where payment was denied by payer	25	7.11	1.77
Number of calls made by consumer	40	0.65	0.26
Number of Web or mobile app visits	15	1.72	0.26
Total effort			2.80

Once all the scores were calculated, it was shown that the diagnoses with the highest amount of effort are not necessarily the ones that are most clinically complex. For example, both sprain of the knee and leg and malignant neoplasm of the breast show up in the top 20 of the list. The typical treatment within a given year for a sprained knee typically consists of less

treatments in a year compared with breast cancer [15,16]. The final results are a list of the most frequently occurring diagnoses sorted, as they related to the effort a consumer must put forth with the health care system to find and pay for care. The list was given to the subject matter experts for validation. The top 25 diagnoses are shown in Table 3.

Table 3. Top 25 high-effort diagnoses.

Diagnosis description (ICD[a] short description)	Average number of people in a household	Average number of out-of-network visits	Average number of claim adjustments	Average number of claim denials	Average number of phone inquiries	Average number of Web or mobile app inquiries	Total effort risk
Chronic kidney disease	1.61	7.10	1.38	15.87	0.57	1.13	5.11
Drug dependence	2.18	11.62	0.97	8.61	0.95	1.79	3.91
Heart failure	1.39	3.64	0.77	12.61	0.41	0.55	3.82
Complic medical care nec/nos	1.61	0.74	0.28	11.89	0.69	1.17	3.58
Sprain of the knee and leg	2.32	5.46	0.60	8.27	0.81	1.72	3.25
Chr airway obstruct nec	1.41	2.45	0.46	10.38	0.35	0.85	3.16
Malig neo female breast	1.74	3.40	0.97	7.89	0.85	2.47	3.11
Intervertebral disc dis	1.91	5.17	0.64	8.00	0.74	1.41	3.06
Somatic dysfunction	2.18	6.17	0.52	7.69	0.56	1.74	3.05
Dis of muscle or lig or fascia	2.00	5.33	0.47	7.49	0.63	1.48	2.91
Other cervical spice dis	1.96	5.04	0.52	7.14	0.68	1.68	2.85
Periph enthesopathies	1.95	4.27	0.52	7.22	0.67	1.78	2.82
Dislocation of the knee	2.14	4.48	0.57	7.11	0.65	1.72	2.80
Osteoarthrosis etal	1.59	3.33	0.66	7.20	0.74	1.64	2.73
Cardiac dysrhythmias	1.56	2.34	0.50	8.03	0.51	1.20	2.69
Nurit or metab or devel symp	2.40	3.97	0.72	6.28	0.75	1.53	2.59
Malign neopl prostrate	1.69	2.52	0.71	7.04	0.60	1.56	2.57
Back disorder nec and nos	1.87	4.04	0.47	6.23	0.60	1.54	2.48
Radius and ulna fracture	2.51	3.41	0.61	5.98	0.71	1.56	2.45
Oth-ill def morbid or mortl	1.55	2.15	0.33	6.80	0.62	1.41	2.43
Joint disorder nec and nos	1.96	3.64	0.46	5.88	0.69	1.42	2.38
Oth dis synov or tend or bursa	1.82	3.37	0.52	5.88	0.61	1.55	2.34
Normal pregnancy	2.19	2.45	0.59	5.16	1.06	1.84	2.34
Oth chr ischemic hrt dis	1.56	1.97	0.53	6.43	0.48	1.26	2.26
Sprain of the back nec or nos	1.84	4.02	0.37	5.39	0.54	1.10	2.17

[a]ICD: International Classification of Diseases.

Discussion

Principal Findings

This study took into consideration the consumer's ability to pay and access health care with regard to their diagnosis. It was determined that having denied medical claims and inability to work with a provider covered by insurance results in higher effort for the consumer. Confusion and inquiries on how to get care also contribute to the need to engage. Removing the amount of work involved with these administrative tasks allows for easier access to the treatment procedures that are more likely to result in better health. Effort in using the health care system impacts a consumer's willingness to accept the system and engage. This amount of effort also ties to diagnosis.

Diagnosis appears to stand out as being a way to determine who puts forth effort when accessing the health care system. Segmenting the population by diagnosis as it relates to effort will allow more customization to consumer needs when designing the system. By focusing the system design to assist those who put forth the most effort when accessing and paying for care, the overall satisfaction and acceptance of the system will be improved. Given that claim denials are the greatest source of effort, the system could help educate consumers on how to avoid this situation. The cost of care can be reduced through improved utilization of the broader health care system, and the cost of the information system within the health insurance organization should also be reduced as there will be

less contacts and rework, thereby resulting in better operational efficiency.

Future studies could verify the belief that designing specifically for effort factors and diagnosis will improve the satisfaction of the consumer in using the health care system. In this study, payment and access systems were the area of focus, and health care consisted of other tasks such as treatment regimens, medication adherence, and clinical-based care. These additional tasks could also add to the level of effort; removing concerns about access and payment will only begin to make overall health care easier for consumers.

The method used in the paper can be replicated in other organizations to assist with guiding consumers toward accessible health care. Although this study included hundreds of variations of health insurance policies, call center representative expertise, provider networks, and population demographics across the United States, the results are likely dependent on the context of these factors in the organization studied. Ideally, a health care organization would take into consideration the structure and processes of their own system to determine the factors that result in effort for their consumers.

Limitations

Limitations of this study include the fact that only 1 years' worth of data were used. It is possible that over a lifetime, consumers with chronic diagnoses put forth more effort. Although this is true, consumers who do not typically use the health care system are often those most confused about how to use it [14]. Many of the highest effort diagnoses were related to injuries such as sprains and fractures, as well as normal pregnancy. Those who are athletic and mothers who are generally healthy would be those likely to experience these diagnoses and therefore would not be frequent users of the health care system.

Conclusions

This study provides an understanding of how consumers must put forth effort when engaging with a health care system to access care. It shows how their diagnosis relates to the amount of effort put forth in administrative tasks such as finding providers and paying for care versus the effort related to undergoing treatments. It is known that for consumers to accept and engage with a system, it must be free from effort and easy to use. Therefore, designing systems using results found in this study is more likely to lead to better consumer engagement. For higher satisfaction and acceptance results, health care payers ideally will design and develop systems that facilitate an understanding of how to avoid denied claims, educate on the payment of claims to avoid adjustments, and quickly find providers of affordable care. This could be done across platforms that provide information for accessing care, such as forms, Web portals, and call centers. Consumers would receive information as part of the system process instead of relying on their own knowledge as a guide for health care navigation. There is a relationship between consumers' ability to access and pay for care with their satisfaction in engagement; by first removing stress and improving satisfaction by finding financially accessible care, we can then gain consumer engagement for treatments and clinically related health and well-being.

Conflicts of Interest

None declared.

References

1. Hhs. 2016 Feb 16. U.S. Department of Health and Human Services URL: https://www.hhs.gov/ [accessed 2017-09-29] [WebCite Cache ID 6tqEtKHjz]
2. Census. United States Census Bureau URL: http://www.census.gov/ [accessed 2017-09-29] [WebCite Cache ID 6tqFH5fZl]
3. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560. [doi: 10.1197/jamia.M2888] [Medline: 19390112]
4. Moen A, Brennan PF. Health@Home: the work of health information management in the household (HIMH): implications for consumer health informatics (CHI) innovations. J Am Med Inform Assoc 2005;12(6):648-656 [FREE Full text] [doi: 10.1197/jamia.M1758]
5. Gruman J, Rovner MH, French ME, Jeffress D, Sofaer S, Shaller D, et al. From patient education to patient engagement: implications for the field of patient education. Patient Educ Couns 2010 Mar;78(3):350-356. [doi: 10.1016/j.pec.2010.02.002] [Medline: 20202780]
6. Taylor HA, Sullivan D, Mullen C, Johnson CM. Implementation of a user-centered framework in the development of a web-based health information database and call center. J Biomed Inform 2011 Oct;44(5):897-908 [FREE Full text] [doi: 10.1016/j.jbi.2011.03.001] [Medline: 21396486]
7. Bevan N. Usability. In: Encyclopedia of Database Systems. New York City, NY: Springer; 2009:3247-3251.
8. Long S, Khairat S. Roadmap for engaging consumers in using health information technology. Stud Health Technol Inform 2014;202:177-180. [Medline: 25000045]
9. Wen-Chih C, Chyuan P, Chin-Chao L. The relationship between technology acceptance model and usability test-case of performing e-learning task with PDA. 2009 Presented at: Information Engineering, 2009. ICIE'09. WASE International Conference on Information Engineering; July 10-11; Taiyuan, Chanxi, China. [doi: 10.1109/ICIE.2009.207]
10. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]

11. Stochel MG, Wawrowski MR, Waskiel JJ. Adaptive agile performance modeling and testing. 2012 Presented at: Computer Software and Applications Conference Workshops (COMPSACW), 2012 IEEE 36th Annual Computer Software and Applications Conference Workshops; July 16-20; Izmir, Turkey. [doi: 10.1109/COMPSACW.2012.85]
12. He Y, Tang X, Chang W. Technical decomposition approach of critical to quality characteristics for product design for six sigma. Qual Reliab Eng Int 2010;26(4):325-339. [doi: 10.1002/qre.1077]
13. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance in information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
14. Brügemann B, Manovskii I. Economics.sas.upenn. Fragility: a quantitative analysis of the US health insurance system URL: http://economics.sas.upenn.edu/~manovski/papers/Fragility.pdf [accessed 2017-10-06] [WebCite Cache ID 6u0aZ2sNk]
15. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, Panel members. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 2013 Sep;24(9):2206-2223 [FREE Full text] [doi: 10.1093/annonc/mdt303] [Medline: 23917950]
16. Logerstedt DS, Snyder-Mackler L, Ritter RC, Axe MJ, Godges JJ. Knee stability and movement coordination impairments: knee ligament sprain. J Orthop Sports Phys Ther 2010 Apr;40(4):A1-A37. [doi: 10.2519/jospt.2010.0303] [Medline: 20357420]

Abbreviations

ICD: International Classification of Diseases

Edited by G Eysenbach; submitted 19.04.17; peer-reviewed by S Zwier, T Copeland; comments to author 29.07.17; revised version received 17.08.17; accepted 06.09.17; published 12.10.17.

Please cite as:
Long S, Monsen KA, Pieczkiewicz D, Wolfson J, Khairat S
Predicting Consumer Effort in Finding and Paying for Health Care: Expert Interviews and Claims Data Analysis
JMIR Med Inform 2017;5(4):e39
URL: http://medinform.jmir.org/2017/4/e39/
doi:10.2196/medinform.7892
PMID:29025697

Original Paper

Patient Portal Utilization Among Ethnically Diverse Low Income Older Adults: Observational Study

Thomas A Arcury[1], PhD; Sara A Quandt[2], PhD; Joanne C Sandberg[1], PhD; David P Miller Jr[3], MS, MD; Celine Latulipe[4], PhD; Xiaoyan Leng[5], MD, PhD; Jenifer W Talton[5], MS; Kathryn P Melius[1], MS; Alden Smith[6], MBA; Alain G Bertoni[2], MPH, MD

[1]Department of Family and Community Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States
[2]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, United States
[3]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States
[4]Department of Software and Information Systems, University of North Carolina at Charlotte, Charlotte, NC, United States
[5]Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, United States
[6]Greene County Health Care, Inc, Snow Hill, NC, United States

Corresponding Author:
Thomas A Arcury, PhD
Department of Family and Community Medicine
Wake Forest School of Medicine
Medical Center Boulevard
Winston-Salem, NC, 27157
United States
Phone: 1 3367169428
Email: tarcury@wakehealth.edu

Abstract

Background: Patient portals can improve patient communication with providers, provide patients with greater health information access, and help improve patient decision making, if they are used. Because research on factors facilitating and limiting patient portal utilization has not been conceptually based, no leverage points have been indicated for improving utilization.

Objective: The primary objective for this analysis was to use a conceptual framework to determine potentially modifiable factors affecting patient portal utilization by older adults (aged 55 years and older) who receive care at clinics that serve low income and ethnically diverse communities. The secondary objective was to delineate how patient portal utilization is associated with perceived usefulness and usability.

Methods: Patients from one urban and two rural clinics serving low income patients were recruited and completed interviewer-administered questionnaires on patient portal utilization.

Results: A total of 200 ethnically diverse patients completed questionnaires, of which 41 (20.5%) patients reported utilizing portals. Education, social support, and frequent Internet utilization improve the odds of patient portal utilization; receiving health care at a rural clinic decreases the odds of portal utilization.

Conclusions: Leverage points to address disparities in patient portal utilization include providing training for older adults in patient portal utilization, involving spouses or other care partners in this training, and making information technology access available at public places in rural and urban communities.

(JMIR Med Inform 2017;5(4):e47) doi:10.2196/medinform.8026

KEYWORDS

electronic health records; electronic personal health information management; health disparities; aging; rural health

Introduction

Background

Electronically supported forms of personal health information management is essential to the future of health care as these approaches facilitate improved health outcomes through improvement in health care quality and efficacy, decrease medical costs, and improve patient-physician communication [1-10]. For the potential of electronic personal health information management approaches to improve patients' communication with their providers and for patients to access greater information that improves decision making, these patients must actually utilize electronic personal health information management applications. Patient portals are one approach to electronic personal health information management that has been discussed for its potential to benefit patients and to reduce health care costs. Patient portals enable secure messaging between patients and health care providers and give patients access to their personal health records [11-13]. Patient portals are a concern for health care providers, as the Centers for Medicare and Medicaid Services has mandated that providers achieve meaningful use of these portals by their patients. Patient portal utilization is especially important for older adults, as aging is associated with a growing number of health issues and disabilities, prescription medicines, and providers.

Research on patient portal utilization has included several approaches. First, analyses of electronic health (eHealth) records indicate a wide variation in the proportion of patients receiving the access code for their portal, activating their accounts, and actually utilizing their portals [14-17]. Although Gordon and Hornbrook [14] found that almost 80% of Kaiser Permanente older adult patients had enrolled in the health system's patient portal, most other analyses have reported lower rates of patient portal utilization, with between 10% and 30% of patients activating or logging into their portals at least once and fewer than 10% being active portal users [15-17]. In addition to more women than men utilizing their patient portals, these secondary analyses consistently found that factors reflecting health disparities, including older age, lack of private health insurance, and minority group membership were related to lower patient portal utilization.

Cross-sectional surveys report similar low levels of patient portal utilization. Fewer than one-third of patients report having logged into their patient portal accounts in the past year [18,19]. Similar to analyses of electronic health records (EHRs), these primary surveys found that measures related to health disparities, including lower educational attainment, older age, minority group membership, and living in rural communities were associated with lower patient portal utilization [16,19]. Furthermore, Peacock and colleagues [18] found that health care providers were less likely to offer patient portal access to minority patients than white patients.

Analyses of factors affecting patients' and caregivers' portal utilization have used qualitative designs based on focus groups and individual in-depth interviews [20-24]. The need for technical assistance [21] and the lack of technological experience and access to technology [20], as well as a lack of facility with keyboards and screens [22] were discussed as barriers to patient portal utilization by patients and their caregivers. Limits to literacy and health literacy (the inability to read or understand information provided through the portal) also reduced patient portal utilization [22,23], as did concerns over information security [20,22,23]. Fear of losing a personal relationship with a health care provider and a preference for in-person communication with health care providers also curbed the desire to utilize a patient portal [20,21,23,24]. At the same time, poor existing relationships and communications with a health care provider increased the desire to utilize a patient portal [24]. Health care providers have also expressed concerns about patient portals, including uncertainties over increased workload [25,26], increased patient confusion, and alienating patients who do not utilize portals [26].

Finally, analyses examining the association of patient portal utilization on health outcomes have shown mixed effects. Portal utilization increases patient satisfaction and improves health services utilization [15,27-29], but data are insufficient for determining the effects of patient portal utilization on health outcomes [30]

The overall picture of patient portal utilization is that only a limited number of patients utilize these portals, with utilization decreasing with patient age. Several personal characteristics that reflect health disparities limit patient portal utilization. However, because much of this research is not conceptually based, no leverage points are indicated for improving patient portal utilization. Our primary aim for this analysis was to use a conceptual framework to determine potentially modifiable factors affecting patient portal utilization by older adults (those aged 55 years and older) who receive care at clinics that primarily serve low income, ethnically diverse communities. Our secondary aim was to delineate how patient portal utilization is associated with perceived usefulness and usability. This analysis is innovative in that it focuses on older adults who receive care from clinics concentrating on patients with limited resources, and it is conceptually based.

Conceptual Framework

Our conceptual framework integrates concepts from Davis' Technology Acceptance Model (TAM) [31] and the Person-Environmental Interaction Model [32] that have been influential for understanding users' adoption of technologies [33-37]. On the basis of Fishbein and Azjen's Theory or Reasoned Action and Theory of Planned Behavior [38], TAM posits that for acceptance of information technology (IT), belief leads to attitude leads to intention leads to behavior. Empirical results have demonstrated a parsimonious model containing two beliefs as the fundamental determinants of technology product utilization: perceived usefulness and perceived usability [35,39-41]. Perceived usefulness is the degree to which an individual believes that using a technology would enhance performance, whereas perceived usability is the degree to which an individual believes that using a technology would involve little effort [31].

Our framework addresses how key factors of (1) the individual user, (2) social support, (3) organizational characteristics, (4) environment, and (5) human-technology interaction influence

the overall adoption process (Figure 1). Individual user characteristics such as education and health literacy are important for understanding patient portal utilization. Lower technology access and use are related to infrastructure barriers as well as demographic characteristics, including ethnicity (African American and Latino), advanced age, and lower income and education [42,43]. Older adults' utilization of patient portals is influenced by their unique characteristics. For instance, age has been inversely related to technology interest and utilization [44,45], Internet utilization [46,47], and broadband access [42,43].

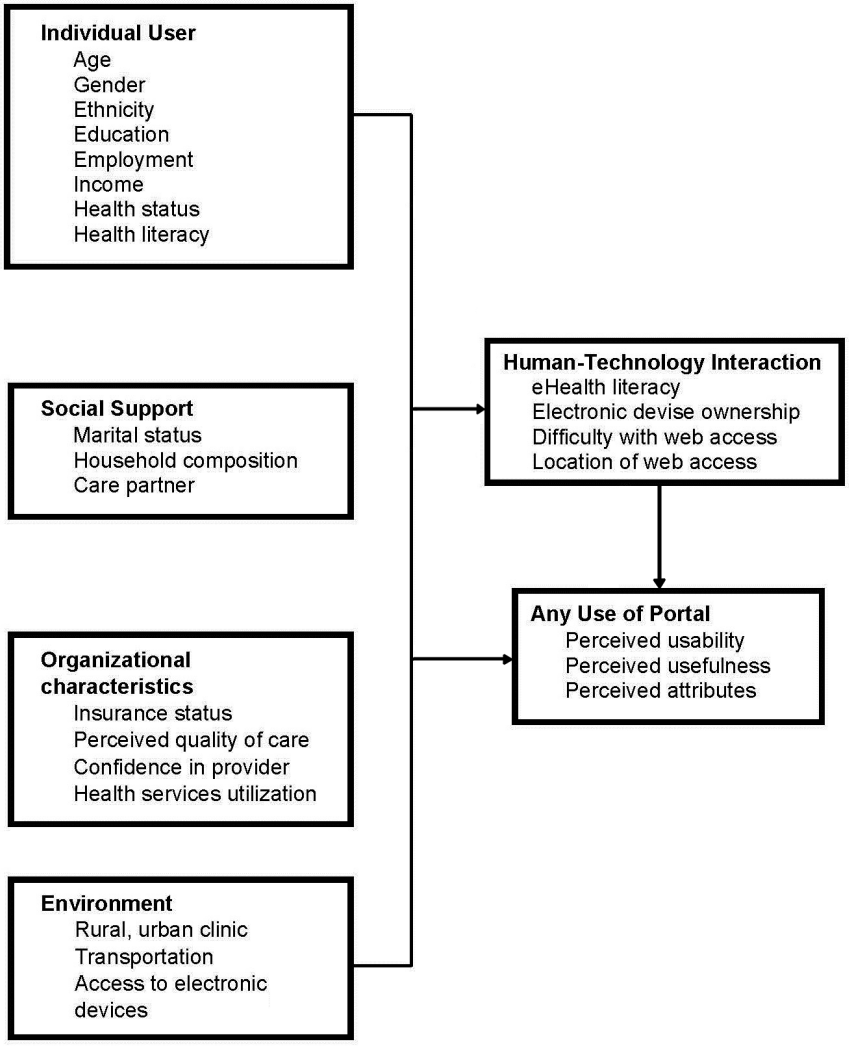
Successful patient portal adoption depends on two other key factors: social support and organizational characteristics. Older adults' social network can be segmented into components (eg, spouse, children, and care partner), each having its own modalities to initiate and sustain patient portal utilization. Social support has been found to be predictive of health information technology utilization, although the findings are not consistent [37,48]. Technology utilization can also improve social support for older adults [49]. The health and communication information field still lacks a clear understanding between older users and the members of their social network to support IT behavior.

Some research suggests the importance of organizational characteristics in using health technologies. Satisfaction with medical care services and confidence in one's health care provider are associated with technology acceptance [37,50,51]. Our rationale is that patient portal adoption will likely be embraced by satisfied patients. The extent to which older patients utilize patient portals likely depends on the frequency of care and having more confidence in one's health care provider.

Users' environmental contexts facilitate or impede patient portal adoption. Technology and contextual setting do not occupy separate domains but are intimately linked. Many individuals without Internet access make use of public resources such as libraries or community centers [52]; thus, public access sites serve to improve usage, contributing positively to patient portal acceptance among rural users.

The framework focuses on the human-technology interaction that is based on TAM. Human-technology interaction variables such as eHealth literacy and technology experience moderate the degree to which older adults utilize technology and their perception of its usability and usefulness [53].

Figure 1. Conceptual framework.



Methods

Recruitment

Participants were recruited from urban and rural clinics that primarily serve low income patients. The urban clinic was the Outpatient Department Clinic, Department of Internal Medicine, Wake Forest Baptist Health, located in Winston-Salem, North Carolina. The outpatient department services ethnically diverse, low income, predominantly Medicare and Medicaid patients. Two members of Community Partners HealthNet were the rural clinics. Community Partners HealthNet is a health center–controlled network that was formed in 1999 to implement practice management systems for community health centers. Greene County Health Care Inc and West Caldwell Health Council Inc serve the rural areas of Greene and Caldwell Counties, North Carolina, respectively. Greene County Health Care Inc has six clinic locations. West Caldwell Health Council Inc has two clinic locations. The patient portal systems of the urban and rural clinics differed; the urban system had been established for several years and included a large number of features, whereas the system used in the rural clinics was new and had fewer features than those used by the urban clinic. Both systems were only available in English, making them difficult to use for patients from North Carolina's growing Latino population, many of whom have limited English language skills.

Inclusion criteria were community-dwelling adults aged 55 years and older, who were being treated for a chronic disease (diabetes, hypertension, dyslipidemia, or cardiovascular disease), who spoke English or Spanish, and were in sufficiently good health to give informed consent and complete the series of interviews. The majority of patients were recruited using a three-step process. With the assistance of clinic staff, lists of patients who met the inclusion criteria were generated and shared with the project team members. Clinic staff reviewed these lists and indicated patients who might be willing to participate in the project. Potential participants were randomly selected from this list and sent letters introducing and describing the study. The letters indicated that patients were eligible to participate because they met the inclusion criteria. Follow-up phone calls were then made to further describe the study and to schedule interviews with those who received the letters. Additionally, Spanish-speaking participants were recruited as they came to one set of rural clinics. The data collector approached individuals fitting the inclusion criteria, described the study, and scheduled interviews for a later date. The study protocol was approved by the Wake Forest Baptist Health Institutional Review Board, and all participants provided signed informed consent.

The patient sample included 200 African American, white, Latino, and other older adult patients who completed baseline interviews (Table 1). Data collectors attempted to contact 628 patients by letter or in person, with a follow-up telephone call (Figure 2). Of the 628 attempted contacts, 110 had a nonworking telephone number, 111 could not be contacted by telephone, 13 were deceased, and 394 were contacted for a contact rate of 62.7% (394/628). Of the 394 who were contacted, 194 refused to participate, for a refusal rate of 49.2% (194/394); and 200 participants were successfully enrolled and completed the interviews, for an overall participation rate of 31.8% (200/628). Common reasons for refusing included not being interested (90 individuals), being too busy (33 individuals), being too ill (29 individuals), caring for a family member (5 individuals), and having changed location (4 individuals). Those who refused to participate were equally divided among women and men. However, more white (42.3%; 82/194) than African American (22.2%; 43/194) or Latino (0.0%) patients refused to participate, and more urban clinic (66.1%; 128/194) than rural clinic (28.9%; 56/194) patients refused to participate.

Data Collection

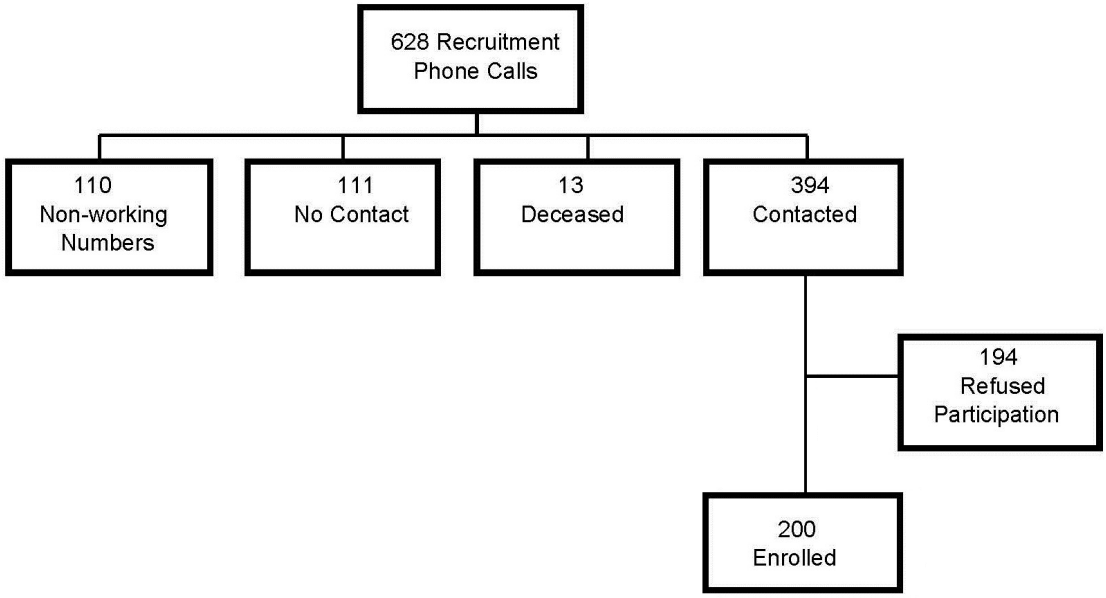
The patient questionnaire included items eliciting information on personal characteristics such as age, race, marital status, and educational attainment; social interaction and social support; health characteristics such as chronic conditions, cognitive status, use of prescription medicines, health-related quality of life, health literacy; Internet and other modes of technology access; access to health care; and orientation and utilization of electronic health information resources such as a patient portal.

Questionnaires were always administered in person by trained interviewers, usually at the participants' homes or at the clinic where they received medical care. Interviews were completed from November 2014 to May 2016. The interview generally took 1 hour to complete and ranged in length from 45 min to 2 hours. Participants were given an incentive of US $20 for completing the interview. Research Electronic Data Capture [54], a secured web-based system, was used to record interview data.

Table 1. Sample size by gender and ethnicity.

Ethnicity	Gender		
	Female	Male	Total
White	43	37	80
African American	53	37	90
Latino	17	9	26
Other	3	1	4
Total	116	84	200

Figure 2. Patient recruitment flowchart.



Measures

Patient portal utilization had the values of ever used versus never used based on patient self-report. Measures of the number of patient portal features used, patient portal positive perceptions, and utilization frequency are included for participants who were ever-users. Participants were asked whether they used 7 patient portal features (send a message to your doctor or nurse, refill prescriptions, view lab or test results, make or change an appointment, request a referral, find information about a health issue, and other). The number of features used were summed and placed in the categories 3 or less and 4 to 7; frequencies of use for the specific features are reported in Multimedia Appendix 1**.** The use frequency item had the values several times a week, several times a month, less than once a month, a few times a year, and never. Responses were placed in the categories at least once a month (once a month, several times a month, and several times a week), and less than one a month (less than once a month, a few times a year, and never). Patient portal positive perceptions included 16 statements adapted from the Technology Acceptance Scale developed by Gardner and Amoroso [55] (eg, "Using my patient portal can enable me to accomplish tasks more quickly" and "Using my patient portal can make it easier to do my tasks."). Agreement with statements was summed, with scores ranging from 0 to 16. Scores were placed in the categories 0 to 12 and 13 to 16 for analysis. Frequencies for the specific positive perceptions are reported in the Multimedia Appendix 1.

Individual user characteristics included age (in the categories 55-59 years, 60-64 years, 65-69 years, and 70 years and older), gender, ethnicity (in the categories white and minority), and education (in the categories high school or less and greater than high school). Employment had the values of not employed or employed (whether part- or full-time). Poverty level was based on the total household income divided by the total number of residents and adjusted for the year in which the data were collected [56]. Poverty level was placed in the categories less than 100% of poverty level, 100% to 200% of poverty level, and greater than 200% of poverty level. Health status was measured with the Charlson Comorbidity Index [57] that includes the self-reported diagnosis of 18 different chronic conditions. Participants were classified as having fewer than 5 versus 5 or more chronic conditions. All of the participants had at least one chronic condition (diabetes, hypertension, dyslipidemia, or cardiovascular disease) that required lab tests or ongoing medication. The categories used differentiate the ill from the very ill. Health literacy was measured with the Newest Vital Sign (NVS) [58]. This scale had a range of 0 to 6 and had a Cronbach alpha of .82. On the basis of the recommendation of the scale developers, those with a score of 4 to 6 were considered to have adequate health literacy, and those with a score of 0 to 3 were considered to have inadequate health literacy.

Three social support measures were included. Marital status had the values currently married and not currently married. Household composition had the values live alone, live with a spouse and no others, live with spouse and adult child, and other. Having a care partner was dichotomous. A care partner was defined for the participants as, "someone who helps (you) with activities and questions about (your) health. These activities and questions include simple things, like reminding you to take your prescription and about an upcoming doctor's appointment, or finding information about something the doctor has told you; they can include more substantial assistance, like taking you to an appointment, helping you take your medicine, and helping you exercise; and they can include personal assistance, like helping you get dressed and bathe. Those who help you can be your spouse, brothers or sisters, children, or friends. The person

who helps you may live with you, but they can also live in another house. They might even live in another town or city and help you by phone or the Internet."

Health insurance, the first organizational measure, had the values private insurance, government insurance (eg, Medicare, Medicaid, and Veterans Administration), and no insurance. Difficulty in contacting the medical office during regular hours was dichotomous. Difficulty in contacting the medical office after regular hours and whether the medical office has night or weekend office hours had the values no, yes, and don't know.

Whether the patient was recruited from a rural (Greene County Health Care Inc or West Caldwell Health Council Inc) or urban (Wake Forest Baptist Health Out Patient Department) clinic was the first environmental measure. Difficulty in accessing email in the county had the values difficult, easy, and don't know.

Human-technology interaction measures included whether the patient sends and receives emails. eHealth literacy was measured with the 8-item eHealth Literacy Scale (eHEALS) [59,60] that had a range of 8 to 40, an overall mean of 22.7, an SD of 9.5, and a Cronbach alpha of .95. Access to e-devices and Internet at home (including desktop, laptop, tablet, and mobile phone) was dichotomous. Number of e-devices and Internet at home categories had the values of 0, 1, and 2 or more. Internet use frequency had the values less than once a day and at least once a day. Stress experienced when using computer was based on the item, "How much stress do you feel when using a computer?" which had the response categories no stress at all to very much stress; responses were placed in the categories no stress (no stress at all) and some stress (low stress, moderate stress, much stress, and very much stress).

Statistical Analysis

All analyses were performed using SAS 9.4 (SAS Institute Inc). Personal characteristics were compared between patients who ever used a patient portal and those who never used a patient portal using chi-square test for categorical variables and student *t* test for continuous variables. A logistic regression model was used to examine association between personal characteristics and patient portal utilization. Factors in each of the conceptual framework key domains (the individual user, social support, organizational characteristics, environment, and human technology) that had a statistically significance associations with patient portal utilization based on chi-square tests and *t* tests were included in the logistic regression model. No organizational characteristics had a statistically significant bivariate association with patient portal utilization; the health insurance variable was selected to represent this key factor in the logistic regression model. Odds ratios (OR) with the corresponding 95% CI were estimated for each characteristic. Furthermore, associations between participants' patient portal utilization and perceived usefulness were examined in terms of factors that remained statistically significant in the multivariable logistic regression model using chi-square test. All tests were two-sided and performed at a significance level of .05.

Results

Factors Associated With Ever Using a Patient Portal

Participant characteristics, organized by domain within the conceptual framework and their association with patient portal utilization are reported in Table 2. A total of 41 (20.5%) participants reported utilizing their patient portals. Patient portal utilization did not differ by participant age or gender. More white participants (37.5%) than minority participants (9.2%), and more with greater than a high school education (47.1%) than with a high school education or less (6.2%) had utilized their patient portal. Employment was not associated with patient portal utilization but poverty level was: 9.0% of those below the poverty level, 25.3% of those at 100% to 200 % of the poverty level, and 46.9% of those above 200% of the poverty level had utilized their patient portal. Those with worse health, as indicated by a Charlson Comorbidity Index of 5 or more, utilized their patient portal more (25.4%) than did those with a score below 5 (9.7%). Over half (50.9%) of those with adequate health literacy (NVS score of 4-6) utilized their patient portal, whereas 9.6% of those with inadequate health literacy utilized their patient portal.

Participants who were married (29.8%) and lived only with their spouse (34.5%) utilized their portal more than those not married (13.8%) and who had other household composition. Identifying a care partner was not associated with patient portal utilization. Type of health insurance, difficulty with contacting the medical office during or after regular hours, and whether the medical office had night or weekend office hours were not associated with portal utilization. Receiving care at an urban (30.0%) rather than a rural (6.3%) clinic was associated with portal utilization. Difficulty of accessing email in the county was not associated with patient portal utilization.

Greater eHealth literacy, as measured by the eHEALS scale, was associated with patient portal utilization; the mean eHEALS value for the entire sample was 22.7 (SD 9.5), whereas it was 31.7 (SD 6.5) for portal users and 20.2 (SD 8.7) for those not utilizing their portal (P<.001). A far greater percentage of participants who send and receive email (51.3%) utilized a patient portal than those who did not use email (0.8%). Those with access to e-devices and Internet in their homes (33.9% vs 1.2%), those with 2 or more e-devices in their home (36.8% vs 12.2% with 1 device and 0% with no device), who use the Internet at least once a day (47.5% vs 8.6%), and who experience no stress when using a computer (50.0% vs 11.3% who experience at least some stress) were more likely to utilize their patient portal.

Table 2. Participant characteristics and their association with patient portal utilization.

Participant characteristics	Overall (N=200) n (%)	Ever used patient portal (n=41) n (%)	*P* value
Individual user			
Age in years			.71
55-59	55 (27.5)	11 (20.0)	
60-64	63 (31.5)	15 (23.8)	
65-69	52 (26.0)	11 (21.2)	
70 and older	30 (15.0)	4 (13.3)	
Gender			.32
Male	84 (42.0)	20 (23.8)	
Female	116 (58.0)	21 (18.1)	
Ethnicity			<.001
White	80 (40.0)	30 (37.5)	
Minority	120 (60.0)	11 (9.2)	
Education			<.001
High school or less	130 (65.0)	8 (6.2)	
Greater than high school	70 (35.0)	33 (47.1)	
Employment			.34
Not employed	160 (80.0)	35 (21.9)	
Employed (part-time or full-time)	40 (20.0)	6 (15.0)	
Poverty level			<.001
Less than 100% of poverty level	89 (46.4)	8 (9.0)	
100-200% of poverty level	71 (37.0)	18 (25.3)	
Greater than 200% of poverty level	32 (16.6)	15 (46.9)	
Charlson Comorbidity Index			.03
Fewer than 5	62 (31.0)	6 (9.7)	
5 or more	138 (69.0)	35 (25.4)	
Newest Vital Sign			<.001
Adequate literacy (score of 4-6)	53 (28.0)	27 (50.9)	
Inadequate literacy (0-3)	136 (72.0)	13 (9.6)	
Social support			
Marital status			<.01
Currently married	84 (42.0)	25 (29.8)	
Not currently married	116 (58.0)	16 (13.8)	
Household composition			.02
Live alone	70 (35.0)	12 (17.1)	
Live with spouse (no other residents)	55 (27.5)	19 (34.5)	
Live with spouse and adult child	25 (12.5)	4 (16.0)	
Other	50 (25.0)	6 (12.0)	
Care partner			.88
No	76 (38.0)	16 (21.1)	
Yes	124 (62.0)	25 (20.2)	

Participant characteristics	Overall (N=200) n (%)	Ever used patient portal (n=41) n (%)	*P* value
Organizational characteristics			
Health insurance			.22
Private insurance	54 (27.0)	15 (27.8)	
Government insurance	121 (60.5)	23 (19.1)	
None	25 (12.5)	3 (12.0)	
Difficulty in contacting the medical office during regular hours			.67
No	126 (63.0)	27 (21.4)	
Yes	74 (37.0)	14 (18.9)	
Difficulty in contacting the medical office after regular hours			.08
No	57 (28.5)	13 (22.8)	
Yes	65 (32.5)	18 (27.7)	
Don’t know	78 (39.0)	10 (12.8)	
Medical office has night or weekend office hours			.16
No	99 (49.5)	25 (25.3)	
Yes	50 (25.0)	6 (12.0)	
Don’t know	51 (25.5)	10 (19.6)	
Environment			
Clinic			<.001
Rural	80 (40.0)	5 (6.3)	
Urban	120 (60.0)	36 (30.0)	
Difficulty in accessing email in the county			.15
Difficult (very difficult, difficult, neutral)	32 (16.0)	7 (21.9)	
Easy (easy, very easy)	133 (66.5)	31 (23.3)	
Don’t know	35 (17.5)	3 (8.6)	
Human technology			
Send and receive email			<.001
No	122 (61.0)	1 (0.8)	
Yes	78 (39.0)	40 (51.3)	
Access to e-devices and Internet at home			<.001
No	82 (41.0)	1 (1.2)	
Yes	118 (59.0)	40 (33.9)	
Number of e-devices and Internet at home			<.001
0	56 (28.0)	0	
1	49 (24.5)	6 (12.2)	
2 or more	95 (47.5)	35 (36.8)	
Internet use frequency			<.001
Less than once a day	139 (69.5)	12 (8.6)	
At least once a day	61 (30.5)	29 (47.5)	
Stress experienced when on computer			<.001
No stress	48 (24.1)	24 (50.0)	
Some stress	151 (75.9)	17 (11.3)	

Multivariate analysis addressing patient portal utilization was conducted using measures from each framework domain (Table 3). Measures selected for this analysis were significant in the bivariate analysis but not collinear with other measures. The organizational characteristic measure health insurance was included in the multivariate analysis even though it did not have a significant bivariate association to have this key factor included. The human-technology measures were all highly intercorrelated; Internet use frequency was selected over other measures (eg, eHEALS score) because asking a patient whether they used the Internet at least once a day is a procedure that a health care provider promoting patient portal utilization could easily accomplish during a patient visit.

Several personal characteristics were not significantly associated with patient portal utilization in the multivariate analysis, including ethnicity, health status or Charlson Index, and type of health insurance. Other personal characteristics sustained statistically significant associations with patient portal utilization. Those with greater than a high school education had greater odds of patient portal utilization (Odds ratio [OR] 5.75, 95% CI 1.94-17.04). Those who were not currently married had lesser odds of patient portal utilization (OR 0.17, 95% CI 0.06-0.52). Receiving care at an urban clinic greatly increased the odds of patient portal utilization (OR 12.21, 95% CI 3.05-48.87). Finally, using the Internet at least daily increased the odds of patient portal utilization (OR 7.08, 95% CI 2.55-19.67).

Utilization of Patient Portal Features

Of 41 participants who utilized their patient portal, almost half (48.8%) utilized at least four portal features (Table 4), such as sending a message to a doctor or nurse or refilling prescriptions. Most (70.7%) users utilized their patient portal at least once a month. Most (61.0%) users had positive perceptions of most patient portal attributes. Users who were not currently married more often (68.8%) utilized at least four portal features than those who were currently married (36.0%; Table 5). Users who received care at urban clinics more often (55.6%) utilized at least four portal features than did those who received care at rural clinics (0.0%). Frequency of portal utilization did not differ among users for the characteristics considered (Table 6). Positive perceptions of patient portal attributes were greater among those who received care at urban clinics (69.4%) than at rural clinics (0.0%) (Table 7).

Table 3. Logistic regression models of patient portal utilization.

Personal characteristics	Ever used patient portal	
	Odds ratio (95% CI)	*P* value
Minority versus white (white as reference)	0.55 (0.19-1.60)	.27
Greater than high school versus high school or less	5.75 (1.94-17.04)	<.01
Charlson Index: 5 or more versus fewer than 5	2.43 (0.65-9.13)	.18
Not currently married versus married	0.17 (0.06-0.52)	.001
Private insurance versus none	0.73 (0.11-5.07)	.86
Government insurance versus none	0.61 (0.09-4.10)	
Urban versus rural clinic	12.21 (3.05-48.87)	<.01
Internet frequency: at least once a day versus less than once a day	7.08 (2.55-19.67)	<.01

Table 4. Participant patient portal utilization and perceived usefulness.

Patient portal utilization and perceived usefulness	n (%)
Patient portal features utilized	
3 or less	21 (51.2)
4 to 7	20 (48.8)
Utilization frequency	
Less than once a month	12 (29.3)
At least once a month	29 (70.7)
Patient portal positive perceptions	
12 or less	16 (39.0)
13 to 16	25 (61.0)

Table 5. Factors associated with patient portal feature utilization.

Participant characteristics	Patient portal features used		
	3 or less, n (%)	4 to 7, n (%)	*P* value
Education			.70
High school or less	5 (62.5)	3 (37.5)	
Greater than high school	16 (48.5)	17 (51.5)	
Marital status			.04
Currently married	16 (64.0)	9 (36.0)	
Not currently married	5 (31.2)	11 (68.8)	
Clinic			.03
Rural	5 (100.0)	0 (0.0)	
Urban	16 (44.4)	20 (55.6)	
Frequency of Internet use			.20
Less than once per day	8 (66.7)	4 (33.3)	
At least once per day	13 (44.8)	16(55.2)	

Table 6. Factors associated with patient portal utilization frequency.

Participant characteristics	Use frequency		
	Less than once per month n (%)	At least once per month n (%)	*P* value
Education			>.99
High school or less	6 (75.0)	2 (25.0)	
Greater than high school	23 (69.7)	10 (30.3)	
Marital status			.73
Currently married	17 (68.0)	8 (32.0)	
Not currently married	12 (75.0)	4 (25.0)	
Clinic			.13
Rural	2 (40.0)	3 (60.0)	
Urban	27 (75.0)	9 (25.0)	
Frequency of Internet use			.28
Less than once per day	7 (58.3)	5 (41.7)	
At least once per day	22 (75.9)	7 (24.1)	

Table 7. Factors associated with patient portal positive perceptions.

Participant characteristics	Patient portal positive perceptions		
	12 or less, n (%)	13 to 16, n (%)	*P* value
Education			.12
High school or less	1 (12.5)	7 (87.5)	
Greater than high school	15 (45.5)	18 (54.56	
Marital status			.41
Currently married	11 (44.0)	14 (56.0)	
Not currently married	5 (31.3)	11 (68.8)	
Clinic			<.01
Rural	5 (100.0)	0 (0.0)	
Urban	11 (30.6)	25 (69.4)	
Frequency of Internet use			.48
Less than once per day	6 (50.0)	6 (50.0)	
At least once per day	10 (34.5)	19 (65.5)	

Discussion

Principal Findings

Only a moderate proportion of patients participating in this study (20.5%) reported utilizing their patient portal when compared with the results of other primary surveys [18,19] and to the analyses of electronic records [14-17]. Patient portal utilization among patients participating in our survey was not comparable to the 80% utilization among older patients in the California Kaiser Permanente system [14]. Contrary to other research, we did not find differences in portal utilization by age or gender among our participating patients. The lack of variation in portal utilization by age may reflect the relatively "young" patients who participated, with few over the age of 75 years. Having insurance, whether private or government, was not associated with patient portal utilization as indicated by Ancker and colleagues [15], perhaps because so many of our participants had government insurance (Medicare) and so few had no insurance. We did find differences in portal utilization by ethnicity, education, poverty level, and rurality; those who are minority, have lower income and education, and are rural utilize patient portals less. Differences in participant portal utilization by these characteristics reflect disparities. Even in this relatively low income population, a social gradient in utilization is apparent [61].

As qualitative analyses suggested, familiarity and use of technology were associated with greater patient portal utilization [20-23]. Those patients with limited access to electronic devices in their homes, who seldom used electronic devices, and who experienced stress when using computers were less likely to have utilized a patient portal. Health literacy and eHealth literacy were associated with patient portal utilization in the bivariate analysis. Health literacy is important in mediating the ability of patients to access their patient portals and to interpret the medical information on their portals, which may influence willingness to use [62,63]. Tieu et al [22,23] report that health literacy was an important factor in their utilization of their patient portals.

Our multivariate analysis delineates factors that are important to patient portal utilization, including education, having a spouse (the most common form of social support), and frequent use of the Internet. It reveals that receiving health care in rural communities is associated with limited patient portal utilization; rural communities have less Internet access [64].

The results for patient portal usefulness and usability are somewhat tautological, that is, those who utilize their patient portal at all generally utilize several features. Most utilize their portal at least several times per month, and they generally have positive perceptions of their portal. Few factors differentiate levels of utilization and perceived usefulness, with the exception of rurality. Rural older adults utilize fewer portal features and have fewer positive perceptions; this reflects the fewer features available. The portal available through the rural clinics is not as robust as that available through the urban clinic. Future research needs to include a clear understanding of patient portal sophistication when comparing utilization and usability among users.

This conceptually based analysis indicates leverage points for improving patient portal utilization in health disparities communities, particularly minority and rural communities, in which a greater portion of members have chronic conditions and less access to health care than in the general population. Rurality limits patient portal utilization. Improving Internet connectively across all rural communities would improve patients' ability to connect their patient portals [64]. Of course, improving Internet connectivity in rural communities would lead to other social and economic benefits [65]. Such infrastructure development for rural communities is not novel; the Rural Electrification Administration [66] provided similar rural infrastructure development in the 1930s.

Ethnicity does not remain a significant limitation to patient portal utilization when we control for other factors. Education remains a significant advantage for patient portal utilization. The complement to education, familiarity and regular use of Internet applications also improves patient portal utilization as documented in this and other analyses [20-23]. Community programs that provide Internet training and access for residents, particularly older adults and members of vulnerable minority populations, will help improve their patient portal utilization. Libraries have become Internet user hubs [67], including those in rural communities, and provide one institution that could be recruited for this purpose. Most communities, including rural communities, have facilities in addition to libraries, such as senior centers, congregate meal sites, recreational centers, and churches, in which computers can be located for Internet training and access; for example, WinstonNET has provided such training in Winston-Salem, North Carolina, for over 15 years [68]. Ensuring that residents know that they can access the Internet at diverse locations in their communities can support patient portal utilization. Making patient portals available in the language in which the patient is most comfortable would also improve utilization.

Familiarity and use of the Internet also provides an easy indicator for clinicians when assessing patients for potential patient portal utilization. Asking patients how often they use the Internet or email is more parsimonious than any multi-item scale for predicting patient portal adoption. Finally, involving a patients' social support, particularly a spouse, can improve patient portal adoption and utilization. Sarkar and Bates [69] note the importance of involving family members or other "care partners" in training older adults to utilize patient portals. Such social support can be extremely important for attaining these skills within health disparities communities. Helping patients involve family members and share health information can improve the role of patient portals in the provision of patient-centered care. However, more research is needed to ensure processes that protect patients' sense of privacy and autonomy [70].

Health care organizations must maintain the position that patient portals are a crucial mechanism to improve patient health and well-being, and they must convey this to patients to ensure utilization. Since this research began, the standards for patient portal meaningful use have deteriorated substantially, simply having a portal and having utilization by a single patient constitutes compliance. Some organizations do see the potential for patients. For example, since we began writing this paper, one organization, Wake Forest Baptist Health, has initiated a marketing campaign to increase patient portal utilization. This campaign includes billboards, radio advertisements, and television advertisements of very high production values.

Limitations

This research should be evaluated within its limitations. The sample was drawn from patients receiving care at three sets of clinics (one urban and two rural). The participation rate was limited. These factors limit the generalizability of the results. The urban and rural clinics differed in the features available in their patient portals and in the time that they had been established before data collection was conducted. These factors could affect the differences between patients in their patient portal utilization. The key outcome measure, patient portal use, is based on self-reported use rather than capturing actual use using EHRs. This could limit the validity of this measure. At the same time, this survey did recruit a large, multi-ethnic, low income sample that included rural and urban patients.

Conclusions

This analysis found that variation in patient portal utilization reflects disparities, even in low income patient populations. Our conceptual approach allows the delineation of leverage points to address these disparities that can be addressed by public policy and health policy, specifically the need to provide training for older adults in the utilization of IT in general and specifically of patient portals, involving family members (spouses) or other care partners in this training [69] and making IT widely (geographically) available at public places in rural as well as urban communities [67]. Future research should examine whether these strategies successfully lead to higher rates of portal use by vulnerable older adults.

Acknowledgments

The research was supported by grant R01 HS021679 from the Agency for Healthcare Research and Quality.

Authors' Contributions

TAA, SAQ, JCS, DPM, CL, KL, AS, and AGB collaborated in the overall design of the research project. KPM managed data collection activities. XL and JT were responsible for data management and data analysis. TAA led writing of this paper. All authors provided critical comments on the content of this paper and agreed on the final document.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental tables.

[PDF File (Adobe PDF File), 29KB - medinform_v5i4e47_app1.pdf]

References

1. Ahern DK, Woods SS, Lightowler MC, Finley SW, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. Am J Prev Med 2011 May;40(5 Suppl 2):S162-S172. [doi: 10.1016/j.amepre.2011.01.005] [Medline: 21521591]
2. Bates DW, Bitton A. The future of health information technology in the patient-centered medical home. Health Aff (Millwood) 2010 Apr;29(4):614-621 [FREE Full text] [doi: 10.1377/hlthaff.2010.0007] [Medline: 20368590]
3. Carter EL, Nunlee-Bland G, Callender C. A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. Perspect Health Inf Manag 2011;8:1b [FREE Full text] [Medline: 21307985]
4. Hargreaves JS. Will electronic personal health records benefit providers and patients in rural America? Telemed J E Health 2010 Mar;16(2):167-176. [doi: 10.1089/tmj.2009.0063] [Medline: 20082592]
5. Herrin J, da Graca B, Nicewander D, Fullerton C, Aponte P, Stanek G, et al. The effectiveness of implementing an electronic health record on diabetes care and outcomes. Health Serv Res 2012 Aug;47(4):1522-1540 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01370.x] [Medline: 22250953]
6. Kim EH, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
7. Krist A, Woolf S, Rothemich S. Healthit.ahrq. Rockville, MD: The Agency for Healthcare Research and Quality (AHRQ); 2011. An interactive preventive health record (IPHR) to promote patient-centered preventive care URL: https://healthit.ahrq.gov/sites/default/files/docs/publication/r18hs017046-krist-final-report-2011.pdf [accessed 2017-02-12] [WebCite Cache ID 6qNdDvaxM]
8. O'Donnell HC, Patel V, Kern LM, Barrón Y, Teixeira P, Dhopeshwarkar R, et al. Healthcare consumers' attitudes towards physician and personal use of health information exchange. J Gen Intern Med 2011 Sep;26(9):1019-1026 [FREE Full text] [doi: 10.1007/s11606-011-1733-6] [Medline: 21584839]
9. Snyder CF, Jensen R, Courtin SO, Wu AW, Website for Outpatient QOL Assessment Research Network. PatientViewpoint: a website for patient-reported outcomes assessment. Qual Life Res 2009 Sep;18(7):793-800 [FREE Full text] [doi: 10.1007/s11136-009-9497-8] [Medline: 19544089]
10. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patientt. Health Aff (Millwood) 2010 Jul 06;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
11. Pearl R. Kaiser Permanente Northern California: current experiences with internet, mobile, and video technologies. Health Aff (Millwood) 2014 Feb;33(2):251-257 [FREE Full text] [doi: 10.1377/hlthaff.2013.1005] [Medline: 24493768]
12. The Health Information Technology Resource Center. Rwjf.: Robert Wood Johnson Foundation; 2012 Jul. The value of personal health records and web portals to engage consumers and improve quality URL: https://www.rwjf.org/en/library/research/2012/07/the-value-of-personal-health-records-and-web-portals-to-engage-c.html [accessed 2017-10-31] [WebCite Cache ID 6uczDR7cV]
13. Otte-Trojel T, de Bont A, van de Klundert J, Rundall TG. Characteristics of patient portals developed in the context of health information exchanges: early policy effects of incentives in the meaningful use program in the United States. J Med Internet Res 2014;16(11):e258 [FREE Full text] [doi: 10.2196/jmir.3698] [Medline: 25447837]
14. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
15. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
16. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient Portals: Who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501. [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
17. Wallace LS, Angier H, Huguet N, Gaudino JA, Krist A, Dearing M, et al. Patterns of electronic portal use among vulnerable patients in a nationwide practice-based research network: from the OCHIN practice-based research network (PBRN). J Am Board Fam Med 2016 Oct;29(5):592-603 [FREE Full text] [doi: 10.3122/jabfm.2016.05.160046] [Medline: 27613792]
18. Peacock S, Reddy A, Leveille SG, Walker J, Payne TH, Oster NV, et al. Patient portals and personal health information online: perception, access, and use by US adults. J Am Med Inform Assoc 2017 Apr 01;24(e1):e173-e177. [doi: 10.1093/jamia/ocw095] [Medline: 27413120]
19. 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]
20. Latulipe C, Gatto A, Nguyen HT, Miller DP, Quandt SA, Bertoni AG, et al. Design considerations for patient portal adoption by low-income, older adults. Proc SIGCHI Conf Hum Factor Comput Syst 2015 Apr;2015:3859-3868 [FREE Full text] [doi: 10.1145/2702123.2702392] [Medline: 27077140]

21. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. "I Want to Keep the Personal Relationship With My Doctor": understanding barriers to portal use among African Americans and Latinos. J Med Internet Res 2016 Oct 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
22. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017;24(e1):e47-e54. [doi: 10.1093/jamia/ocw098] [Medline: 27402138]
23. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
24. Zickmund SL, Hess R, Bryce CL, McTigue K, Olshansky E, Fitzgerald K, et al. Interest in the use of computerized patient portals: role of the provider-patient relationship. J Gen Intern Med 2008 Jan;23 Suppl 1:20-26 [FREE Full text] [doi: 10.1007/s11606-007-0273-6] [Medline: 18095039]
25. Keplinger LE, Koopman RJ, Mehr DR, Kruse RL, Wakefield DS, Wakefield BJ, et al. Patient portal implementation: resident and attending physician attitudes. Fam Med 2013 May;45(5):335-340 [FREE Full text] [Medline: 23681685]
26. Miller Jr DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
27. 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]
28. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
29. Lyles CR, Sarkar U, Schillinger D, Ralston JD, Allen JY, Nguyen R, et al. Refilling medications through an online patient portal: consistent improvements in adherence across racial/ethnic groups. J Am Med Inform Assoc 2016 Apr;23(e1):e28-e33 [FREE Full text] [doi: 10.1093/jamia/ocv126] [Medline: 26335983]
30. Goldzweig CL, Orshansky G, Paige NM, Miake-Lye IM, Beroes JM, Ewing BA, et al. Electronic health record-based interventions for improving appropriate diagnostic imaging: a systematic review and meta-analysis. Ann Intern Med 2015 Apr 21;162(8):557-565 [FREE Full text] [doi: 10.7326/M14-2600] [Medline: 25894025]
31. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
32. Lawton MP. Behavior-relevant ecological factors. In: Schaie KW, Schooler C, editors. Social Structure and Aging: Psychological Processes. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1989:57-78.
33. Dalcher I, Shine J. Extending the new technology acceptance model to measure the end user information systems satisfaction in a mandatory environment: a bank's treasury. Technol Anal Strateg Manage 2003 Dec;15(4):441-455. [doi: 10.1080/0953732030000136033]
34. Bruner GC, Kumar A. Explaining consumer acceptance of handheld Internet devices. J Bus Res 2005 May;58(5):553-558. [doi: 10.1016/j.jbusres.2003.08.002]
35. Chau PY. An empirical investigation on factors affecting the acceptance of CASE by systems developers. Inf Manag 1996 Sep;30(6):269-280 [FREE Full text] [doi: 10.1016/S0378-7206(96)01074-9]
36. 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]
37. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Aug;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
38. Fishbein M, Azjen I. Belief, Attitude, Intention, Behavior: An introduction to Theory and Research. Reading, MA: Addison-Wesley Publishing Co; 1975:1-578.
39. Adams DA, Nelson RR, Todd PA. Perceived usefulness, ease of use, and usage of information technology: a replication. MIS Q 1992:227-247 [FREE Full text] [doi: 10.2307/249577]
40. Lu HP, Gustafson DH. An empirical study of perceived usefulness and perceived ease of use on computerized support system use over time. Int J Inf Manage 1994 Oct;14(5):317-329 [FREE Full text] [doi: 10.1016/0268-4012(94)90070-1]
41. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
42. Horrigan JB. Online.wsj. New York, NY: Wall Street Journal; 2010. Broadband adoption and use in America URL: http://online.wsj.com/public/resources/documents/FCCSurvey.pdf [accessed 2011-09-16] [WebCite Cache ID 6qKjAqXFe]
43. LaRose R, Gregg JL, Strover S, Straubhaar J, Carpenter S. Closing the rural broadband gap: promoting adoption of the Internet in rural America. Telecommun Policy 2007 Jul;31(6-7):359-373 [FREE Full text] [doi: 10.1016/j.telpol.2007.04.004]
44. Czaja SJ, Moen P. Technology and employment. In: Pew RW, Van Hemel SB, editors. Technology for Adaptive Aging. Washington DC: National Academies Press; 2004:150-178.
45. Kinsella K, Velkoff VA. Census. Washington, DC: US Government Printing Office; 2001. An aging world: 2001 URL: https://www.census.gov/prod/2001pubs/p95-01-1.pdf [accessed 2017-02-12] [WebCite Cache ID 6qNilRU7E]

46. Fox S. Pewinternet. Washington, DC: Pew Internet and American Life Project; 2009. The social life of health information, 2011 URL: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2017-02-12] [WebCite Cache ID 6qNgYyx9i]
47. Ntia.doc. Washington, DC: US Government Printing Office; 2002. A nation online: how Americans are expanding their use of the Internet URL: https://www.ntia.doc.gov/legacy/ntiahome/dn/anationonline2.pdf [accessed 2017-05-11] [WebCite Cache ID 6qNfvx4jb]
48. Duyck P, Pynoo B, Devolder P, Voet T, Adang L, Vercruysse J. User acceptance of a picture archiving and communication system. Applying the unified theory of acceptance and use of technology in a radiological setting. Methods Inf Med 2008;47(2):149-156. [Medline: 18338086]
49. Czaja SJ, Boot WR, Charness N, Rogers WA, Sharit J. Improving social support for older adults through technology: findings from the PRISM randomized controlled trial. Gerontologist 2017 Feb 15 [Epub ahead of print] (forthcoming) [FREE Full text] [doi: 10.1093/geront/gnw249] [Medline: 28201730]
50. Li X, Hess TJ, Valacich JS. Why do we trust new technology? A study of initial trust formation with organizational information systems. J Strateg Inf Syst 2008 Mar;17(1):39-71. [doi: 10.1016/j.jsis.2008.01.001]
51. Merritt SM, Ilgen DR. Not all trust is created equal: dispositional and history-based trust in human-automation interactions. Hum Factors 2008 Apr;50(2):194-210 [FREE Full text] [doi: 10.1518/001872008X288574] [Medline: 18516832]
52. Bell P, Reddy P, Rainie L. Pewinternet. Washington, DC: Pew Internet & American Life Project; 2004. Rural Americans' Internet use has grown but they continue to lag behind others URL: http://www.pewinternet.org/2004/02/17/rural-americans-internet-use-has-grown-but-they-continue-to-lag-behind-others/ [accessed 2017-05-09] [WebCite Cache ID 6qKivCylg]
53. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
54. 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]
55. Gardner C, Amoroso D. Development of an instrument to measure the acceptance of internet technology by consumers. Washington DC: IEEE; 2004 Presented at: Proceedings of the 37th Hawaii International Conference on System Sciences; January 5-8, 2004; Big Island, HI, USA. [doi: 10.1109/HICSS.2004.1265623]
56. United States Census Bureau. Census. Poverty URL: http://www.census.gov/topics/income-poverty/poverty.html [accessed 2017-05-09] [WebCite Cache ID 6qKib8GQx]
57. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [Medline: 3558716]
58. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
59. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
60. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol 2011 Apr;64(4):366-379. [doi: 10.1016/j.jclinepi.2010.04.005] [Medline: 20638235]
61. Deaton A. Policy implications of the gradient of health and wealth. Health Aff (Millwood) 2002;21(2):13-30 [FREE Full text] [Medline: 11900153]
62. Czaja SJ, Zarcadoolas C, Vaughon WL, Lee CC, Rockoff ML, Levy J. The usability of electronic personal health record systems for an underserved adult population. Hum Factors 2015 May;57(3):491-506. [doi: 10.1177/0018720814549238] [Medline: 25875437]
63. Ackerman SL, Sarkar U, Tieu L, Handley MA, Schillinger D, Hahn K, et al. Meaningful use in the safety net: a rapid ethnography of patient portal implementation at five community health centers in California. J Am Med Inform Assoc 2017 Sep 1;24(5):903-912. [doi: 10.1093/jamia/ocx015] [Medline: 28340229]
64. Gao. Washington, DC: Government Accounting Office; 2006. Broadband deployment is extensive throughout the United States, but it is difficult to assess the extent of deployment gaps in rural areas URL: http://www.gao.gov/new.items/d06426.pdf [accessed 2017-05-09] [WebCite Cache ID 6qKiQwfo8]
65. Brown J. Govtech.: Government Technology (GT) Network; 2014. The challenges of broadband in rural libraries URL: http://www.govtech.com/network/The-Challenges-of-Broadband-in-Rural-Libraries.html [accessed 2017-05-09] [WebCite Cache ID 6qKiBNFaa]
66. Beall RT. Naldc.nal.usda. Washington, DC: United States Department of Agriculture; 1940. Rural electrification URL: https://naldc.nal.usda.gov/download/IND43893747/PDF [accessed 2017-02-20] [WebCite Cache ID 6qKhzaWt2]
67. Clark L, Perry KA. After access: libraries & digital empowerment. Building digitally inclusive communities.: American Library Association Ala URL: http://www.ala.org/advocacy/sites/ala.org.advocacy/files/content/ALA%20DI%20After%20Access_final_12%2017%2015.pdf [accessed 2017-05-09] [WebCite Cache ID 6qKhSFa6K]

68. Winstonnet. A community resource for computer access and training URL: http://www.winstonnet.org/index.html [accessed 2017-05-09] [WebCite Cache ID 6qKh4w30e]
69. Sarkar U, Bates DW. Care partners and online patient portals. J Am Med Assoc 2014;311(4):357-358. [doi: 10.1001/jama.2013.285825] [Medline: 24394945]
70. Crotty BH, Walker J, Dierks M, Lipsitz L, O'Brien J, Fischer S, et al. Information sharing preferences of older patients and their families. J Am Med Assoc Intern Med 2015 Sep;175(9):1492-1497. [doi: 10.1001/jamainternmed.2015.2903] [Medline: 26147401]

Abbreviations

eHealth: electronic health
eHEALS: eHealth Literacy Scale
EHR: electronic health record
IT: information technology
NVS: Newest Vital Sign
OR: odds ratio
TAM: technology acceptance model

Edited by T Kool; submitted 11.05.17; peer-reviewed by T Giardina, N Gordon, M de Mul; comments to author 12.06.17; revised version received 03.07.17; accepted 21.09.17; published 14.11.17.

Please cite as:
Arcury TA, Quandt SA, Sandberg JC, Miller Jr DP, Latulipe C, Leng X, Talton JW, Melius KP, Smith A, Bertoni AG
Patient Portal Utilization Among Ethnically Diverse Low Income Older Adults: Observational Study
JMIR Med Inform 2017;5(4):e47
URL: http://medinform.jmir.org/2017/4/e47/
doi:10.2196/medinform.8026
PMID:29138129

Original Paper

Ranking Medical Terms to Support Expansion of Lay Language Resources for Patient Comprehension of Electronic Health Record Notes: Adapted Distant Supervision Approach

Jinying Chen[1], PhD; Abhyuday N Jagannatha[2], MSc; Samah J Fodeh[3], PhD; Hong Yu[1,4], FACMI, PhD

[1]Department of Quantitative Health Sicences, University of Massachusetts Medical School, Worcester, MA, United States

[2]School of Computer Science, University of Massachusetts, Amherst, MA, United States

[3]Yale Center for Medical Informatics, Yale University, New Haven, CT, United States

[4]Bedford Veterans Affairs Medical Center, Bedford, MA, United States

Corresponding Author:
Jinying Chen, PhD
Department of Quantitative Health Sicences
University of Massachusetts Medical School
368 Plantation Street
Worcester, MA, 01605
United States
Phone: 1 774 455 3527
Fax: 1 508 856 8993
Email: jinying.chen@umassmed.edu

Abstract

Background: Medical terms are a major obstacle for patients to comprehend their electronic health record (EHR) notes. Clinical natural language processing (NLP) systems that link EHR terms to lay terms or definitions allow patients to easily access helpful information when reading through their EHR notes, and have shown to improve patient EHR comprehension. However, high-quality lay language resources for EHR terms are very limited in the public domain. Because expanding and curating such a resource is a costly process, it is beneficial and even necessary to identify terms important for patient EHR comprehension first.

Objective: We aimed to develop an NLP system, called adapted distant supervision (ADS), to rank candidate terms mined from EHR corpora. We will give EHR terms ranked as high by ADS a higher priority for lay language annotation—that is, creating lay definitions for these terms.

Methods: Adapted distant supervision uses distant supervision from consumer health vocabulary and transfer learning to adapt itself to solve the problem of ranking EHR terms in the target domain. We investigated 2 state-of-the-art transfer learning algorithms (ie, feature space augmentation and supervised distant supervision) and designed 5 types of learning features, including distributed word representations learned from large EHR data for ADS. For evaluating ADS, we asked domain experts to annotate 6038 candidate terms as important or nonimportant for EHR comprehension. We then randomly divided these data into the target-domain training data (1000 examples) and the evaluation data (5038 examples). We compared ADS with 2 strong baselines, including standard supervised learning, on the evaluation data.

Results: The ADS system using feature space augmentation achieved the best average precision, 0.850, on the evaluation set when using 1000 target-domain training examples. The ADS system using supervised distant supervision achieved the best average precision, 0.819, on the evaluation set when using only 100 target-domain training examples. The 2 ADS systems both performed significantly better than the baseline systems (P<.001 for all measures and all conditions). Using a rich set of learning features contributed to ADS's performance substantially.

Conclusions: ADS can effectively rank terms mined from EHRs. Transfer learning improved ADS's performance even with a small number of target-domain training examples. EHR terms prioritized by ADS were used to expand a lay language resource that supports patient EHR comprehension. The top 10,000 EHR terms ranked by ADS are available upon request.

(JMIR Med Inform 2017;5(4):e42) doi:10.2196/medinform.8531

KEYWORDS

electronic health records; natural language processing; lexical entry selection; transfer learning; information extraction

Introduction

Significance and Background

Online patient portals have been widely adopted in the United States in a nationwide effort to promote patient-centered care [1-3]. Many health organizations also allow patients to access their full electronic health record (EHR) notes through patient portals, with early evidence showing improved medical comprehension and health care outcomes [4-6]. However, medical terms—abundant in EHR notes—remain a major obstacle for patients to comprehend medical text, including EHRs [7-12]. In addition, an estimated 36% of adult Americans have limited health literacy [13]. Limited health literacy has been identified as one major barrier to patient use of EHRs [3,14-17]. Misinterpretation of EHR content may result in unintended increases in service utilization and change of patient-provider relationships.

Textbox 1 shows an excerpt from a typical clinical note. The medical terms that may hinder patients' comprehension are italicized. Here we show a subset of medical terms identified by the Unified Medical Language System (UMLS) lexical tool MetaMap [18] for illustration purposes only.

Textbox 1. Illustration of medical terms in a sample clinical note.

Her *creatinine* has shown a steady rise over the past four years. She does have *nephrotic range proteinuria*. The likely *etiology* of her *nephrotic range proteinuria* is her *diabetes*.

She was on an *ACE inhibitor*, which was just stopped in August due to the *elevated creatinine* of 4.41. Given the severity of her *nephrotic syndrome*, her chronic kidney disease is likely permanent; however, I will repeat a *chem-8* now that she is off the *ACE inhibitor*. I will also get a *renal duplex scan* to make sure she does not have any *renal artery stenosis*.

There has been long-standing research interest in developing health information technologies that promote health literacy and consumer-centered communication of health information [19,20]. Natural language processing (NLP)-enabled interventions have also been developed to link medical terms in EHRs to lay terms [21,22] or definitions [23], showing improved comprehension [22,23]. Although there is a substantial amount of health information available on the Internet, many Internet users face challenges accessing and selecting relevant high-quality information [24-27]. The aforementioned NLP-enabled interventions have the advantage of reducing patients' information-seeking burden by integrating authorized health-related information in a single place, and thereby helping patients easily read through and understand their EHR notes.

However, high-quality lay language resources—the cornerstone of such interventions—are very limited in the public domain. The readability levels of health educational materials on the Internet often exceed the level that is easily understood by the average patient [28-30]. Definitions of medical terms provided by controlled health vocabularies, such as those included in the UMLS, often themselves contain complex medical concepts. For example, the term "nephrotic syndrome" in Textbox 1 is defined in the US National Cancer Institute vocabulary as "A collection of symptoms that include severe edema, proteinuria, and hypoalbuminemia; it is indicative of renal dysfunction," where the medical concepts "edema," "proteinuria," "hypoalbuminemia," and "renal dysfunction" may not be familiar to patients.

The consumer health vocabulary (CHV) [31] is a valuable lay language resource that has been integrated into the UMLS and has also been used in EHR simplification [21,22]. CHV contains consumer health terms (which were used by lay people to query online health information) and maps these terms to UMLS concepts. As a result, it contains both lay terms and medical terms, and links between these 2 types of terms. In addition, it provides lay definitions for some medical terms. From our current work, however, we found that CHV alone is not sufficient for comprehending EHR notes, as many medical terms in EHRs do not exist in CHV, and many others exist in CHV but do not have lay terms or lay definitions. For example, among the 19,503 unique terms identified by MetaMap [18] from a corpus of 7839 EHR notes, 4680 (24.0%) terms do not appear in CHV, including "focal motor deficit," "Hartmann procedure," "titrate," and "urethrorectal fistula" (see Multimedia Appendix 1 for more results).

We are building a lay language resource for EHR comprehension by including medical terms from EHRs and creating lay definitions for those terms. This is a time-consuming process that involves collecting candidate definitions from authorized health educational resources, and curating and simplifying these definitions by domain experts. Since the number of candidate terms mined from EHRs is large (hundreds of thousands of terms), we ranked candidate terms based on how important they are for patients' comprehension of EHRs, and therefore prioritized the annotation effort of lexical entries based on those important terms.

The goal of this study was to develop an NLP system to automate the process of lexical entry selection. This task was challenging because the distinctions between important and nonimportant EHR terms in our task were more subtle than that between medical terms and nonmedical terms (detailed below in the Important Terms for Electronic Health Record Comprehension subsection). To achieve this goal, we developed a new NLP system, called adapted distant supervision (ADS), which uses distant supervision from the CHV and uses transfer learning to adapt itself to the target domain to rank terms from EHRs. We aimed to empirically show that ADS is effective in ranking EHR terms at the corpus level and outperforms supervised learning.

Related Work

Natural Language Processing to Facilitate Creation of Lexical Entries

Previous studies have used both unsupervised and supervised learning methods to prioritize terms for inclusion in biomedical and health knowledge resources [32-35]. Term recognition methods, which are widely used unsupervised methods for term extraction, use rules and statistics (eg, corpus-level word and term frequencies) to prioritize technical terms from domain-specific text corpora. Since these methods do not use manually annotated training data, they have better domain portability but are less accurate than supervised learning [32]. The contribution of this study is to propose a new learning-based method for EHR term prioritization, which is more accurate than supervised learning while also having good domain portability.

Our work is also related to previous studies that have used distributional semantics for lexicon expansion [35-37]. In this work, we used word embedding, one technique for distributional semantics, to generate one type of learning features for the ADS system to rank EHR terms.

Ranking Terms in Electronic Health Records

We previously developed NLP systems to rank and identify important terms from each EHR note of individual patients [38,39]. This study is different in that it aimed to rank terms at the EHR corpus level for the purpose of expanding a lay language resource to improve health literacy and EHR comprehension of the general patient population. Notice that both types of work are important for building NLP-enabled interventions to support patient EHR comprehension. For example, a real-world application can link all medical jargon terms in a patient's EHR note to lay terms or definitions, and then highlight the terms most important for this patient and provide detailed information for these important terms.

Distant Supervision

Our ADS system uses distant supervision from the CHV. Distant supervision refers to the learning framework that uses information from knowledge bases to create labeled data to train machine learning models [40-42]. Previous work often used this technique to address context-based classification problems such as named entity detection and relation detection. In contrast, we used it to rank terms without considering context. However, our work is similar in that it uses heuristic rules and knowledge bases to create training data. Although training data created this way often contain noise, distant supervision has been successfully applied to several biomedical NLP tasks to reduce human annotation efforts, including extraction of entities [40,41,43], relations [44-46], and important sentences [47] from the biomedical literature. In this study, we made novel use of the non-EHR-centric lexical resource CHV to create training data for ranking terms from EHRs. This approach has greater domain portability than conventional distant supervision methods due to fewer demands on the likeness between the knowledge base and the target-domain learning task. On the other hand, learning from the distantly labeled data with a mismatch to the target task is more challenging. We address this challenge by using transfer learning.

Transfer Learning

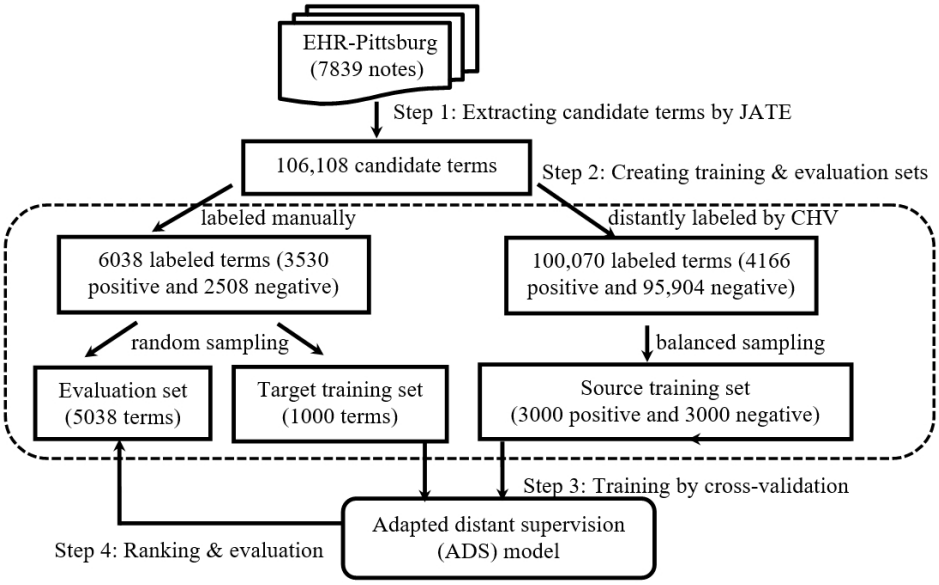
Transfer learning is a learning framework that transfers knowledge from the source domain D_S (the training data derived from the CHV, in our case) to the target domain D_T to help improve the learning of the target-domain task T_T [48]. We followed Pan and Yang [48] to distinguish between inductive transfer learning, where the source- and target-domain tasks are different, and domain adaptation, where the source- and target-domain tasks are the same but the source and target domains (ie, data distributions) are different. Our approach belongs to the first category because our source-domain and target-domain tasks define positive and negative examples in different ways. Transfer learning has been applied to important bioinformatics tasks such as DNA sequence analysis and gene interaction network analysis [49]. It has also been applied to several clinical and biomedical NLP tasks, including part-of-speech tagging [50] and key concept identification for clinical text [51], semantic role labeling for biomedical articles [52] and clinical text [53], and key sentence extraction from biomedical literature [47]. In this work, we investigated 2 state-of-the-art transfer learning algorithms that have shown superior performance in recent studies [47,53]. We aimed to empirically show that they, in combination with distant supervision, are effective in ranking EHR terms.

Methods

Electronic Health Record Corpus and Candidate Terms

We used 7839 discharge summary notes (5.4 million words) from the University of Pittsburgh NLP Repository (using these data requires a license) [54], called EHR-Pittsburgh for convenience, for this study. We applied the linguistic filter of the Java Automatic Term Extraction (JATE) toolkit (version 1.11) [55] to EHR-Pittsburgh to extract candidate terms (see step 1 in Figure 1). JATE's linguistic filter uses a word extractor, a noun phrase extractor, and a stop word list to select high-quality words and noun phrases as candidate terms. We extracted a total of 106,108 candidate terms and further used them to identify and rank medical terms.

Figure 1. Overview of development of the adapted distant supervision (ADS) natural language processing system to rank candidate terms mined from electronic health record (EHR) corpora: data extraction (steps 1 and 2), ADS (step 3), and evaluation (step 4). CHV: consumer health vocabulary.



Consumer Health Vocabulary

CHV was developed by collaborative research to address vocabulary discrepancies between lay people and health care professionals [56-59]. CHV incorporates terms extracted from various consumer health sites, including queries submitted to MedlinePlus, a consumer-oriented online knowledge resource maintained by the US National Library of Medicine [60,61]. CHV contains 152,338 terms, most of which are consumer health terms [60-62]. Zeng et al [60] mapped these terms to UMLS concepts by a semiautomatic approach. As a result of this work, CHV encompasses lay terms (eg, "low blood sugar level" and "heart attack"), as well as corresponding medical terms (eg, "hypoglycemia" and "myocardial infarction"). In this study, we used CHV to create distantly labeled training data for ADS.

Important Terms for Electronic Health Record Comprehension

We defined important terms as those terms that, if understood by the patients, would significantly improve their EHR comprehension. In practice, we used 4 criteria, unithood, termhood, unfamiliarity, and quality of compound term (defined with examples in Multimedia Appendix 2), to judge term importance.

Except for unithood, which is a general criterion for lexical entry selection, the other 3 criteria all measure term importance from the perspective of patient EHR comprehension (details in Multimedia Appendix 2). For example, familiar terms are not important because they are already known by the average patient. High-quality compound terms are those terms whose meanings are beyond the simple sum of their component words (eg, "community-acquired pneumonia"). These terms are important and should be annotated with lay definitions because otherwise patients would not understand them even if they know all the individual words in these terms.

Distant Supervision from Consumer Health Vocabulary

We used CHV to select positive examples to train ADS (see step 2 in Figure 1). Specifically, we assumed that medical terms that occur in both EHRs and CHV (called EHR-CHV terms) are important for patient EHR comprehension. We chose CHV for distant supervision for 3 reasons. First, terms in CHV have been curated and thus all satisfy the unithood criterion. Second, recall that medical terms existing in CHV are synonyms of consumer health terms initially identified from queries and postings generated by patients in online health forums. Therefore, we expect most of these terms to bear clear and significant clinical meanings for patients and thus satisfy the termhood criterion. Third, CHV assigns familiarity scores to 57.89% (88,189 out of 152,338) of its terms for extended usability, which can be used to distinguish between medical terms and lay terms. CHV familiarity scores estimate the likelihood that a term can be understood by an average reader [63] and take values between 0 and 1 (with 1 being most familiar and 0 being least familiar). CHV provides different types of familiarity scores [21]. Following Zeng-Treitler et al [21], we used the combined score and used a heuristic rule (ie, CHV familiarity score ≤0.6) to identify medical terms.

Despite the aforementioned merits, CHV is not perfect in labeling the training data. First, there is not a clear boundary between familiar and unfamiliar terms if their CHV familiarity scores are close to 0.6. For example, "congestive heart failure" and "atypical migraine" have familiarity scores of 0.64 and 0.61; therefore, they would be labeled as negative examples by CHV. However, these 2 terms were judged by domain experts as important terms that need lay definitions. Second, some compound terms in CHV (eg, "knee osteoarthritis," "brain MRI," "aspirin allergy"), although labeled as positive examples by CHV, were judged by domain experts as being not high-quality compound terms from the perspective of efficiently expanding a lay language resource and thus did not need immediate treatment for adding lay definitions.

Transfer Learning Algorithms

Problem Formalization

Since CHV-labeled training data are noisy, we used transfer learning to adapt the system distantly supervised by CHV to the target-domain task. More formally, we defined the training data derived from CHV as the source-domain data $D_S=\{(x^s_1, y^s_1), (x^s_2, y^s_2), \ldots, (x^s_N, y^s_N)\}$ and the target-domain data as $D_T=\{(x^t_1, y^t_1), (x^t_2, y^t_2), \ldots, (x^t_M, y^t_M)\}$, where N is the number of source-domain instances, (x^s, y^s) is the paired feature vector and class label of an instance in the source domain, and M and (x^t, y^t) are defined similarly for the target domain. Notice that we refer to CHV-labeled candidate terms as the source-domain data by following the convention of transfer learning, although these terms were extracted from EHRs. In our study, we used all the N source-domain instances and at most K ($K \ll M$) target-domain instances to train the model. The goal of transfer learning is to make an optimal use of the $N+K$ training data to improve model performance on the M-K target-domain test data.

In this study, we investigated 2 state-of-the-art transfer learning methods: feature space augmentation (FSA) and supervised distant supervision (SDS).

Feature Space Augmentation

FSA [64] has shown the best performance in semantic role labeling for clinical text [53].

This approach assumes that D_S and D_T share the same feature space $X= R^F$ (ie, each feature vector is an F-dimension real-valued vector) and defines an augmented feature space $X^+= R^{3F}$. It then defines 2 feature mapping functions, Φ_S and Φ_T: $X \rightarrow X^+$, by Equation 1 (Figure 2) to respectively map feature vectors from D_S and D_T to the augmented feature space. The motivation is to make the learning easier by separating the general features (ie, the first F dimensions in the augmented feature space, which are useful to learn examples in both D_S and D_T) and the domain-specific features (the second and third F dimensions in the augmented feature space). In addition, it allows a single model to regulate jointly the trade-off between the general and domain-specific feature weights.

Figure 2. Equations for feature mapping functions used in feature space augmentation (1), objective function used in supervised distant supervision (2), and average precision (3).

$$\Phi_S(x)=<x, x, \mathbf{0}>, \Phi_T(x)=<x, \mathbf{0}, x> \quad (1)$$

where $\mathbf{0}$ is a zero vector $\in R^F$; x is a feature vector $\in R^F$; F is the dimension of the original feature space; Φ_S (Φ_T) maps the source-domain (target-domain) feature vector x to the augmented feature space

$$\hat{\epsilon}_\alpha(h) = \sum_{(x,y)\in D_T} \hat{\epsilon}(h(x), y) + \alpha \sum_{(x,y)\in D_S} \hat{p}_T(y|x)\hat{\epsilon}(h(x), y) \quad (2)$$

where α is the hyperparameter to weight the source-domain (D_S) and target-domain (D_T) training data at the corpus level; h is the prediction function learned by the classifier; $\hat{\epsilon}(h(x), y)$ is the instance-level error; and $\hat{p}_T(y|x)$ is the probability of a source-domain label being correct (ie, being consistent with the target-domain label), which is used to weight the source-domain training data at the instance level. We estimated $\hat{p}_T(y|x)$ by a log-linear model learned only from the target-domain training data.

$$Average\ Precision = \sum_{k=1}^{n} P(k)\Delta_r(k) \quad (3)$$

where $P(k)$ is precision of the ranking at rank k, and $\Delta_r(k)$ is the increase of recall of the ranking at rank k compared with the recall at rank k-1

Supervised Distant Supervision

SDS is an extension of the algorithm recently proposed by Wallace et al [47]. It minimizes an objective function that combines empirical source-domain and target-domain errors, as defined in Equation 2 (Figure 2).

Our algorithm differs from that of Wallace et al [47] in that it does not assume that only positive examples in the source domain are unreliable and is therefore more generalizable.

Implementation Issues

We implemented 2 versions of the ADS system, ADS-fsa and ADS-sds, by incorporating the 2 transfer learning algorithms. We used the log-linear model as the base of all the models (including the baseline models introduced in the subsection Baseline Systems) and used L2 regularization for model training. The output from the log-linear models is probabilities of a candidate term being a positive example and can be used to rank candidate terms directly. We used grid search and cross-validation on the target-domain training data to set the hyperparameters α (the corpus weighting parameter in Equation 2; Figure 2) and C (the hyperparameter of the log-linear model

to control the regularization strength; a small *C* corresponds to a strong regularization). In our experiments, we set $\alpha=\beta(K / N)$ (*N* and *K* are the size of the source- and target-domain training data) and searched β in [0.01, 0.1, 1, 10, 100]. We searched *C* in [1,0.1,0.001,0.0001].

Training and Evaluation Datasets

Data Annotation

We derived the training and evaluation datasets from the 106,108 candidate terms extracted from EHR-Pittsburgh as follows.

First, 3 people with a postgraduate level of education in biology, public health, and biomedical informatics reviewed candidate terms among the terms ranked as high by the nonadapted distant supervision model (ie, among the top 10,000 terms) or by the term recognition algorithm C-value [65] (ie, among the top 5000 terms). We chose top-ranked terms, which were likely to contain more important terms than randomly sampled terms, to speed up the whole annotation process. We used the output from 2 methods to increase the diversity of terms used for evaluation and used more terms from the distant supervision model because a manual review suggested that it outperformed C-value. We adopted the expert annotation approach because nonexperts may lack sufficient knowledge to judge the domain relevance and quality of a candidate term, which could potentially introduce noise to the data and slow down the annotation process.

Each term was annotated by 1 primary reviewer and then reviewed by another reviewer based on the 4 criteria introduced in the subsection Important Terms for Electronic Health Record Comprehension (details in Multimedia Appendix 2). Difficult cases were discussed and resolved within the group. Using this procedure, we obtained 6038 annotated terms (3530 positive examples and 2508 negative examples) before starting this study and used all of them for our experiments. To compute the interannotator agreement, 2 reviewers independently annotated 500 candidate terms and achieved a .71 kappa coefficient on this dataset.

Target-Domain Training and Evaluation Sets

We used 1000 examples randomly sampled from the 6038 annotated terms as the target-domain training set and used the remaining 5038 terms as the evaluation set. We did not use stratified sampling because in practice we did not know the class distribution of the target-domain data or the test data. In transfer learning, the target-domain training data are critical to system performance. Therefore, we repeated the above procedure 100 times to obtain 100 pairs of <target training set, evaluation set> for system evaluation to take into account the variance of the target training set. To test the effects of the size of the target-domain training data, we reported system performance by using *L* (*L*=100, 200, 500, 1000) examples randomly selected from the full target training set.

Source-Domain Training Set

We first obtained 100,070 terms by removing the 6038 manually labeled terms from the 106,108 candidate terms. We then automatically labeled the 100,070 terms based on whether a term was an EHR-CHV medical term (ie, positive term) or not (ie, negative term). In this way, we obtained 4166 positive terms and 95,904 negative terms. Because we did not know the distribution of the target-domain data, we randomly sampled 3000 positive and 3000 negative terms from these data to form a balanced source-domain training set. We set the size of the source training set to 6000 by following previous work [66].

Baseline Systems

We employed 2 baselines commonly used to evaluate transfer learning methods [47,53,64]: *SourceOnly* or nonadapted distant supervision model, which was trained by using only source-domain training data, and *TargetOnly*, which was trained by using only target-domain training data.

Features

Word Embedding

Word embedding is the distributed vector representation of words. It has emerged as a powerful technique for word representation and proved beneficial in a variety of biomedical and clinical NLP tasks. We used word2vec software to create the skip-gram word embeddings [67,68] and trained word2vec using a combined text corpus (over 3 billion words) of English Wikipedia, articles from PubMed Central Open Access Subset, and 99,735 EHR notes from the University of Pittsburgh NLP Repository [54]. We set the training parameters by following Jagannatha et al [37] and Pyysalo et al [69]. Specifically, we used 200-dimension vectors with a window size of 6 and used hierarchical soft-max with a subsampling threshold of 0.001 for training. We represented multiword terms (ie, compound terms) by the mean of the vectors of their component words by following Jagannatha et al [37] and Chen and colleagues [38,39].

Semantic Type

We mapped candidate terms to UMLS concepts and included semantic types for those concepts that had an exact match or a head-noun match as features. Each semantic type is a 0-1 binary feature. This type of feature has been used to identify domain-specific medical terms [23,33] and to rank medical terms from individual EHR notes [38].

Automatic Term Recognition

We used the confidence scores from 2 term-recognition algorithms: corpus-level term frequency-inverse document frequency [55] and C-value [65].

General-Domain Term Frequency

We generated 4 features from the Google Ngram corpus [70]: the average, minimum, and maximum frequencies of a term's component words and the term frequency. Corpus frequency has proved to be a strong indicator for term familiarity [63,71]. The Google Ngram corpus is a database of unigram and n-gram counts of words collected from over 15 million books containing over 5 billion pages. We used the top 4.4 million high-frequency words from this corpus and their unigram, bigram, and trigram matches to derive our features.

Term Length

Term length is the number of words in a term. Because a long candidate term may not be a good compound term but rather a simple concatenation of shorter terms (eg, "left heart cardiac catheterization"), this feature may help the ADS system to identify and rank as low the low-quality compound terms.

Evaluation Metrics

Average Precision

This metric averages precision *P(k)* at rank *k* as a function of recall *r*, as defined in Equation 3 (Figure 2).

Area Under the Receiver Operating Characteristic Curve

The area under the receiver operating characteristic curve (AUC-ROC) is computed; this curve plots the true positive rate (y-coordinate) against the false positive rate (x-coordinate) at various threshold settings.

Recall that we have 100 pairs of <target training set, evaluation set> randomly sampled from the 6038 labeled terms. When evaluating a system, we averaged its performance scores on the 100 pairs of datasets and report the averaged values.

We used sklearn.metrics to compute the average precision and AUC-ROC scores. Scikit-learn is an open source Python library widely used for machine learning [72]. In this study, we only reported the paired-samples *t* test results for performance difference between the ADS systems and the baselines because the baselines were expected to be better than a random classifier. The AUC-ROC score of each individual system tested in our experiments was significantly better than 0.5—that is, the AUC-ROC score of a random classifier ($P<.001$).

Statistical Analysis

We used the paired-samples *t* test to test the significance of the performance difference between a pair of systems. We used scipy.stats to conduct the paired *t* test. SciPy is an open source Python library widely used for scientific computing [73].

Results

ADS Ranking Performance on Evaluation Set

Table 1 shows the evaluation results, where the 2 ADS systems outperformed the 2 baselines significantly (t_{99} ranges from 4.84 to 133.31, $P<.001$) for AUC-ROC and average precision under all 4 conditions (ie, using 4 different sizes of target training data). The performance scores of the ADS systems continuously improved with increased size of target training data. When comparing the 2 ADS systems, ADS-fsa performed significantly better than ADS-sds when using 1000 target-domain training examples for transfer learning and performed worse than ADS-sds when using 100 or 200 target-domain training examples (see bottom 2 rows in Table 1 for *t* and *P* values).

Table 1. Performance of different natural language processing systems on the evaluation set under 4 conditions using 100, 200, 500, and 1000 target-domain training examples[a].

System	AUC-ROC[b]				Average precision			
	100	200	500	1000	100	200	500	1000
SourceOnly	0.739	0.739	0.739	0.739	0.811	0.811	0.811	0.811
TargetOnly	0.728	0.749	0.769	0.782	0.799	0.816	0.833	0.844
ADS-fsa[c]	0.746	0.756	*0.776*	*0.790*	0.815	0.823	*0.839*	*0.850*
ADS-sds[d]	*0.751*	*0.759*	0.775	0.786	*0.819*	*0.826*	0.838	0.847
ADS-fsa vs ADS-sds[e]								
t_{99}	4.25	2.79		8.78	3.81	3.04		11.58
P values	<.001	.01		<.001	<.001	.003		<.001

[a]The highest performance scores are italicized.

[b]AUC-ROC: area under the receiver operating characteristic curve.

[c]ADS-fsa: adapted distant supervision-feature space augmentation.

[d]ADS-sds: adapted distant supervision-supervised distant supervision.

[e]The *P* values for difference between ADS-fsa and SourceOnly, ADS-sds and SourceOnly, ADS-fsa and TargetOnly, and ADS-sds and TargetOnly are <.001 (t_{99} ranges from 4.84 to 133.31) for all metrics under all conditions. We report the *P* values (if the *P* value ≤.05) and the corresponding t_{99} values for difference between ADS-fsa and ADS-sds.

The average familiarity level or score of top-ranked terms measures one important aspect of ranking quality. However, because many terms in the evaluation set did not have CHV familiarity scores, we could not compute this value directly. A manual review of the top 500 terms ranked by the best system—that is, ADS-fsa trained using 1000 target-domain training examples—did find many unfamiliar medical terms, including "autoimmune enteropathy," "ileostomy," "myasthenia gravis," "nifedipine," "parathyroid hormone," and "phototherapy."

Effects of Individual Features on ADS Ranking Performance

In addition to evaluating system performance, we tested the contribution of each individual feature to system performance by using feature ablation experiments. Table 2 shows that

ADS-sds's performance dropped significantly (P<.001 for both measures under all 4 conditions) when respectively dropping word embedding, general-domain term frequency, and term length. Dropping the semantic features had mixed results, slightly decreasing performance when the target-domain training set was large and increasing performance when the target-domain training set was small. Dropping features derived from automatic term recognition had no statistically significant effects. The effects of dropping individual features on ADS-fsa's performance were similar (see the first table in Multimedia Appendix 3).

Table 2. Performance of different ADS-sds[a] systems implemented by using all types of features or by dropping each individual type of feature, under 4 conditions using 100, 200, 500, and 1000 target-domain training examples[b].

ADS-sds system	AUC-ROC[c]				Average precision			
	100	200	500	1000	100	200	500	1000
ADS-sds-ALL[d]	0.751	0.759	0.775	0.786	0.819	0.826	0.838	0.847
ADS-sds-woWE[e]	0.711	0.718	0.726	0.733	0.780	0.785	0.793	0.799
ADS-sds-woWE vs ADS-sds-ALL								
t_{99}	30.37	32.74	59.92	112.25	36.61	39.63	81.04	124.15
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
ADS-sds-woSem[f]	0.753	0.760	0.772	0.782	0.823	0.829	0.838	0.845
ADS-sds-woSem vs ADS-sds-ALL								
t_{99}			4.63	12.28	3.18	4.00		4.55
P value			<.001	<.001	.002	<.001		<.001
ADS-sds-woATR[g]	0.751	0.759	0.774	0.786	0.819	0.826	0.838	0.847
ADS-sds-woGTF[h]	0.740	0.749	0.765	0.777	0.813	0.821	0.833	0.842
ADS-sds-woGTF vs ADS-sds-ALL								
t_{99}	13.04	9.50	14.85	22.55	8.12	6.49	11.52	23.07
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
ADS-sds-woTL[i]	0.741	0.751	0.767	0.778	0.807	0.815	0.829	0.838
ADS-sds-woTL vs ADS-sds-ALL								
t_{99}	11.21	10.81	19.78	25.58	16.43	17.15	34.50	41.72
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

[a]ADS-sds: adapted distant supervision-supervised distant supervision.

[b]We report the P values (if the P value ≤.05) and the corresponding t_{99} values for differences between each implementation and ADS-sds-ALL.

[c]AUC-ROC: area under the receiver operating characteristic curve.

[d]ADS-sds-ALL: ADS-sds with all types of features.

[e]ADS-sds-woWE: ADS-sds without word embedding.

[f]ADS-sds-woSem: ADS-sds without semantic features.

[g]ADS-sds-woATR: ADS-sds without features derived from automatic term recognition.

[h]ADS-sds-woGTF: ADS-sds without general-domain term frequency.

[i]ADS-sds-woTL: ADS-sds without term length.

Discussion

Principal Results

In an effort to build a lexical resource that provides lay definitions for medical terms in EHRs, we developed the ADS system to rank candidate terms mined from an EHR corpus and prioritized our efforts to collect and curate lay definitions for top-ranked terms. Given only 100 labeled target training examples, the best ADS system, ADS-sds, achieved 0.751 AUC-ROC and 0.819 average precision on the evaluation set, which are significantly better (P<.001) than the corresponding performance scores of supervised learning (Table 1, ADS-sds vs TargetOnly). When using 1000 target-domain training examples, the best ADS system, ADS-fsa, achieved 0.790 AUC-ROC and 0.850 average precision, also significantly better (P<.001) than that achieved by supervised learning (Table 1, ADS-fsa vs TargetOnly).

Our evaluation set was challenging, because terms included in this set had been prefiltered (ie, ranked as high) by 2

term-ranking methods (details in the Training and Evaluation Datasets subsection). In other words, we evaluated ADS on a set of candidate terms that had higher quality than the average candidate terms mined from EHRs, for which the boundaries between positive and negative examples were more subtle. For example, some candidate terms (eg, "metastatic carcinoid tumor," "normal serum calcium," and "acute cardiac ischemia"), although registered as medical terms in UMLS, were judged nonimportant or nonurgent for lay definition creation because their meanings could be easily inferred from their component words.

The evaluation results on this dataset suggest that our ADS system is effective in ranking EHR terms and can be used to facilitate the expansion of lexical resources that support EHR comprehension. In particular, it can be used to alleviate the data sparseness problem when there are very few target-domain training data and can be used to boost the performance of supervised learning when the size of the training data increases.

Effects of Target-Domain Training Data

Our evaluation results also suggested that using more target-domain training data is beneficial for system performance (rows 2-4 in Table 1). In an additional experiment (details in Multimedia Appendix 4), we found that the performance of ADS-fsa, the best system when using 1000 target training data, continued to improve with increased target training data and began to plateau when the number of target training examples reached 2500.

Effects of Individual Features

The results of our feature ablation experiment (Table 2) indicate that word embedding contributes mostly to system performance, followed by general-domain term frequency and term length. Although dropping semantic features had mixed effects, the results from further analysis indicate that semantic features are useful when excluding word embedding from the feature set. Specifically, adding semantic features on the 3 other types of features (ie, automatic term recognition, general-domain term frequency, and term length) significantly improved system performance (t_{99} ranges from 12.74 to 128.11, P<.001 for 2 measures under 4 conditions; see the second table in Multimedia Appendix 3 for details). This suggests that most information provided by the semantic features for ranking terms is subsumed by that provided by word embedding (but not vice versa). Different from the semantic features, the automatic term recognition features had little additional effect on the performance even without counting word embedding. A likely reason is that our evaluation data set was created by including terms already ranked as high (top 5%) by the automatic term recognition algorithm C-value [65], which may have diminished the effect of this type of feature on this dataset.

Comparing Different Transfer Learning Methods

Although ADS-fsa and ADS-sds were both effective in ranking EHR terms (Table 1), ADS-fsa had small gains over ADS-sds when the size of target training data was large (1000 examples) and vice versa when the size of the target training data was small (100 and 200 examples). The 2 systems used different methods, SDS and FSA, to balance the source- and target-domain training data. Specifically, SDS allows fine-grained weighting of training data from source and target domains at the instance level; FSA, by using an augmented feature space, allows redistribution of feature weights for source, target, and "shared" domains. Our results suggest that instance weighting (ie, ADS-sds) can be more effective when the target-domain training data are very limited.

Error Analysis

We identified three major types of errors through error analysis on the top-rank and low-rank terms (using 300 as the rank threshold) that were ranked by the ADS-sds system that used 1000 target-domain training examples for transfer learning. Error analysis for ADS-fsa showed similar results. First, we found that most errors were caused by compound terms. Specifically, ADS-sds ranked some terms (such as "malignant cell," "chronic rhinitis," and "viral bronchitis") as high, even though their meanings could be easily inferred from their component words. It also ranked certain good compound terms (eg, "community-acquired pneumonia," "end-stage kidney failure," and "left ventricular ejection fraction") as low when these terms contained familiar words. This suggests that advanced features generated by a compound term detector may improve the system's performance, which we may explore in the future. Second, ADS-sds missed certain terms that are lay terms in the general domain but bear unfamiliar clinical meanings (eg, "baseline," "vehicle," and "family history"). Third, ADS-sds ranked some common medical terms (eg, "aspirin," "vitamin," and "nerve") as high, although these terms are likely to be already known by the average patient. The second and third types of errors may be reduced by including domain-specific knowledge about term familiarity as additional features, which we will study in the future.

Conclusion

We report a novel ADS system for ranking and identifying medical terms important for patient EHR comprehension. We empirically show that the ADS system outperforms strong baselines, including supervised learning, and transfer learning can effectively boost its performance even with only 100 target-domain training examples. The EHR terms prioritized by our model have been used to expand a comprehensive lay language lexical resource that supports patient EHR comprehension. The top 10,000 EHR terms ranked by ADS are available upon request.

Acknowledgments

This work was supported by the Institutional National Research Service Award (T32) 5T32HL120823-02 from the US National Institutes of Health (NIH) and the Health Services Research & Development Service of the US Department of Veterans Affairs

Investigator Initiated Research (1I01HX001457-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH, the US Department of Veterans Affairs, or the US Government.

We thank Weisong Liu, Elaine Freund, Emily Druhl, and Victoria Wang for technical support in data collection. We also thank the anonymous reviewers for their constructive comments and suggestions.

Authors' Contributions

HY and JC designed the study. JC and ANJ collected the data. JC designed and developed the ADS system, conducted the experiments, and drafted the manuscript. ANJ contributed substantially to feature generation for ADS. HY and SJF provided important intellectual input into system evaluation and content organization. All authors contributed to the writing and revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Analysis results of consumer health vocabulary's coverage of terms in electronic health record notes.

[PDF File (Adobe PDF File), 334KB - medinform_v5i4e42_app1.pdf]

Multimedia Appendix 2

Criteria used for manual selection of terms important for patient comprehension of electronic health record notes.

[PDF File (Adobe PDF File), 553KB - medinform_v5i4e42_app2.pdf]

Multimedia Appendix 3

Effects of features on performance of adapted distant supervision.

[PDF File (Adobe PDF File), 419KB - medinform_v5i4e42_app3.pdf]

Multimedia Appendix 4

Effects of increasing target-domain training data on system performance.

[PDF File (Adobe PDF File), 431KB - medinform_v5i4e42_app4.pdf]

References

1. Steinbrook R. Health care and the American Recovery and Reinvestment Act. N Engl J Med 2009 Mar 12;360(11):1057-1060. [doi: 10.1056/NEJMp0900665] [Medline: 19224738]
2. Wright A, Feblowitz J, Samal L, McCoy AB, Sittig DF. The Medicare Electronic Health Record Incentive Program: provider performance on core and menu measures. Health Serv Res 2014 Feb;49(1 Pt 2):325-346 [FREE Full text] [doi: 10.1111/1475-6773.12134] [Medline: 24359554]
3. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
4. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
5. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to electronic health records: results from a survey of veterans. Med Care 2013 Mar;51(3 Suppl 1):S52-S56. [doi: 10.1097/MLR.0b013e31827808db] [Medline: 23407012]
6. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
7. Pyper C, Amery J, Watson M, Crook C. Patients' experiences when accessing their on-line electronic patient records in primary care. Br J Gen Pract 2004 Jan;54(498):38-43 [FREE Full text] [Medline: 14965405]
8. Keselman A, Slaughter L, Smith CA, Kim H, Divita G, Browne A, et al. Towards consumer-friendly PHRs: patients' experience with reviewing their health records. AMIA Annu Symp Proc 2007:399-403 [FREE Full text] [Medline: 18693866]
9. Chapman K, Abraham C, Jenkins V, Fallowfield L. Lay understanding of terms used in cancer consultations. Psychooncology 2003 Sep;12(6):557-566. [doi: 10.1002/pon.673] [Medline: 12923796]

10. Lerner EB, Jehle DV, Janicke DM, Moscati RM. Medical communication: do our patients understand? Am J Emerg Med 2000 Nov;18(7):764-766. [doi: 10.1053/ajem.2000.18040] [Medline: 11103725]
11. Jones RB, McGhee SM, McGhee D. Patient on-line access to medical records in general practice. Health Bull (Edinb) 1992 Mar;50(2):143-150. [Medline: 1517087]
12. Baldry M, Cheal C, Fisher B, Gillett M, Huet V. Giving patients their own records in general practice: experience of patients and staff. Br Med J (Clin Res Ed) 1986 Mar 1;292(6520):596-598 [FREE Full text] [Medline: 3081187]
13. Kutner M, Greenberg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 National Assessment of Adult Literacy (NCES 2006-483). Washington, DC: Department of Education, National Center for Educational Statistics (NCES); 2006 Sep. URL: http://nces.ed.gov/pubs2006/2006483.pdf [accessed 2016-07-15] [WebCite Cache ID 6lwUr7mOK]
14. 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]
15. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
16. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
17. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
18. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
19. McCray AT. Promoting health literacy. J Am Med Inform Assoc 2005;12(2):152-163 [FREE Full text] [doi: 10.1197/jamia.M1687] [Medline: 15561782]
20. 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]
21. Zeng-Treitler Q, Goryachev S, Kim H, Keselman A, Rosendale D. Making texts in electronic health records comprehensible to consumers: a prototype translator. AMIA Annu Symp Proc 2007:846-850 [FREE Full text] [Medline: 18693956]
22. Kandula S, Curtis D, Zeng-Treitler Q. A semantic and syntactic text simplification tool for health content. AMIA Annu Symp Proc 2010;2010:366-370 [FREE Full text] [Medline: 21347002]
23. Polepalli RB, Houston T, Brandt C, Fang H, Yu H. Improving patients' electronic health record comprehension with NoteAid. Stud Health Technol Inform 2013;192:714-718 [FREE Full text] [Medline: 23920650]
24. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
25. Connolly KK, Crosby ME. Examining e-Health literacy and the digital divide in an underserved population in Hawai'i. Hawaii J Med Public Health 2014 Feb;73(2):44-48 [FREE Full text] [Medline: 24567867]
26. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
27. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
28. Graber MA, Roller CM, Kaeble B. Readability levels of patient education material on the World Wide Web. J Fam Pract 1999 Jan;48(1):58-61. [Medline: 9934385]
29. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA 2001;285(20):2612-2621 [FREE Full text] [Medline: 11368735]
30. Kasabwala K, Agarwal N, Hansberry DR, Baredes S, Eloy JA. Readability assessment of patient education materials from the American Academy of Otolaryngology--Head and Neck Surgery Foundation. Otolaryngol Head Neck Surg 2012 Sep;147(3):466-471 [FREE Full text] [doi: 10.1177/0194599812442783] [Medline: 22473833]
31. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]
32. Zeng QT, Tse T, Divita G, Keselman A, Crowell J, Browne AC, et al. Term identification methods for consumer health vocabulary development. J Med Internet Res 2007 Feb 28;9(1):e4 [FREE Full text] [doi: 10.2196/jmir.9.1.e4] [Medline: 17478413]
33. Spasić I, Schober D, Sansone S, Rebholz-Schuhmann D, Kell DB, Paton NW. Facilitating the development of controlled vocabularies for metabolomics technologies with text mining. BMC Bioinformatics 2008 Apr 29;9 Suppl 5:S5 [FREE Full text] [doi: 10.1186/1471-2105-9-S5-S5] [Medline: 18460187]

34. Doing-Harris KM, Zeng-Treitler Q. Computer-assisted update of a consumer health vocabulary through mining of social network data. J Med Internet Res 2011 May 17;13(2):e37 [FREE Full text] [doi: 10.2196/jmir.1636] [Medline: 21586386]
35. Ahltorp M, Skeppstedt M, Kitajima S, Henriksson A, Rzepka R, Araki K. Expansion of medical vocabularies using distributional semantics on Japanese patient blogs. J Biomed Semantics 2016 Sep 26;7(1):58 [FREE Full text] [doi: 10.1186/s13326-016-0093-x] [Medline: 27671202]
36. Henriksson A, Conway M, Duneld M, Chapman WW. Identifying synonymy between SNOMED clinical terms of varying length using distributional analysis of electronic health records. AMIA Annu Symp Proc 2013;2013:600-609 [FREE Full text] [Medline: 24551362]
37. Jagannatha A, Chen J, Yu H. Mining and ranking biomedical synonym candidates from Wikipedia. 2015 Presented at: Sixth International Workshop on Health Text Mining and Information Analysis (Louhi); Sep 17-21, 2015; Lisbon, Portugal p. 142-151 URL: https://aclweb.org/anthology/W/W15/W15-2619.pdf
38. Chen J, Zheng J, Yu H. Finding important terms for patients in their electronic health records: a learning-to-rank approach using expert annotations. JMIR Med Inform 2016 Nov 30;4(4):e40 [FREE Full text] [doi: 10.2196/medinform.6373] [Medline: 27903489]
39. Chen J, Yu H. Unsupervised ensemble ranking of terms in electronic health record notes based on their importance to patients. J Biomed Inform 2017 Apr;68:121-131. [doi: 10.1016/j.jbi.2017.02.016] [Medline: 28267590]
40. Craven M, Kumlien J. Constructing biological knowledge bases by extracting information from text sources. Proc Int Conf Intell Syst Mol Biol 1999:77-86. [Medline: 10786289]
41. Morgan AA, Hirschman L, Colosimo M, Yeh AS, Colombe JB. Gene name identification and normalization using a model organism database. J Biomed Inform 2004 Dec;37(6):396-410 [FREE Full text] [doi: 10.1016/j.jbi.2004.08.010] [Medline: 15542014]
42. Mintz M, Bills S, Snow R, Jurafsky D. Distant supervision for relation extraction without labeled data. 2009 Presented at: Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2; Aug 2-7, 2009; Suntec, Singapore p. 1003-1011 URL: https://www.aclweb.org/anthology/P09-1113
43. Usami Y, Cho H, Okazaki N, Tsujii J. Automatic acquisition of huge training data for bio-medical named entity recognition. 2011 Presented at: 2011 Workshop on Biomedical Natural Language Processing; Jun 23-24, 2011; Portland, OR, USA p. 65-73 URL: http://www.aclweb.org/anthology/W11-0208
44. Buyko E, Beisswanger E, Hahn U. The extraction of pharmacogenetic and pharmacogenomic relations--a case study using PharmGKB. Pac Symp Biocomput 2012:376-387 [FREE Full text] [Medline: 22174293]
45. Bobic T, Klinger R, Thomas P, Hofmann-Apitius M. Improving distantly supervised extraction of drug-drugprotein-protein interactions. 2012 Presented at: Joint Workshop on UnsupervisedSemi-Supervised Learning in NLP; Apr 23-27, 2012; Avignon, France p. 35-43 URL: http://www.aclweb.org/anthology/W12-0705
46. Liu M, Ling Y, An Y, Hu X. Relation extraction from biomedical literature with minimal supervision and grouping strategy. 2014 Presented at: 2014 IEEE International Conference on Bioinformatics and Biomedicine; Nov 2-5, 2014; Belfast, UK p. 444-449. [doi: 10.1109/BIBM.2014.6999198]
47. Wallace BC, Kuiper J, Sharma A, Zhu MB, Marshall IJ. Extracting PICO sentences from clinical trial reports using supervised distant supervision. J Mach Learn Res 2016;17 [FREE Full text] [Medline: 27746703]
48. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010 Oct;22(10):1345-1359 [FREE Full text] [doi: 10.1109/TKDE.2009.191]
49. Xu Q, Yang Q. A survey of transfer and multitask learning in bioinformatics. J Comput Sci Eng 2011;5(3):257-268. [doi: 10.5626/JCSE.2011.5.3.257]
50. Ferraro JP, Daumé H, Duvall SL, Chapman WW, Harkema H, Haug PJ. Improving performance of natural language processing part-of-speech tagging on clinical narratives through domain adaptation. J Am Med Inform Assoc 2013;20(5):931-939 [FREE Full text] [doi: 10.1136/amiajnl-2012-001453] [Medline: 23486109]
51. Zheng J, Yu H. Identifying key concepts from EHR notes using domain adaptation. 2015 Presented at: Sixth International Workshop on Health Text Mining and Information Analysis (Louhi); Sep 17-21, 2015; Lisbon, Portugal p. 115-119 URL: https://aclweb.org/anthology/W/W15/W15-2615.pdf
52. Dahlmeier D, Ng HT. Domain adaptation for semantic role labeling in the biomedical domain. Bioinformatics 2010 Apr 15;26(8):1098-1104 [FREE Full text] [doi: 10.1093/bioinformatics/btq075] [Medline: 20179074]
53. Zhang Y, Tang B, Jiang M, Wang J, Xu H. Domain adaptation for semantic role labeling of clinical text. J Am Med Inform Assoc 2015 Sep;22(5):967-979 [FREE Full text] [doi: 10.1093/jamia/ocu048] [Medline: 26063745]
54. Mowery D, Wiebe J, Visweswaran S, Harkema H, Chapman WW. Building an automated SOAP classifier for emergency department reports. J Biomed Inform 2012 Feb;45(1):71-81 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.020] [Medline: 21925286]
55. Zhang Z, Iria J, Brewster C, Ciravegna F. A comparative evaluation of term recognition algorithms. 2008 Presented at: LREC 2008: Sixth International Conference on Language Resources Evaluation; May 26-Jun 1, 2008; Marrakech, Morocco p. 2108-2113 URL: http://www.lrec-conf.org/proceedings/lrec2008/pdf/538_paper.pdf

56. McCray AT, Loane RF, Browne AC, Bangalore AK. Terminology issues in user access to Web-based medical information. Proc AMIA Symp 1999:107-111 [FREE Full text] [Medline: 10566330]
57. Zeng Q, Kogan S, Ash N, Greenes RA. Patient and clinician vocabulary: how different are they? Stud Health Technol Inform 2001;84(Pt 1):399-403 [FREE Full text] [Medline: 11604772]
58. Patrick TB, Monga HK, Sievert ME, Houston HJ, Longo DR. Evaluation of controlled vocabulary resources for development of a consumer entry vocabulary for diabetes. J Med Internet Res 2001;3(3):e24 [FREE Full text] [doi: 10.2196/jmir.3.3.e24] [Medline: 11720966]
59. Zeng Q, Kogan S, Ash N, Greenes RA, Boxwala AA. Characteristics of consumer terminology for health information retrieval. Methods Inf Med 2002;41(4):289-298. [Medline: 12425240]
60. Zeng QT, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying consumer-friendly display (CFD) names for health concepts. AMIA Annu Symp Proc 2005:859-863 [FREE Full text] [Medline: 16779162]
61. Keselman A, Smith CA, Divita G, Kim H, Browne AC, Leroy G, et al. Consumer health concepts that do not map to the UMLS: where do they fit? J Am Med Inform Assoc 2008;15(4):496-505 [FREE Full text] [doi: 10.1197/jamia.M2599] [Medline: 18436906]
62. Tse T, Soergel D. Exploring medical expressions used by consumers and the media: an emerging view of consumer health vocabularies. AMIA Annu Symp Proc 2003:674-678 [FREE Full text] [Medline: 14728258]
63. Zeng Q, Kim E, Crowell J, Tse T. A text corpora-based estimation of the familiarity of health terminology. 2005 Presented at: Sixth International Conference on Biological and Medical Data Analysis; Nov 10-11, 2005; Aveiro, Portugal p. 184-192 URL: https://lhncbc.nlm.nih.gov/files/archive/pub2005041.pdf [doi: 10.1007/11573067_19]
64. Daume H. Frustratingly easy domain adaption. 2004 Presented at: 45th Annual Meeting of the Association of Computational Linguistics; Jun 23-30, 2007; Prague, Czech Republic p. 256-263 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.2062&rep=rep1&type=pdf
65. Frantzi K, Ananiadou S, Mima H. Automatic recognition of multi-word terms: the C-value/NC-value method. Int J Digit Libr 2000 Aug 1;3(2):115-130 [FREE Full text] [doi: 10.1007/s007999900023]
66. Jiang J, Zhai C. Instance weighting for domain adaption in NLP. 2007 Presented at: 45th Annual Meeting of the Association of Computational Linguistics; Jun 23-30, 2007; Prague, Czech Republic p. 264-271 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.8018&rep=rep1&type=pdf
67. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: Advances in Neural Information Processing Systems (NIPS 2013); Dec 5-10, 2013; Lake Tahoe, NV, USA p. 3111-3119 URL: https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf
68. Mikolov T, Chen K, Corrado G, Dean J. ArXiv13013781 Cs. 2013 Sep 7. Efficient estimation of word representations in vector space URL: http://arxiv.org/abs/1301.3781 [accessed 2017-06-18] [WebCite Cache ID 6m6NhZqFz]
69. Pyysalo S, Ginter F, Moen H, Salakoski T, Ananiadou S. Distributional semantics resources for biomedical text processing. 2013 Presented at: The 5th International Symposium on Languages in Biology and Medicine (LBM 2013); December 12-13, 2013; Tokyo, Japan p. 39-43 URL: http://bio.nlplab.org/pdf/pyysalo13literature.pdf
70. Michel J, Shen YK, Aiden AP, Veres A, Gray MK, Google Books Team, et al. Quantitative analysis of culture using millions of digitized books. Science 2011 Jan 14;331(6014):176-182 [FREE Full text] [doi: 10.1126/science.1199644] [Medline: 21163965]
71. Elhadad N. Comprehending technical texts: predicting and defining unfamiliar terms. AMIA Annu Symp Proc 2006:239-243 [FREE Full text] [Medline: 17238339]
72. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12(Oct):2825-2830 [FREE Full text]
73. Jones E, Oliphant T, Peterson P. SciPy: Open source scientific tools for Python. 2001-. SciPy.org.: SciPy developers; 2017. URL: https://www.scipy.org/ [accessed 2017-10-20] [WebCite Cache ID 6uMcmi5ru]

Abbreviations

ADS: adapted distant supervision
AUC-ROC: area under the receiver operating characteristic curve
CHV: consumer health vocabulary
EHR: electronic health record
FSA: feature space augmentation
JATE: Java Automatic Term Extraction
NLP: natural language processing
SDS: supervised distant supervision
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 22.07.17; peer-reviewed by L Cui, J Luo, T Abdulai, C Fincham, I Mircheva; comments to author 06.09.17; revised version received 19.09.17; accepted 20.09.17; published 31.10.17.

Please cite as:
Chen J, Jagannatha AN, Fodeh SJ, Yu H
Ranking Medical Terms to Support Expansion of Lay Language Resources for Patient Comprehension of Electronic Health Record Notes: Adapted Distant Supervision Approach
JMIR Med Inform 2017;5(4):e42
URL: http://medinform.jmir.org/2017/4/e42/
doi:10.2196/medinform.8531
PMID:29089288

Review

Open-Source Electronic Health Record Systems for Low-Resource Settings: Systematic Review

Assel Syzdykova[1,2*], MSc; André Malta[1*], MSc; Maria Zolfo[3], PhD; Ermias Diro[4], PhD; José Luis Oliveira[1], PhD

[1]University of Aveiro, Department of Electronics, Telecommunications and Informatics (DETI) / Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Aveiro, Portugal

[2]BMD Software, Aveiro, Portugal

[3]Institute of Tropical Medicine, Antwerp, Belgium

[4]University of Gondar, Gondar, Ethiopia

[*]these authors contributed equally

Corresponding Author:
José Luis Oliveira, PhD
University of Aveiro
Department of Electronics, Telecommunications and Informatics (DETI) / Institute of Electronics and Informatics Engineering of Aveiro (IEETA)
Campus Universitario de Santiago
Aveiro, 3810 193
Portugal
Phone: 351 234370523
Email: jlo@ua.pt

Abstract

Background: Despite the great impact of information and communication technologies on clinical practice and on the quality of health services, this trend has been almost exclusive to developed countries, whereas countries with poor resources suffer from many economic and social issues that have hindered the real benefits of electronic health (eHealth) tools. As a component of eHealth systems, electronic health records (EHRs) play a fundamental role in patient management and effective medical care services. Thus, the adoption of EHRs in regions with a lack of infrastructure, untrained staff, and ill-equipped health care providers is an important task. However, the main barrier to adopting EHR software in low- and middle-income countries is the cost of its purchase and maintenance, which highlights the open-source approach as a good solution for these underserved areas.

Objective: The aim of this study was to conduct a systematic review of open-source EHR systems based on the requirements and limitations of low-resource settings.

Methods: First, we reviewed existing literature on the comparison of available open-source solutions. In close collaboration with the University of Gondar Hospital, Ethiopia, we identified common limitations in poor resource environments and also the main requirements that EHRs should support. Then, we extensively evaluated the current open-source EHR solutions, discussing their strengths and weaknesses, and their appropriateness to fulfill a predefined set of features relevant for low-resource settings.

Results: The evaluation methodology allowed assessment of several key aspects of available solutions that are as follows: (1) integrated applications, (2) configurable reports, (3) custom reports, (4) custom forms, (5) interoperability, (6) coding systems, (7) authentication methods, (8) patient portal, (9) access control model, (10) cryptographic features, (11) flexible data model, (12) offline support, (13) native client, (14) Web client,(15) other clients, (16) code-based language, (17) development activity, (18) modularity, (19) user interface, (20) community support, and (21) customization. The quality of each feature is discussed for each of the evaluated solutions and a final comparison is presented.

Conclusions: There is a clear demand for open-source, reliable, and flexible EHR systems in low-resource settings. In this study, we have evaluated and compared five open-source EHR systems following a multidimensional methodology that can provide informed recommendations to other implementers, developers, and health care professionals. We hope that the results of this comparison can guide decision making when needing to adopt, install, and maintain an open-source EHR solution in low-resource settings.

(JMIR Med Inform 2017;5(4):e44) doi:10.2196/medinform.8131

KEYWORDS

electronic health record; EHR; software; eHealth; open source

Introduction

Electronic Health Records

The world has been moving toward global health and equity in accessing health care services for many decades [1-3]. However, the majority of low- and middle-income countries (LMICs) still face challenges in providing comprehensive medical care compared with the developed countries. Many factors influence health care services in low-resource settings such as the lack of financing, absence of health care policy, and limited technical and human resources [4-9]. As a result, the adoption of eHealth in poorly resourced areas such as LMICs is a challenging task because of their limitations and regional specificities. To bring positive changes to public medical care in less developed economies, the World Health Organization recommends the adoption of electronic health (eHealth) solutions starting from national eHealth strategy [10] and electronic health records (EHRs) [11].

As information and communication technologies have rapidly become involved in health care, particularly in patient management, this has increased the efficiency and effectiveness of services provided by medical care facilities [12-14]. The EHRs era began in the middle of the 1960s when health care professionals changed the direction of patient management toward digital format. EHR systems not only increase accuracy and reduce mistakes through allergy alerts, access to laboratory data, and immunization history but also improve organizational and societal outcomes [13]. Moreover, the collected patients' data can be used in research [15-19], giving an opportunity to study diseases and extract knowledge from clinical data. Despite the advantages of EHRs, the majority of developing countries cannot afford expensive proprietary software from big vendors and demand cost-effective solutions. Thus, the adoption of open-source EHR (OS EHR) is a possible solution for LMICs. Moreover, it is important to consider other accompanying expenses such as power supply, Internet connection, staff training, and system support.

One of the pioneers of OS EHR systems was the Veterans Information Systems and Technology Architecture (VistA) developed by the US Department of Veterans Affairs implemented in the late 1970s [20]. The system is still functioning and serves several million patients in US hospitals. Almost four decades after the first OS EHR system was introduced in developed countries, it has reached the developing world too [21].

eHealth in Low-Resource Settings

In several studies conducted on eHealth adoption in LMICs, most were focused on EHR implementation [22-32]. Aminpour et al [33] pointed out that open-source EHRs have been widely used by resource-limited regions in all continents, particularly in Sub-Saharan Africa and South America. The authors concluded that HIS create opportunities to improve national health care, particularly in developing countries with minimal financial resources.

The major obstacles to OS EHR utilization in low-resource settings were discussed by Fraser and Blaya [5] after their experience in developing countries. The authors shared lessons they learned, highlighting key ideas such as the need for clinical staff involvement to provide qualitative data entry, the reuse of validated EHR implementations, and planning of offline data entry, which is particularly relevant because of an unstable power supply and Internet connection.

In studying the success criteria of OS EHRs for resource-limited settings, Fritz et al [24] concluded that a system's functionality and technical infrastructure play an essential role, whereas financing is not a major aspect because of donors' sponsorship in developing countries. The authors also mentioned the importance of organizational facets such as training availability, project management, and staff involvement.

Other studies focused on the regional distribution of OS EHR in developing countries, which included Africa [34-36], Latin America [37], and Asia [38,39]. Several studies focused on applying health record systems to concrete diseases such as human immunodeficiency virus (HIV) [40,41], tuberculosis [42], and diabetes [43]. For instance, Manders et al reported on their experience in the adaptation of an open-source health record system in a hospital in Mozambique.

The study by Tilahun et al [44] in Ethiopia was based on interviewing health care professionals in 5 hospitals. The findings showed clinical staff's dissatisfaction, for example, with low service quality and entering data twice (both on paper and electronic medical records [EMR]). The authors suggest focusing on training, user support, and providing enough computers to increase the use of EMR systems by health care professionals.

Open-Source Solutions

In a recent comparison of OS EHR systems [45-47], the authors used a multicriteria selection approach to compare and score the tools. The authors examined 13 chosen systems and compared 25 of their features.

One of the biggest open-source software (OSS) communities in health care is the Open Medical Record System (OpenMRS) led by Regenstrief Institute and Partners in Health; it focuses on the development of an electronic medical record solution to be used in resource-constrained environments. Initially, the system was implemented in Kenya and then was rapidly adopted by other health care organizations [35,48-58]. Starting from a simple data model, currently, OpenMRS has become an EHR platform with developers all over the world [59], contributing to its Java-based service-oriented architecture [60].

Bahmni is an integrated clinical software, which combines three open-source platforms: (1) OpenMRS for patient records, (2) OpenELIS (OpenELIS Foundation) for laboratory management, and (3) OpenERP (Odoo SA) for hospitals' accounting

operations. Bahmni was implemented by the ThoughtWorks company in 2012. The main goal of the software is to provide a universal solution for health care institutions. Bahmni has been deployed mainly in low-resource settings such as India, Bangladesh, and Nepal [61]. However, the number of implementations is growing and spreading to other regions [62].

GNU Health (GNU Solidario) is another OS EHR system that provided hospitals with a wide range of functionalities for patient management. The system was introduced in 2008 and has been adopted in several developing countries such as Paraguay, Kenya, the Philippines, and Bangladesh [63-66].

Open Source Clinical Application Resource (OSCAR) is a nonproprietary EHR system, developed by McMaster University, Canada. It is licensed under the GLv2, allowing other programmers to reuse the system code by acknowledging copyright and giving access to their modifications under the same license agreement. Unlike many EHR tools, OSCAR has been implemented mainly by doctors rather than developers. However, it has become widespread among health care providers [67].

OpenVistA (Medsphere Systems Corporation) is another OS EHR system based on the VistA software [55], one of the first EHR systems implemented for American veterans. OpenVistA allows multiple clinicians to simultaneously access various patient data in real time. The system provides progress notes, various templates, ordering and reporting tools, audit capabilities, electronic signature, document management, and data integration tools, among other features of OpenVistA.

FreeMED (FreeMED Software Foundation) is a modular open-source system for medical data and was implemented almost two decades ago in the United States. It also has an external billing subsystem called REMITT.

OpenClinic GA (Medical eXchange Solutions) is an open-source system that includes EHR, laboratory management, and pharmacy management. It was implemented in 2006 in Belgium and has been used in many countries worldwide. The system has been used in several resource-limited settings, including hospitals in Rwanda, Mali, and Burundi.

Overall, there are many open-source projects available for health care providers. However, we selected five OS EHR systems to make detailed research of each and focus on the functionalities provided, which are pointed out in the next section.

Methods

Identification of Key Features

In this section, we reviewed OS EHR systems discussing their strengths and weaknesses according to a set of predefined parameters. The evaluation methodology was defined in close collaboration with the University of Gondar Hospital, Ethiopia, and it included two main aspects: (1) user requirements and (2) systems' selection.

The selection of the most important features of these systems was performed through interactive discussion sessions with experienced physicians from Gondar and partners from the EUROLEISH-NET research network, with great experience in low-resource settings, both in Asia and in Africa. In this phase, the health care professionals' help was crucial because of the specificity of medical workflows and specialized vocabularies. Moreover, we analyzed previous publications related to eHealth development in LMICs, where the authors also tackled the specificity of the infrastructure and services. From this study, we obtained the following set of features:

1. Integrated applications referring to the type of subsystems that are integrated in each solution
2. Configurable reports, general report templates may be generated by an end user, without programming skills
3. Custom reports, specific report templates can be created by a user with basic programming skills
4. Custom forms allow making a form template to gather patient data or other clinical information
5. Interoperability, ability to exchange data with other applications (it includes the support for standards such as Health Level-7 [HL7], Fast Healthcare Interoperability Resources [FHIR], and Digital Imaging and Communications in Medicine [DICOM])
6. Coding systems, referring to the utilization of medical terminology classification, such as SNOMED (Systematized Nomenclature of Medicine) or International Classification of Diseases (ICD)-10
7. Authentication methods enable users to access systems without credentials by using the authentication from third-party application (eg, single sign-on, lightweight directory access protocol [LDAP])
8. Patient portal can be accessed by patients to consult their own records
9. Access control model, management of users' access, roles, and permissions
10. Cryptographic features refer to applying security techniques to conceal patient data
11. Flexible data model can cope with various and dynamic data characteristics
12. Offline support, still operating on the client side, when disconnected from the server
13. Web client, the application can be accessed through a common Web browser
14. Native client, has a client that runs natively on a specific desktop platform (eg, in Windows or Linux)
15. Other clients refer to more than one client version to cope with hardware limitations
16. Code-based language, programming language used for system's source code (eg, Java)
17. Development activity refers to how often the software is updated and to the developers' support
18. Software modularity refers to software engineering quality, namely, how manageable the system is
19. User interface refers to the usability of the system and how the user interface is perceived by end users
20. Community support includes discussion forum activities and community contributions such as translation of user interface to other languages
21. Customization, how easy modifications are without rewriting the core code of a system to reach established requirements

We ranked several features such as modularity, user interface, community support, and custom forms to show our evaluation of the EHR systems. The modularity rank, for instance, highlights whether the platform architecture is well thought out to simplify future modifications and additions from developers. Community support is another important aspect of open-source software that helps volunteer developers and IT teams to improve and share their contributions.

Software Selection Process

The selection of open-source EHR software for this review was based on a representative set of open-source EHR systems found in the literature, as well as popular and active projects in open-source repositories. The methodology, which is depicted in Figure 1, included the following steps: (1) literature, (2) code repositories, and (3) features filtering.

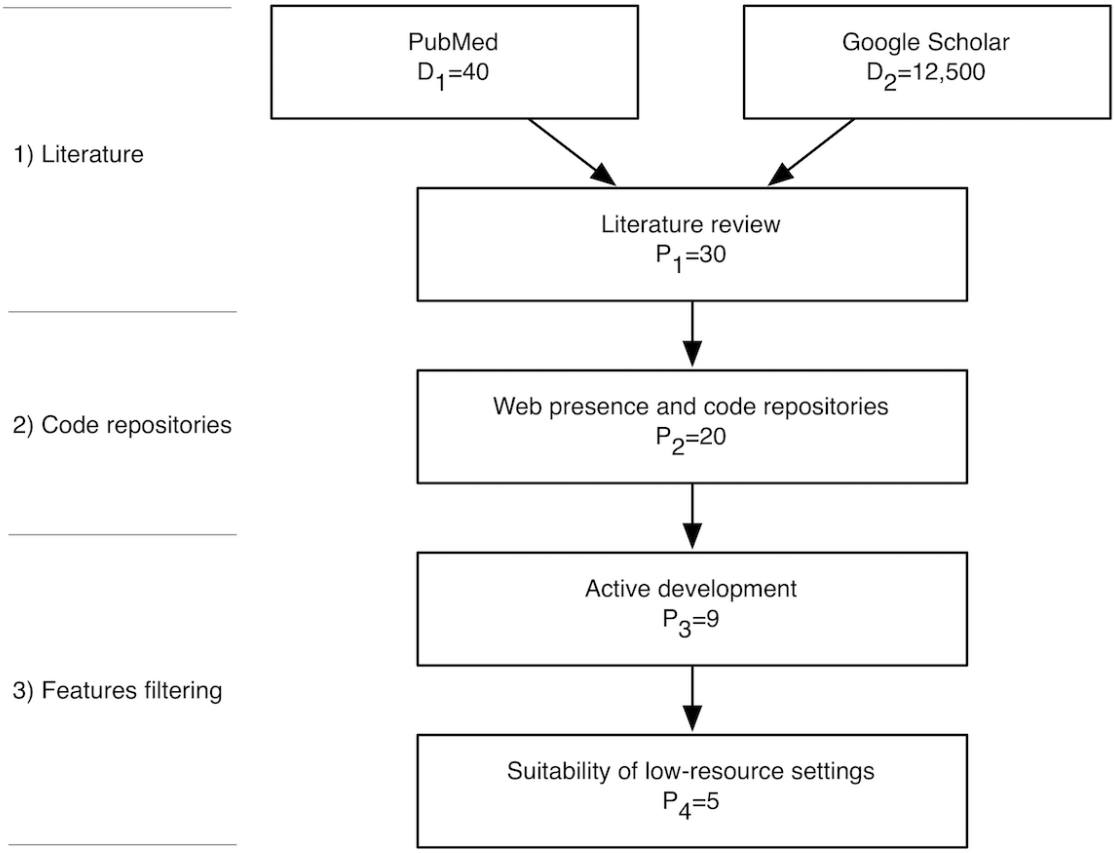
In the first step, we searched on PubMed for papers referring to OS EHR software using the terms "open source" and "EMR or EHR," combined with the MeSH term "computer software." We validated the content of the retrieved publications leading to a total of 40 documents (D_1=40). We also queried Google Scholar using an advanced search composed by the term "software" combined with the exact expression "open source" and containing at least one of the terms "electronic medical record, electronic health record, EHR, and EMR." From the returned results (D_2=12,500), we evaluated the top 100 documents to verify its relevance to our study. From this assessment, we identified 30 potential solutions (P_1=30).

In the second phase, a common Web search engine (Google) was used, mainly for obtaining information about EHR solutions, for example, project status and resources. For each, we gathered the availability of public websites (24) and public source code repositories (23), from which we ruled out the ones outside the intersection of these (P_2=20).

The last was the filtering step. First, we excluded the projects with lower activity (code updates, discussion forums, case studies, etc), which resulted in a more reduced set (P_3=9). For each of these solutions, we assessed their capability to fulfil the limitations of a low-resource setting (eg, localization support and cross-platform support), ending up with the following group (P_4=5): GNU Health, OpenEMR, FreeMED, OpenMRS, and Bahmni.

The selected systems were then installed and extensively evaluated against the requirements previously discussed. In the following sections, the characteristics of these five systems are discussed and compared.

Figure 1. Flowchart of the open source electronic health record selection process.



Results

GNU Health

GNU Health is an EHR software and health information system (HIS) developed by GNU Solidario, a nonprofit and nongovernmental organization for health and education development. The main areas of GNU Health supports are (1) individual management, (2) patient management, (3) health center management, and (4) information management.

Individual management refers to a person's basic information, genealogy, household, domiciliary infrastructure, and so on. The GNU Health project emphasizes on the importance of social and economic development, with a strong focus on relevant factors such as level of education, occupation, living conditions, and family relations.

Patient management comprises the collection of different types of data related to a patient's health, for instance, lifestyle parameters such as diet and exercise, addictions, sexuality, or safety. It also includes management of the patient's health care regarding encounters and evaluations, medical procedures, laboratory test requests or results, and so on.

Health center management mainly comprises enterprise resource planning (ERP), for instance, financial and human resource management, sales, invoicing and accounting, inventory, stock, and supply chain management. It also features a focus on pharmacy and laboratory management. The managed subject can either be the institution as a whole or individual department.

Information management covers the tasks that combine data produced on the platform. This includes building reports or presenting demographics with configurable parameters, which are useful for epidemiology studies or crossing different data, for instance.

The areas explained above generally involve teams from different fields of activity in which members generally only work within their corresponding domain of information.

As the information stored by GNU Health is highly sensitive, the platform is developed with a focus on data confidentiality and integrity. It provides serialization, hashing, signing, and verification, which are commonly used in features that handle or produce sensitive information such as reports, prescriptions, or laboratory tests. It allows doctors to digitally sign death certificates or users to verify the authenticity of reports compiled in other GNU Health systems. This is achieved via a bundled client-side plug-in for pretty good privacy that uses GNU Privacy Guard implementation.

Most of GNU Health's functionality is provided by the official add-on modules, which extend the core module. These are separated into general domains or specialties, for example, pediatrics, surgery, or genecology. Modules add new functionality to these domains by reusing models of given entities and creating specification classes designed in other modules. As this creates dependencies between modules, in order for a target module to work, its parent modules need to be installed as well. In GNU Health, the way to provide new functionality such as custom forms for collecting patient information is through developing new modules. As this requires programming knowledge, building and adding custom forms becomes a difficult task for nondevelopers.

GNU Health user interface contains a global menu where functionalities are presented in a tree-style structure (Figure 2). The menu supports filtering by text, providing easier and quicker ways to find the desired functionality. Several items can be active at the same time, making use of tabs for arranging the multiple content panels. It is common for some features to present a multitude of functionalities separated into their own tabs, adding to navigation and display complexity.

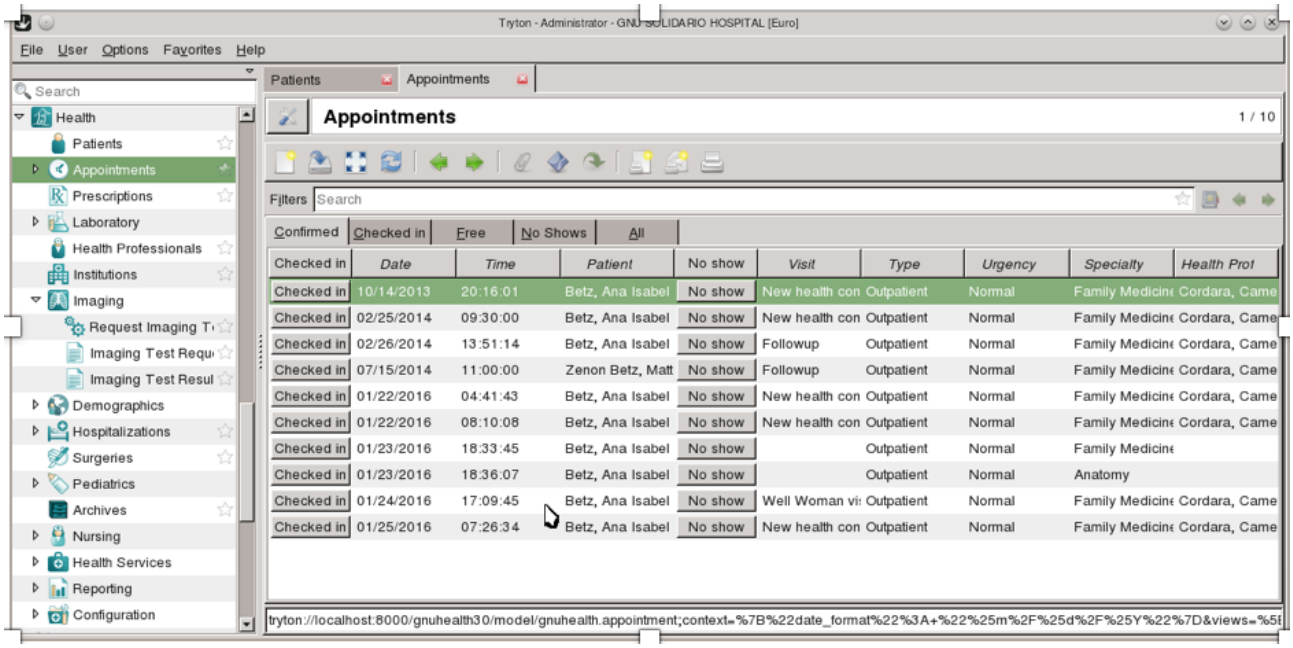
GNU Health client does not provide offline support, that is, interactions made in the client while the server is disconnected are lost. There is a workaround by deploying a GNU Health server on each client machine, synchronized in a centralized mode with the main server. Replicating sensitive data in each client device would raise complexity, security, and privacy issues and would also require different and more expensive hardware than that needed for running just a GNU Health thin client.

OpenEMR

OpenEMR is an EHR software, medical practice management software (PMS), and ERP software mainly supported by OEMR 4, a nonprofit organization (NPO) whose goal is to ensure that all people have access to high-quality medical care through its software and services [68]. OpenEMR is licensed under GNU general public license (GPL) and comes with a great amount of functionality to cover a wide range of areas of an HIS in addition to patient records and management, including support for billing and claims management, ERP, and information management. The system is written in PHP, a scripting language that is widely supported and has cross-platform runtimes. The use of Web technologies makes the platform accessible through many different devices, provided they have a compatible Web browser.

With respect to patient records, clinical observations can be made through a variety of forms already provided. Observations captured on a patient's different visits can be presented later through graphical charts or tables showing changes in values over time. Free-text type notes can be attached to a patient's encounter and medical issues registered in a flexible system of coding, which supports Current Procedural Terminology (CPT), Healthcare Common Procedure Coding System (HCPCS), ICD-9, ICD-10, and SNOMED Clinical Terms (SNOMED CT) codes. In the prescription management area, doctors can register prescriptions and their frequency, which can then be printed, faxed, or sent electronically to a few supported third-party platforms. This can be complemented with the pharmacy dispensary module, which enables in-house drug dispensing with support for inventory, integration with drug databases, and stock tracking. In patients' history, medical procedures, immunizations, and laboratory tests made can be registered along with results.

Figure 2. User interface of GNU Health's client application.



OpenEMR provides a robust calendar subsystem, which is capable of handling different types of events, such as appointments, notifications, alerts, and automated report generation. With this subsystem, it is possible, for example, to automatically send an email reminder to a patient each time he needs to take his medication, or automatically run reports of weight assessment for children and adolescents each month.

OpenEMR provides a native portal aimed at use by patients and other registered users, supporting integration with other external patient portals. The reporting subsystem comes with an extensible profile of reports that cover many of the platform's different areas. Some of these are configurable, for instance, patient reports can include medical history, encounters, billing, communications, or other patient-related information. Adding custom reports to the OpenEMR platform is possible through the same method as the one used by the preloaded reports. These are built with PHP programming for which a specific PHP template for coding reports is provided. There is no alternative method to create custom reports; consequently, programming knowledge and other advanced computer skills are required for the creation of new reports.

For interoperability, OpenEMR supports the HL-7 standard, which is used for exchanging patient data between other systems, and the ANSI X12 Electronic Data Interchange standard, mainly used in billing and invoicing in the ERP.

The OpenEMR platform can support more than 20 languages and is compatible with Unicode Transformation Format (UTF)-8. The display language of the user interface is user dependent, meaning that multiple users can use the platform simultaneously in their desired languages.

User interaction with OpenEMR is via a Web interface (Figure 3). The user interface arranges active content in a dual panel mode. The first panel is always shown and is used to display primary content of the active area or context. The second panel either shows a given global functionality of the active context such as past encounters and documents while browsing the patient's context or is used to show selected content in addition to the first panel. This proves useful for showing related information side by side but quickly builds up complexity in the interface. Some functionality also opts to open a separate frame inside the content panel, or even a new pop-up window to display or gather information, again adding complexity.

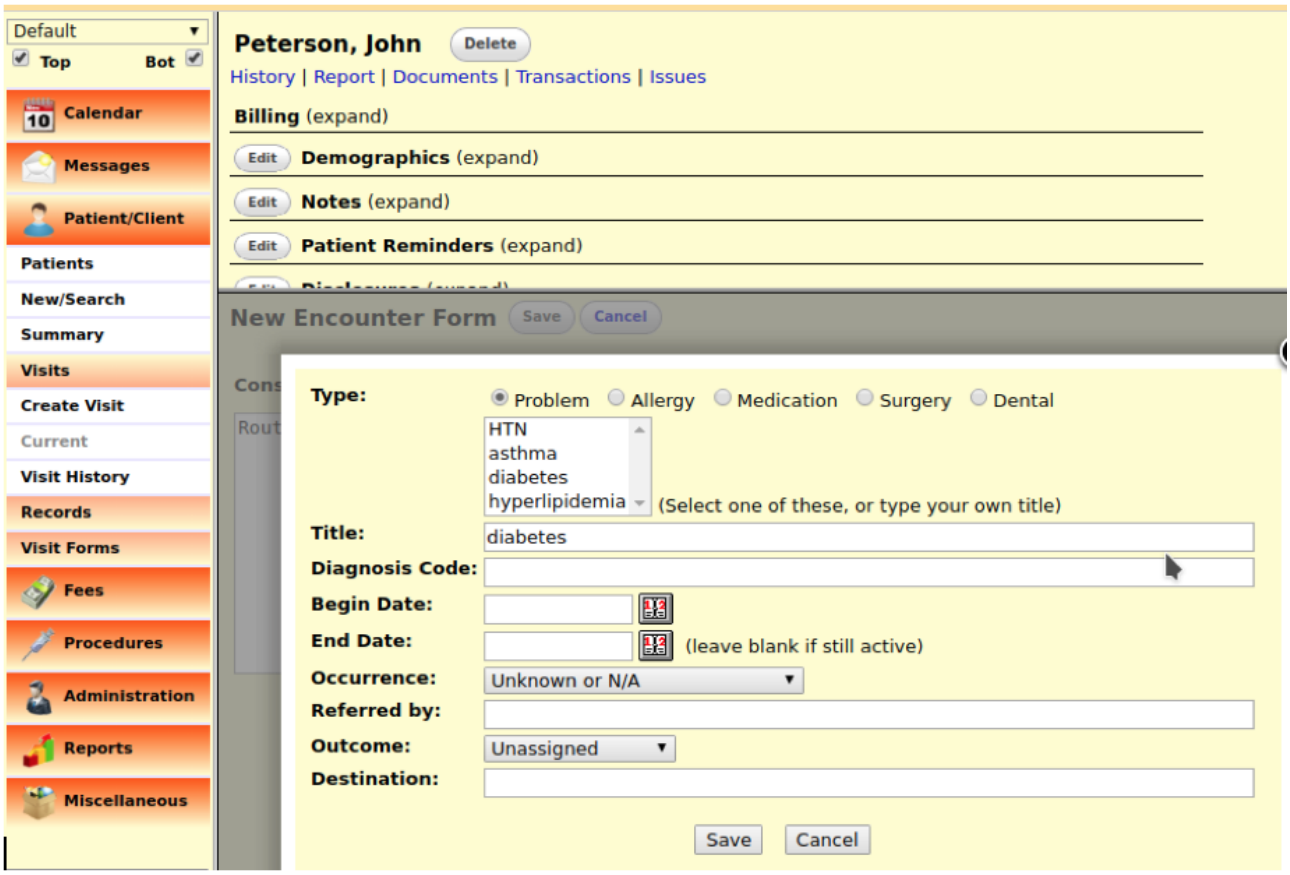
The OpenEMR project development is active, with an average of two major releases per year. The project has a strong community of users and backing companies, with many OpenEMR modules developed by third parties and contributed to OpenEMR, and these are now part of the project. The documentation has extensive user and implementation guides but falls short for developers.

FreeMED

FreeMED is an EHR software and medical PMS by FreeMED Software Foundation 7 and is licensed under GPL. FreeMED has a modular architecture that is designed with functionality and interface separation, which allows customization or development of modules without having to rewrite core components of the system.

A patient record module contains personal and demographic information. Preloaded templates are provided, which can be used for collecting clinical observations during patient visits. Medical assistance is provided by the diagnosis family module and can be coded using CPT and ICD codes. It is possible to assign medication to a patient and manage a registry with them. Patients can be arranged in groups that are assigned to doctors or referenced in programs. There is an area for triage and call-ins useful for registration secretaries, which can be complemented by the appointment scheduling subsystem. Appointments can be scheduled for either an individual or groups of patients, and the interface has safeguards for double booking, presenting warnings in case 2 patients get booked for the same provider at the same time.

Figure 3. User interface of the OpenEMR platform.



FreeMED uses HL-7 for the exchange of medical data. This support is obtained via integration with Mirth Connect, a cross-platform HL-7 interface engine.

FreeMED already comes with some configured reports, but adding new ones requires related technical skills and also knowledge of the FreeMED data model. The reporting subsystem is dependent on the Java library Jasper Reports for report generation. There is a lack of documentation on how to add custom forms to FreeMED. User interaction with the platform is through a Web interface. After language selection and user authentication, different functionalities are displayed depending on the user's access clearance. Access is controlled with Access Control List (ACL), in the same way as described previously for OpenEMR, also making use of the phpGACL library.

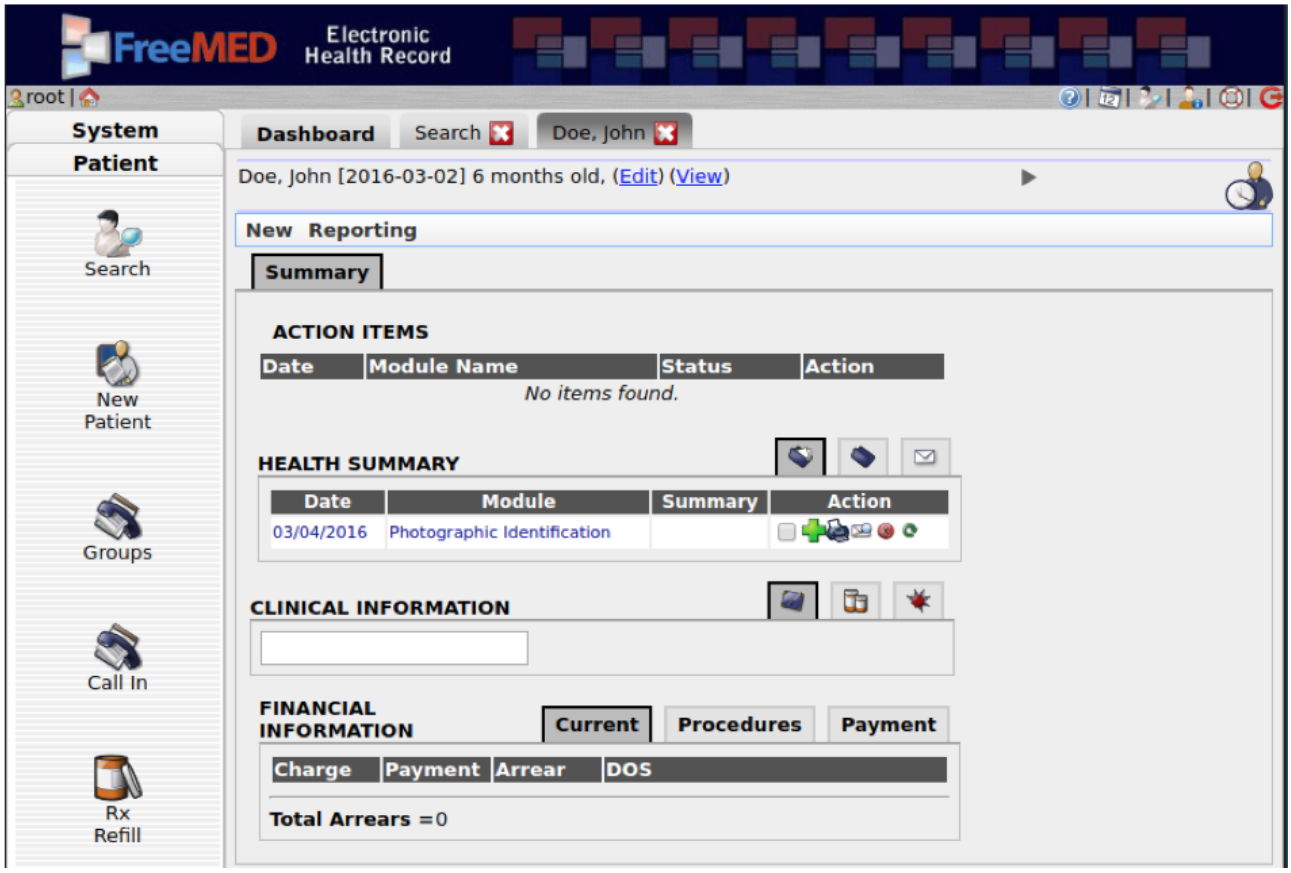
The user interface has an always-present menu, which groups functionalities in categories (Figure 4). The active functionality is displayed on a content panel. Several items can be opened at the same time, being grouped into tabs that keep its current state in the background without interrupting the user experience. The content has a clean presentation, and while some more complex features may show different functionalities grouped into their own tabs, there is an option to choose the arrangement between tabbed and flat presentation.

FreeMED is written mostly in PHP and Perl. Additionally, it requires a Java runtime environment for the report subsystem and JavaEE for REMITT, as they are written in Java. As all the technologies used in the software are cross-platform, FreeMED can be run on different operating systems, ranging from POSIX-compatible such as Linux or BSD to Microsoft Windows. The software is designed in a client-server application model. The platform does not provide offline support, so if the communication is interrupted between the client and the server, users need to wait for the reestablishment of the connection. Akin to OpenEMR, access to the database is through the ADOdb PHP library, which allows transparent usage of many different database management systems (DBMS). Nevertheless, parts of the code base are not DBMS-agnostic, and thus, the project recommends using MySQL.

FreeMED supports localization, and the project claims multiple languages support in addition to English, including German, French, and Polish. There is also an online public project to contribute translations, which has already attained complete localization in Portuguese and Russian [69]. FreeMED supports multiple ISO character set formats and stores entry data with it, making it possible to maintain a demographic database in one language and ISO set, independent of the localization that displays.

The source code is available on a public repository website, which also provides a brief installation guide and a Virtual Machine image of FreeMED for demo purposes. There are no development guides, and the project lacks proper documentation overall. There is also a FreeMED demo available online, but there are no user guides on the official site or in the FreeMED software itself.

Figure 4. User interface of the FreeMED platform.



OpenMRS

OpenMRS is an EHR software led by two collaborative NPOs, Regenstrief Institute and Partners in Health. One of the main objectives of the project is to build a robust solution for health care with a special focus on resource-constrained environments. OpenMRS is a highly modular platform, meaning that features are external to the platform and added through add-on modules. An application programming interface (API) plays a major role in this by allowing access to OpenMRS platform functionality. The platform core uses a centralized dictionary of concepts, a customizable part of OpenMRS that strongly defines each implementation. In the dictionary, concepts are metadata of given entities, defined by name, data type, appropriate attributes, and relationships to other concepts. These are used as models to be instanced later in some user form or report. For example, a value collected during a patient encounter is an instance of a given concept that takes the form of an observation. As a model, concepts work as the building blocks that describe forms, orders, clinical summaries, reports, and so on.

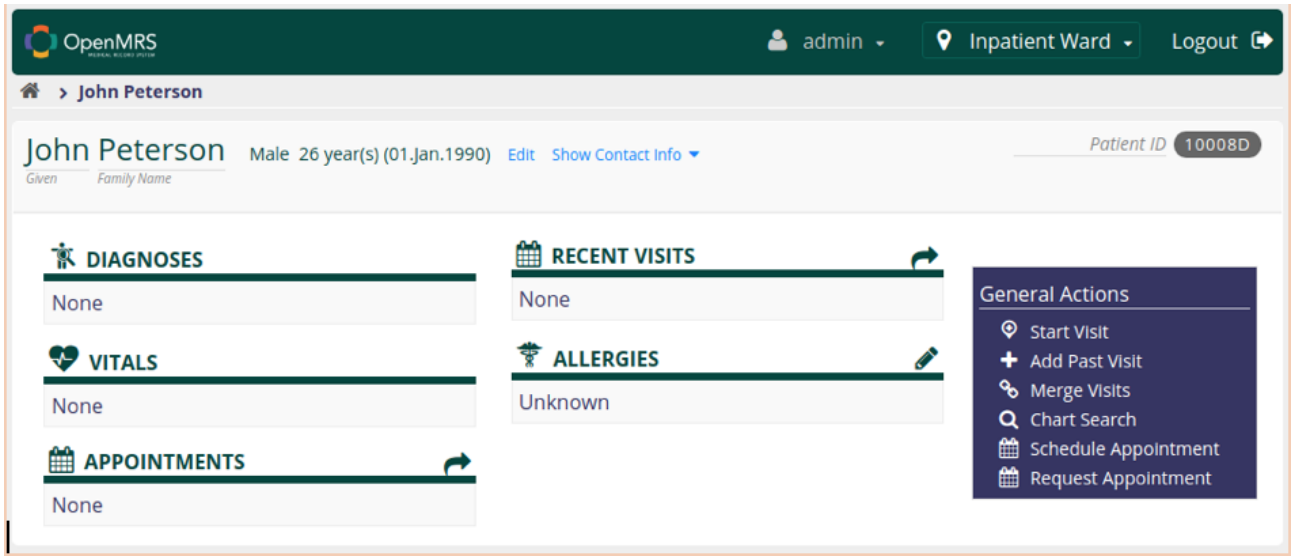
OpenMRS is officially released in two different distributions, one which contains just the OpenMRS platform, and another called the reference application, which packs selected modules that extend and add features to the platform. The framework standardizes modules and defines their interface presentation, communication, behavior, and security, enabling the creation of an OpenMRS application ecosystem. Through modules, OpenMRS extends its features, providing appointment scheduling and management, an event system for notifications, a subsystem for allergy registration, representation of patient observations through charts, and so on (Figure 5).

OpenMRS has a robust subsystem for the creation of custom forms. Any form relies on a schema that defines the fields and concepts it will use, which are supported by the concept dictionary. Although custom forms can be created and their form schema can be designed from within OpenMRS legacy user interface, the methods for designing the forms are independent of the platform.

A reporting subsystem allows three ways of building reports, that is, indicator reports, row-per-patient reports, or custom reports. A report is built through associations with dataset definitions that act as a key to values to include in the report. The dataset can generally refer to person or patient properties and similar attributes that contain only one recorded value stored since its last entry, commonly used in row-per-person reports. The combination of values from multiple observations is possible and is used for building indicator reports. A report definition can be programmed to filter (cohort) a defined set of patients for use as the input for report generation.

Access management follows a role-based access control (RBAC), which covers both the user interface and access to OpenMRS API via REST services. Importing and exporting metadata between platforms can be performed either manually or automatically, using the metadata sharing module. The module is advanced enough to provide automatic synchronization mechanisms, with support for record merging and conflict resolution.

Figure 5. OpenMRS' modern user interface.



OpenMRS has native support for the HL-7 standard and can export data directly to other applications. There is also support for DICOM and radiology standards via a third-party module, in conjunction with dcm4chee, a medical imaging archive and manager. Native support for FHIR is also present natively in the platform from version 2.0.

Officially, the project provides an extensive *wiki* with features content for developers, implementers, and end users. Two online demos are provided, one of the reference application and another with a demo for management of drug-resistant tuberculosis, which can show the level of customization that OpenMRS offers. Currently, more than 175 separate modules are available, but their compatibility depends on the core version because of changes in the OpenMRS API.

OpenMRS has full UTF-8 support and some extra localizations in addition to English, with the possibility to add more. The project provides an online portal for collaborative translation of its user interface, reference application, and commonly used third-party modules [57].

Bahmni

Bahmni is a hospital system for low-resource settings developed by ThoughtWorks Global Health under an AGPL license. Bahmni is built through separate OpenMRS modules, and its user interface communicates with these using the OpenMRS API via REST. The system incorporates most of the OpenMRS platform functionality and extends it with support for picture archiving and communication system (PACS), laboratory information system (LIS), and ERP features. This is done via integration with three separate applications: dcm4chee, which provides the first; OpenELIS, which provides the second; and OpenERP (currently named Odoo), which provides the third.

In a registration process, information such as identity, demographics, photograph, contacts, and other socioeconomic details of a patient can be captured and associated with a person via a newly generated patient identifier. Different types of relationships can be registered, including genealogical or patient-provider relationships. When a patient registers with an institution, there is a strong possibility that he already has health-related documents stored in other institutions, for example, record scans, x-ray images, or other types. Bahmni EHR supports storing these within the system and attaching them to a patient's profile. A patient's profile is displayed through a modular dashboard, where customizable widgets display the most recent information from several different categories of the patient's medical records. These can be, for instance, laboratory tests and results, programs enrolled in, or last visit observations. From observations made in several encounters, the history of changes can be presented.

Clinical services are centered around the patient's profile. These services become available in the consultation area of the dashboard after a patient is active in the system, that is, the patient is in a given encounter with a provider. From here, several types of services are provided: capturing observations, making laboratory or radiology orders, registering diagnosis and medication, managing patient disposition, or taking consultation notes (Figure 6).

With respect to clinical observations, Bahmni already provides multiple forms by default. The form subsystem is very customizable and can easily be extended with more forms. These can be built by creating and changing specific concepts of the dictionary. The form's layout and other parameters can be defined in JSON and, if necessary, there is support for integration with a server-side framework for advanced customization of a form's logic and other aspects. Current and past diagnosis can be associated with a patient and registered along with attributes such as order, confidence, and status. Diagnosis made can also be mapped to ICD-10 codes.

Regarding laboratory orders, Bahmni provides a panel of tests divided in several categories. The laboratory subsystem is flexible, making it possible to choose just the needed tests and categories or arrange them all together. New laboratory tests can be easily created and added just by using the concepts dictionary.

Radiology orders containing patient and investigation details can be requested within Bahmni EHR, which are sent to a modality via HL-7. The results are integrated with the patient

profile even if they come as native unsupported formats, in which references to them are shown. For instance, results that come as DICOM images are archived in an accessible PACS and referenced from the patient profile from which they can be opened in compatible DICOM viewers.

Medication management includes registration of past and current medication from a customizable drug registry, which has the route, frequency, dosage, and instructions for each. There is a drug dispensing service, which is integrated with OpenERP and supports in-house dispensing with segregation of stocks. Patient disposition refers to a patient's state usually set at the end of a provider encounter.

OpenERP is used predominantly for sales and purchase management, including inventory and accounting, and it is integrated to some extent in Bahmni EHR. For instance, registration fees or payment of drugs dispensed is registered in OpenERP under patient billing and inventory.

The reporting area comes with many preloaded reports, which can be run within a given period and exported to CSV, PDF, HTML, or Excel formats. Additionally, the platform provides integration with Jasper Reports, giving support for custom, flexible, and more varied reports.

The access control on Bahmni EHR is provided by the OpenMRS platform. However, there are new access roles associated with the new functionalities and services of Bahmni EHR. Additionally, user access can now be set to use two-factor authentication.

Bahmni EHR supports offline network alerts by displaying a notification in the user interface when the connection is down. This is helpful to avoid loss of filled-in data by submitting it unknowingly when a connection to the server is down. True offline support is also provided via a browser extension or a mobile app, offering the possibility of filling in forms and other user interface functionalities without an active connection to the server. The data produced with this functionality are saved locally on the device and can be uploaded later. The community is closely tied with the OpenMRS community as both come under the same forum and are very active.

Figure 6. Bahmni patient dashboard.

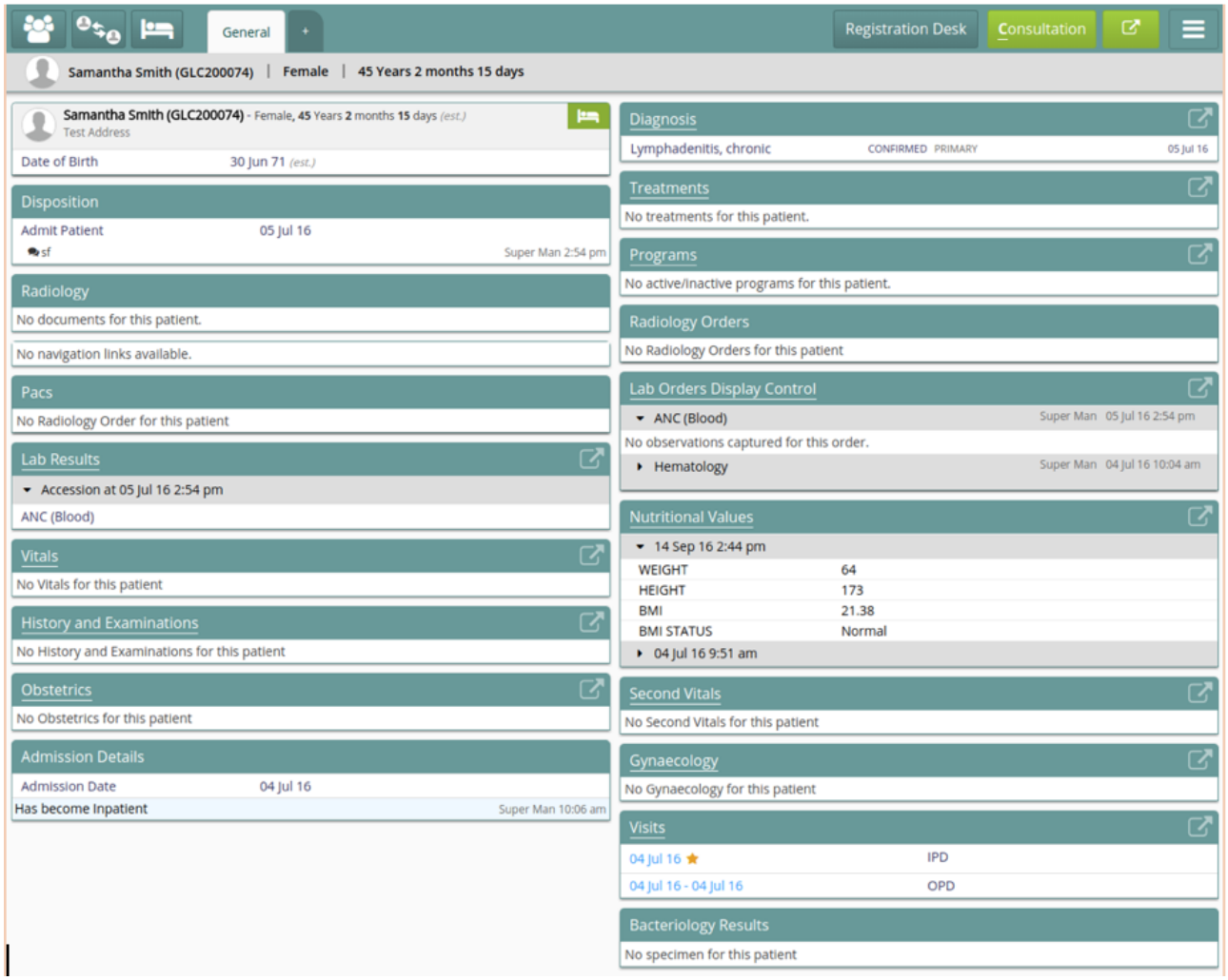


Table 1. Matrix with several aspects of the evaluated electronic health records software. Some features were evaluated according to a ranking that varies between 1 (low) and 3 (high).

Feature or system		GNU Health	OpenEMR	FreeMED	OpenMRS	Bahmni
1	Integrated applications	EHR[a], HIS[b]	EHR, PMS[c], ERP[d]	EHR, PMS	EHR	EHR, PMS, ERP, LIS[e], PACS[f]
2	Configurable reports	Yes	Yes	No	Yes	Yes
3	Custom reports	No	No	No	Yes	Yes
4	Custom forms	-	1	-	3	3
5	Interoperability	FHIR[g], custom	HL7[h]	HL7, DICOM[i]	HL7, DICOM, FHIR	HL7, DICOM, FHIR
6	Coding systems	ICD-10[j]	ICD-9/10, SNOMED[k], CPT[l], HCPCS[m]	ICD-10, CPT, LOINC[n], ATC[o]	CIEL/MVP[p], LOINC ICD-10	CIEL/MVP, ICD-10, SNOMED
7	Authentication methods	LDAP[q]	LDAP, AD[r]	-	-	-
8	Patient portal	No	Yes	No	No	No
9	Access control model	RBAC[s]	ACL[t]	ACL	RBAC	RBAC
10	Cryptographic features	Sign, encrypt	Encrypt	-	-	-
11	Flexible data model	No	No	No	Yes	Yes
12	Offline support	Yes	No	No	No	Yes
13	Web client	Yes	Yes	Yes	Yes	Yes
14	Native client	Yes	No	No	No	No
15	Other clients	Yes	No	No	No	Yes
16	Code-based language	Python	PHP	PHP	Java	Java
17	Development activity	3	3	2	3	3
18	Software modularity	3	1	2	3	3
19	User interface	2	1	3	2	3
20	Community support	3	3	1	3	3
21	Customization	1	2	1	3	3

[a]EHR: electronic health record.

[b]HIS: health information system.

[c]PMS: practice management software.

[d]ERP: enterprise resource planning.

[e]LIS: laboratory information system.

[f]PACS: picture archiving and communication system.

[g]FHIR: Fast Healthcare Interoperability Resources.

[h]HL7: Health Level-7.

[i]DICOM: Digital Imaging and Communications in Medicine.

[j]ICD: International Classification of Diseases.

[k]SNOMED: Systematized Nomenclature of Medicine.

[l]CPT: Current Procedural Terminology.

[m]HCPCS: Healthcare Common Procedure Coding System.

[n]LOINC: Logical Observation Identifiers Names and Codes.

[o]ATC: Anatomical Therapeutic Chemical.

[p]CIEL/MVP: Columbia International eHealth Laboratory/Millennium Villages Project concept dictionary.

[q]LDAP: lightweight directory access protocol.

[r]AD: Active Directory.

[s]RBAC: role-based access control.

[t]ACL: Access Control List.

Discussion

Systems' Comparison

The above assessment allowed us to undertake a systematic review of five popular EHRs that can be used in low-resource settings without being an economic burden. Due to constraints in LMIC and according to the workflow of health care institutions, it is important to consider key operation functionalities and the adaptability of each system. Table 1 summarizes these key aspects, presenting for each system the characteristics defined in our evaluation methodology.

Conclusions

Despite the wide adoption of EHR and eHealth solutions in many developed countries, the scenario is quite unbalanced when compared with health digitalization in low-resource settings. This work was performed in close collaboration with the Hospital of Gondar, Ethiopia, with which the evaluation methodology was defined. We conducted a local assessment of the used solutions and identified a scenario that is very dependent on IT solutions from nongovernmental organizations, which, despite their good will, led to partial and incompatible solutions. There is a clear demand for open-source, reliable, and flexible EHR systems in these health care facilities. Hence, the process of selecting a suitable solution that covers their needs is a central task. In this study, we have evaluated and compared five open-source EHR systems following a multidimensional methodology. We hope that the results of this comparison can guide decision making when needing to adopt, install, and maintain an open-source EHR solution in low-resource settings.

Acknowledgments

This study has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 642609.

Conflicts of Interest

None declared.

References

1. Who.int. 1978 Sep 12. Declaration of Alma-Ata URL: http://www.who.int/publications/almaata_declaration_en.pdf [accessed 2017-11-03] [WebCite Cache ID 6ugtPfy0M]
2. UN.: United Nations; 2010. News on millennium development goals URL: http://www.un.org/millenniumgoals/ [accessed 2017-11-01] [WebCite Cache ID 6ueDSb4Rc]
3. Sustainabledevelopment.un.: Citeseer Sustainable development goals URL: https://sustainabledevelopment.un.org/?menu=1300 [accessed 2017-11-01] [WebCite Cache ID 6ueDfhmcc]
4. Akhlaq A, Sheikh A, Pagliari C. Barriers and facilitators to health information exchange in low- and middle-income country settings: a systematic review protocol. J Innov Health Inform 2015 Mar 27;22(2):284-292. [doi: 10.14236/jhi.v22i2.98]
5. Fraser HS, Blaya J. Implementing medical information systems in developing countries, what works and what doesn't. AMIA Annu Symp Proc 2010;2010:232-236.
6. Chen W, Akay M. Developing EMRs in developing countries. IEEE Trans Inf Technol Biomed 2011 Jan;15(1):62-65. [doi: 10.1109/TITB.2010.2091509]
7. Oluoch T, Santas X, Kwaro D, Were M, Biondich P, Bailey C, et al. The effect of electronic medical record-based clinical decision support on HIV care in resource-constrained settings: a systematic review. Int J Med Inform 2012 Oct;81(10):e83-e92. [doi: 10.1016/j.ijmedinf.2012.07.010] [Medline: 22921485]
8. Luna D, Almerares A, Mayan 3rd JC, González Bernaldo de Quirós F, Otero C. Health informatics in developing countries: going beyond pilot practices to sustainable implementations: a review of the current challenges. Healthc Inform Res 2014 Jan;20(1):3-10 [FREE Full text] [doi: 10.4258/hir.2014.20.1.3] [Medline: 24627813]
9. Luna D, Otero C, Marcelo A. Health informatics in developing countries: systematic review of reviews. Contribution of the IMIA Working Group Health Informatics for Development. Yearb Med Inform 2013;8:28-33. [Medline: 23974545]
10. World Health Organization, International Telecommunication Union. Who.int. 2012. National eHealth strategy toolkit URL: http://www.who.int/ehealth/publications/overview.pdf [accessed 2017-11-03] [WebCite Cache ID 6uguBN7pY]
11. World Health Organization. Wpro.who.int/. 2006. Electronic health records: manual for developing countries URL: http://www.wpro.who.int/publications/docs/EHRmanual.pdf [accessed 2017-11-03] [WebCite Cache ID 6uguIzmXd]
12. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 11;144(10):742-752. [doi: 10.7326/0003-4819-144-10-200605160-00125]
13. Alsaffar M, Yellowlees P, Odor A, Hogarth M. The state of open source electronic health record projects: a software anthropology study. JMIR Med Inform 2017 Feb 24;5(1):e6 [FREE Full text] [doi: 10.2196/medinform.5783] [Medline: 28235750]
14. Reynolds CJ, Wyatt JC. Open source, open standards, and health care information systems. J Med Internet Res 2011 Feb;13(1):e24 [FREE Full text] [doi: 10.2196/jmir.1521] [Medline: 21447469]

15. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy 2011;4:47-55 [FREE Full text] [doi: 10.2147/RMHP.S12985] [Medline: 22312227]
16. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]
17. Lobach DF, Detmer DE. Research challenges for electronic health records. Am J Prev Med 2007 May;32(5 Suppl):S104-S111. [doi: 10.1016/j.amepre.2007.01.018] [Medline: 17466814]
18. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 2;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
19. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, et al. Electronic health records: new opportunities for clinical research. J Intern Med 2013 Dec;274(6):547-560. [doi: 10.1111/joim.12119] [Medline: 23952476]
20. McDonald CJ, Schadow G, Barnes M, Dexter PR, Overhage JM, Mamlin BW, et al. Open source software in medical informatics—why, how and what. Int J Med Inform 2003 Mar;69(2-3):175-184. [doi: 10.1016/S1386-5056(02)00104-1]
21. Fraser H, Biondich P, Moodley D, Choi S, Mamlin B, Szolovits P. Implementing electronic medical record systems in developing countries. J Innov Health Inform 2005 Jun;13(2):83-95. [doi: 10.14236/jhi.v13i2.585]
22. Jawhari B, Ludwick D, Keenan L, Zakus D, Hayward R. Benefits and challenges of EMR implementations in low resource settings: a state-of-the-art review. BMC Med Inform Decis Mak 2016 Sep 6;16:116 [FREE Full text] [doi: 10.1186/s12911-016-0354-8] [Medline: 27600269]
23. Jawhari B, Keenan L, Zakus D, Ludwick D, Isaac A, Saleh A, et al. Barriers and facilitators to electronic medical record (EMR) use in an urban slum. Int J Med Inform 2016 Oct;94:246-254 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.07.015] [Medline: 27573333]
24. Fritz F, Tilahun B, Dugas M. Success criteria for electronic medical record implementations in low-resource settings: a systematic review. J Am Med Inform Assoc 2015 Mar;22(2):479-488. [doi: 10.1093/jamia/ocu038] [Medline: 25769683]
25. Blaya JA, Fraser HS, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251 [FREE Full text] [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
26. Piette JD, Lun KC, Moura Jr LA, Fraser HS, Mechael PN, Powell J, et al. Impacts of e-health on the outcomes of care in low- and middle-income countries: where do we go from here? Bull World Health Organ 2012 May 1;90(5):365-372 [FREE Full text] [doi: 10.2471/BLT.11.099069] [Medline: 22589570]
27. Millard PS, Bru J, Berger CA. Open-source point-of-care electronic medical records for use in resource-limited settings: systematic review and questionnaire surveys. Br Med J Open 2012 Jul;2(4):e000690 [FREE Full text] [doi: 10.1136/bmjopen-2011-000690] [Medline: 22763661]
28. Kalogriopoulos NA, Baran J, Nimunkar AJ, Webster JG. Electronic medical record systems for developing countries: review. Conf Proc IEEE Eng Med Biol Soc 2009;2009:1730-1733. [doi: 10.1109/IEMBS.2009.5333561]
29. Clifford GD, Blaya JA, Hall-Clifford R, Fraser HS. Medical information systems: a foundation for healthcare technologies in developing countries. Biomed Eng Online 2008 Jun 11;7:18 [FREE Full text] [doi: 10.1186/1475-925X-7-18] [Medline: 18547411]
30. Sood SP, Nwabueze SN, Mbarika VW, Prakash N, Chatterjee S, Ray P, et al. Electronic medical records: a review comparing the challenges in developed and developing countries. : IEEE; 2008 Presented at: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008); January 7-10, 2008; Waikoloa, HI, USA. [doi: 10.1109/HICSS.2008.141]
31. Williams F, Boren SA. The role of electronic medical record in care delivery in developing countries. Int J Inf Manage 2008 Dec;28(6):503-507. [doi: 10.1016/j.ijinfomgt.2008.01.016]
32. Tilahun B, Fritz F. Modeling antecedents of electronic medical record system implementation success in low-resource setting hospitals. BMC Med Inform Decis Mak 2015 Aug 1;15:61 [FREE Full text] [doi: 10.1186/s12911-015-0192-0] [Medline: 26231051]
33. Aminpour F, Sadoughi F, Ahamdi M. Utilization of open source electronic health record around the world: a systematic review. J Res Med Sci 2014 Jan;19(1):57-64. [Medline: 24672566]
34. Dillon DG, Pirie F, Rice S, Pomilla C, Sandhu MS, Motala AA, African Partnership for Chronic Disease Research (APCDR). Open-source electronic data capture system offered increased accuracy and cost-effectiveness compared with paper methods in Africa. J Clin Epidemiol 2014 Dec;67(12):1358-1363 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.06.012] [Medline: 25135245]
35. Akanbi MO, Ocheke AN, Agaba PA, Daniyam CA, Agaba EI, Okeke EN, et al. Use of electronic health records in sub-Saharan Africa: progress and challenges. J Med Trop 2012;14(1):1-6. [Medline: 25243111]
36. Mostert-Phipps N, Pottas D, Korpela M. Improving continuity of care through the use of electronic records: a South African perspective. S Afr Fam Pract 2012;54(4):326-331. [doi: 10.1080/20786204.2012.10874244]
37. Fernández A, Oviedo E. 20.iadb. 2011. e-Health in Latin America and the Caribbean: progress and challenges URL: http://www20.iadb.org/intal/catalogo/PE/2011/08957.pdf [accessed 2017-11-03] [WebCite Cache ID 6ugvXXS0P]

38. Scholl J, Syed-Abdul S, Ahmed LA. A case study of an EMR system at a large hospital in India: challenges and strategies for successful adoption. J Biomed Inform 2011 Dec;44(6):958-967 [FREE Full text] [doi: 10.1016/j.jbi.2011.07.008] [Medline: 21846508]
39. Aminpour F, Sadoughi F, Ahmadi M. Towards the application of open source software in developing national electronic health record-narrative review article. Iran J Public Health 2013 Dec;42(12):1333-1339.
40. Manders EJ, José E, Solis M, Burlison J, Nhampossa JL, Moon T. Implementing OpenMRS for patient monitoring in an HIV/AIDS care and treatment program in rural Mozambique. Stud Health Technol Inform 2010;160(Pt 1):411-415. [doi: 10.3233/978-1-60750-588-4-411] [Medline: 20841719]
41. Puttkammer N, Zeliadt S, Balan JG, Baseman J, Destiné R, Domerçant JW, et al. Development of an electronic medical record based alert for risk of HIV treatment failure in a low-resource setting. PLoS One 2014 Nov 12;9(11):e112261 [FREE Full text] [doi: 10.1371/journal.pone.0112261] [Medline: 25390044]
42. Fraser HS, Thomas D, Tomaylla J, Garcia N, Lecca L, Murray M, et al. Adaptation of a web-based, open source electronic medical record system platform to support a large study of tuberculosis epidemiology. BMC Med Inform Decis Mak 2012 Nov 7;12:125 [FREE Full text] [doi: 10.1186/1472-6947-12-125] [Medline: 23131180]
43. Ali MK, Shah S, Tandon N. Review of electronic decision-support tools for diabetes care: a viable option for low- and middle-income countries? J Diabetes Sci Technol 2011 May 1;5(3):553-570 [FREE Full text] [doi: 10.1177/193229681100500310] [Medline: 21722571]
44. Tilahun B, Fritz F. Comprehensive evaluation of electronic medical record system use and user satisfaction at five low-resource setting hospitals in Ethiopia. JMIR Med Inform 2015 May 25;3(2):e22 [FREE Full text] [doi: 10.2196/medinform.4106] [Medline: 26007237]
45. Kiah ML, Haiqi A, Zaidan BB, Zaidan AA. Open source EMR software: profiling, insights and hands-on analysis. Comput Methods Programs Biomed 2014 Nov;117(2):360-382. [doi: 10.1016/j.cmpb.2014.07.002] [Medline: 25070757]
46. Zaidan AA, Zaidan BB, Al-Haiqi A, Kiah ML, Hussain M, Abdulnabi M. Evaluation and selection of open-source EMR software packages based on integrated AHP and TOPSIS. J Biomed Inform 2015 Feb;53:390-404 [FREE Full text] [doi: 10.1016/j.jbi.2014.11.012] [Medline: 25483886]
47. Zaidan AA, Zaidan BB, Hussain M, Haiqi A, Mat Kiah ML, Abdulnabi M. Multi-criteria analysis for OS-EMR software selection problem: a comparative study. Decis Support Syst 2015 Oct;78:15-27. [doi: 10.1016/j.dss.2015.07.002]
48. Mamlin BW, Biondich PG, Wolfe BA, Fraser H, Jazayeri D, Allen C, et al. Cooking up an open source EMR for developing countries: OpenMRS - a recipe for successful collaboration. AMIA Annu Symp Proc 2006;2006:529-533. [Medline: 17238397]
49. Mohammed-Rajput NA, Smith DC, Mamlin B, Biondich P, Doebbeling BN, Open MRS Collaborative Investigators. OpenMRS, a global medical records system collaborative: factors influencing successful implementation. AMIA Annu Symp Proc 2011;2011:960-968. [Medline: 22195155]
50. Tierney WM, Achieng M, Baker E, Bell A, Biondich P, Braitstein P, Tanzania-Uganda Openmrs Consortium. Experience implementing electronic health records in three East African countries. Stud Health Technol Inform 2010;160(Pt 1):371-375. [doi: 10.3233/978-1-60750-588-4-371]
51. Catalani C, Green E, Owiti P, Keny A, Diero L, Yeung A, et al. A clinical decision support system for integrating tuberculosis and HIV care in Kenya: a human-centered design approach. PLoS One 2014 Aug 29;9(8):e103205 [FREE Full text] [doi: 10.1371/journal.pone.0103205] [Medline: 25170939]
52. Anokwa Y, Ribeka N, Parikh T, Borriello G, Were MC. Design of a phone-based clinical decision support system for resource-limited settings. 2012 Presented at: Proceedings of the Fifth International Conference on Information and Communication Technologies and Development, ICTD '12; March 12-15, 2012; Atlanta, Georgia, USA p. 13-24.
53. Kasthurirathne SN, Mamlin B, Kumara H, Grieve G, Biondich P. Enabling better interoperability for healthcare: lessons in developing a standards based application programing interface for electronic medical record systems. J Med Syst 2015 Nov;39(11):182. [doi: 10.1007/s10916-015-0356-6] [Medline: 26446013]
54. Haskew J, Rø G, Saito K, Turner K, Odhiambo G, Wamae A, et al. Implementation of a cloud-based electronic medical record for maternal and child health in rural Kenya. Int J Med Inform 2015 May;84(5):349-354. [doi: 10.1016/j.ijmedinf.2015.01.005] [Medline: 25670229]
55. Allen C, Jazayeri D, Miranda J, Biondich PG, Mamlin BW, Wolfe BA, et al. Experience in implementing the OpenMRS medical record system to support HIV treatment in Rwanda. Stud Health Technol Inform 2007;129(Pt 1):382-386. [Medline: 17911744]
56. Seebregts CJ, Mamlin BW, Biondich PG, Fraser HS, Wolfe BA, Jazayeri D, et al. Human factors for capacity building: lessons learned from the OpenMRS implementers network. Yearb Med Inform 2010:13-20. [Medline: 20938564]
57. Thompson A, Castle E, Lubeck P, Makarfi PS. Experience implementing OpenMRS to support maternal and reproductive health in Northern Nigeria. Stud Health Technol Inform 2010:332-336. [doi: 10.3233/978-1-60750-588-4-332]
58. Wolfe BA, Mamlin BW, Biondich PG, Fraser HS, Jazayeri D, Allen C, et al. The OpenMRS system: collaborating toward an open source EMR for developing countries. AMIA Annu Symp Proc 2006;2006:1146. [Medline: 17238765]

59. Seebregts CJ, Mamlin BW, Biondich PG, Fraser HS, Wolfe BA, Jazayeri D, OpenMRS Implementers Network. The OpenMRS Implementers Network. Int J Med Inform 2009 Nov;78(11):711-720. [doi: 10.1016/j.ijmedinf.2008.09.005] [Medline: 19157968]
60. Maglogiannis I. Towards the adoption of open source and open access electronic health record systems. J Healthc Eng 2012;3(1):141-161. [doi: 10.1260/2040-2295.3.1.141]
61. Bahmni. Implementations URL: https://www.bahmni.org/implementations [accessed 2017-11-01] [WebCite Cache ID 6ueGVeCW0]
62. Syzdykova A, Zolfo M, Malta A, Diro E, Oliveira JL. Customization of OpenMRS for leishmaniasis research and treatment center in Ethiopia. J Int Soc Telemed eHealth 2017;5:e65 [FREE Full text]
63. Avelino JN, Hebron CT, Laranang AL, Paje PN, Bautista MM, Caro JD. Requirements gathering as an essential process in customizing health information systems for small scale health care facilities. 2014 Presented at: 5th International Conference on Information, Intelligence, Systems and Applications (IISA); July 7-9, 2014; Chania, Greece. [doi: 10.1109/IISA.2014.6878729]
64. Karopka T, Schmuhl H, Demski H. Free/Libre open source software in health care: a review. Healthc Inform Res 2014 Jan;20(1):11-22 [FREE Full text] [doi: 10.4258/hir.2014.20.1.11] [Medline: 24627814]
65. Prodhan UK, Rahman MZ, Abid A, Bellah M. Development of a telemedicine model with low cost portable tool kit for remote diagnosis of rural people in Bangladesh. 2016 Presented at: International Conference on Innovations in Science, Engineering and Technology (ICISET); October 28-29, 2016; Dhaka, Bangladesh. [doi: 10.1109/ICISET.2016.7856513]
66. de la Torre I, Martinez B, Lopez-Coronado M. Analyzing open-source and commercial EHR solutions from an international perspective. 2013 Presented at: 2013 IEEE 15th International Conference on e-Health Networking, Applications & Services (Healthcom); October 9-12, 2013; Lisbon, Portugal. [doi: 10.1109/HealthCom.2013.6720708]
67. Safadi H, Chan D, Dawes M, Roper M, Faraj S. Open-source health information technology: a case study of electronic medical records. Health Policy Technol 2015 Mar;4(1):14-28. [doi: 10.1016/j.hlpt.2014.10.011]
68. ONC-Authorized Certification Body. Chpl.healthit. Certified health IT product list URL: https://chpl.healthit.gov/ [accessed 2017-11-01] [WebCite Cache ID 6ueHJkBQ3]
69. FreeMED Software Foundation. Transifex. FreeMED open source EMR/PM system URL: https://www.transifex.com/freemed/freemed/ [accessed 2017-11-01] [WebCite Cache ID 6ueHS7j7L]

Abbreviations

ACL: Access Control List
ATC: Anatomical Therapeutic Chemical
API: application programming interface
CIEL/MVP: Columbia International eHealth Laboratory/Millennium Villages Project
CPT: Current Procedural Terminology
DBMS: database management systems
DICOM: Digital Imaging and Communications in Medicine
eHealth: electronic health
EHR: electronic health record
EMR: electronic medical record
ERP: enterprise resource planning
FHIR: Fast Healthcare Interoperability Resources
GPL: general public license
HCPCS: Healthcare Common Procedure Coding System
HIS: health information system
HIV: human immunodeficiency virus
HL7: Health Level-7
ICD: International Classification of Diseases
LIS: laboratory Information system
LDAP: lightweight directory access protocol
LMICs: low- and middle-income countries
LOINC: Logical Observation Identifiers Names and Codes
NPO: nonprofit organization
OpenMRS: open medical record system
OS EHR: open-source EHR
OSS: open-source software
OSCAR: open source clinical application resource
PMS: practice management software
RBAC: role-based access control

SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms
UTF: Unicode Transformation Format
VistA: Veterans and Information System &Technology Architecture

Edited by G Eysenbach; submitted 04.06.17; peer-reviewed by V Maojo, D Revere; comments to author 12.07.17; revised version received 14.08.17; accepted 13.09.17; published 13.11.17.

Please cite as:
Syzdykova A, Malta A, Zolfo M, Diro E, Oliveira JL
Open-Source Electronic Health Record Systems for Low-Resource Settings: Systematic Review
JMIR Med Inform 2017;5(4):e44
URL: http://medinform.jmir.org/2017/4/e44/
doi:10.2196/medinform.8131
PMID:29133283

Viewpoint

Promoting Secondary Analysis of Electronic Medical Records in China: Summary of the PLAGH-MIT Critical Data Conference and Health Datathon

Peiyao Li[1*], MS; Chen Xie[2*], MEng (Biomed Eng); Tom Pollard[2], PhD; Alistair Edward William Johnson[2], PhD; Desen Cao[1], BS; Hongjun Kang[3], MD, PhD; Hong Liang[1], MS; Yuezhou Zhang[1], MS; Xiaoli Liu[4], MS; Yong Fan[5], MS; Yuan Zhang[5], MS; Wanguo Xue[6], MS; Lixin Xie[7], MD, PhD; Leo Anthony Celi[2], MD, MPH, MS; Zhengbo Zhang[1], PhD

[1]Department of Biomedical Engineering, Chinese People's Liberation Army General Hospital, Beijing, China

[2]Laboratory of Computational Physiology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, United States

[3]Department of Critical Care Medicine, Chinese People's Liberation Army General Hospital, Beijing, China

[4]School of Biological Science and Medical Engineering, Beihang University, Beijing, China

[5]Chinese People's Liberation Army Medical School, Beijing, China

[6]Department of Computer Application and Management, Chinese People's Liberation Army Medical School, Beijing, China

[7]Department of Respiratory and Critical Care Medicine, Chinese People's Liberation Army General Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zhengbo Zhang, PhD
Department of Biomedical Engineering
Chinese People's Liberation Army General Hospital
28, Fuxing Road
Haidian District
Beijing, 100853
China
Phone: 86 13693321644
Fax: 86 1066937983
Email: zhengboz@mit.edu

Abstract

Electronic health records (EHRs) have been widely adopted among modern hospitals to collect and track clinical data. Secondary analysis of EHRs could complement the traditional randomized control trial (RCT) research model. However, most researchers in China lack either the technical expertise or the resources needed to utilize EHRs as a resource. In addition, a climate of cross-disciplinary collaboration to gain insights from EHRs, a crucial component of a learning healthcare system, is not prevalent. To address these issues, members from the Massachusetts Institute of Technology (MIT) and the People's Liberation Army General Hospital (PLAGH) organized the first clinical data conference and health datathon in China, which provided a platform for clinicians, statisticians, and data scientists to team up and address information gaps in the intensive care unit (ICU).

(JMIR Med Inform 2017;5(4):e43) doi:10.2196/medinform.7380

KEYWORDS

electronic health record; datathon; database; intensive care units

The Potential of Healthcare Data

On June 21, 2016, the Chinese State Council promulgated its Guiding Opinions on Promoting and Regulating the Development of the Application of Healthcare Big Data. The report identified healthcare data to be a strategic national resource and that its development would significantly impact healthcare and medical treatment. It also raised the importance of gathering and utilizing healthcare data to a national level.

Healthcare data comes from many sources, giving rise to various fields (ie, medical and health informatics, translational bioinformatics, sensor informatics, and imaging informatics), and within each field, new analytic tools to understand health and disease [1]. In particular, the analysis of data contained within electronic health records (EHRs) is a promising avenue of research for clinicians and data scientists. EHRs contain large volumes of data regarding patient care, both structured and unstructured, making them highly valuable resources for knowledge discovery [2].

For decades, clinical research has relied on randomized controlled trials (RCTs) as the authoritative methodology—conclusions from RCTs are deemed inherently more reliable compared to those of observational studies. While medical societies rely on RCTs to develop clinical practice guidelines, they have some well-known and notable drawbacks [3]. They are costly, labor-intensive, generally take a long time to complete, and tend to contain restrictive inclusion and exclusion criteria, which leads to limited generalizability.

With the spread of EHRs across hospitals, there is an interest in harnessing the data contained therein to power longitudinal, population-based studies without the artificial conditions imposed by RCTs [4]. Once the data infrastructure is created, further research costs tend to be low and these databases lend themselves to iterative analyses—findings can be tested against a wide variety of populations, circumstances, potential confounders, and timeframes, all contained within the existing data. New insights pertinent to the day-to-day practice of clinicians may be gleaned from the troves of digital documentation [2,5]. Like RCTs, retrospective clinical research using EHRs should be hypothesis-driven. However, unlike RCTs, the variables that are included in the analysis are not limited to those pre-defined during the design phase of the RCT. In addition, post-hoc analysis of patient subsets in an RCT for hypothesis generation is limited by the fixed sample size. This is typically not the case for secondary analysis of EHRs. However, the absence of randomization makes retrospective studies problematic and requires more complex causal inference methodologies.

Compared to other countries, China enjoys several advantages when it comes to EHR-based clinical research. First, China has a large population, growing by 5.84% between 2000 and 2010 to 1.34 billion. To meet the needs of this growing population, China has a very large healthcare system [6]. The National Health and Family Planning Commission of the People's Republic of China (NHFPC) statistics from June 2016 show that there were 989,403 health institutions in China, including 28,261 hospitals, 927,147 community health centers, and 30,814 specialized public health institutions [7,8]. In the first half of 2016, over 3.84 billion visits were made to these health institutions, including over 1.56 billion hospital visits [9]. The volume of health information generated by this enormous amount of healthcare consumption presents a great opportunity to understand determinants of patient outcomes and identify treatments most effective among the Chinese population.

China also has relatively robust clinical information systems. These systems have been built and refined since the 1997 launch of the Military No.1 project, which was invented and promoted by the People's Liberation Army General Hospital (PLAGH). This project established a software framework with a set of standards targeted towards medium and large hospitals for hospital, staff, and patient administration, and was adopted by several hundred hospitals over the next decade. The primary functions of its EHR system included keeping track of patient visits, medical history, treatments, and drug prescriptions [10].

In the central government's 12th Five-Year Plan (2011-2016), a series of social and economic development initiatives, the NHFPC declared its "3521" Project for electronic health (e-Health), which was subsequently revised to "4631-2." Here, the "4" denotes the establishment of 4 levels of healthcare administration platforms: national, provincial, prefecture, and county. The "6" represents the strengthening of 6 key services: public health, medical services, medical security, drug administration, family planning, and integrated health management. The "3" denotes the development of 3 fundamental database systems: the electronic medical record, the EHR, and a national database containing the healthcare-related information of the entire population. The "1" denotes the establishment of one centralized communication network to integrate all the previously listed elements. The final "2" denotes the establishment of health information system standards and the protection of healthcare information [7,11,12]. With this government mandate, the EHR adoption rate increased from 25% in 2008 to 47% in 2013 [11].

Challenges to Secondary Use of Health Data in China

Current EHRs are designed to support administrative and billing functions, frequently resulting in their inability to capture clinical information in a structured manner, not to mention poor usability, which hampers clinician efficiency.

The 2015 hospital information systems survey by the China Hospital Information Management Association (CHIMA) received comprehensive responses from 536 hospitals, which is about 2% of the national total [13]. In accordance with the NHFPC's 3-tier classification system, there were 342 (63.8%, 342/536) tier 3 hospitals, and 194 (36.2%, 194/536) lower tier hospitals, where a higher tier indicates a more advanced hospital. The metrics used to define the tiers are hospital capacity based on staff and bed numbers, the level of technology employed, the equipment and facilities available, the quality of the management and logistics, and the quality of care. The tier distribution of the survey sample differs heavily from the national distribution: 12.4% tier 3 and 87.6% tier 2 or below. Hence, the patterns drawn from the survey give insight into China's more technologically and financially advanced hospitals, rather than the national trend.

Among the surveyed hospitals, the most frequently used hospital information system services and the percentage of hospitals using each service are as follows: pharmaceutical administration (78%), emergency department billing and administration (77%), inpatient medication administration (77%), and outpatient scheduling (76%). Correspondingly, the most prevalent issues

raised against the hospital information systems and the percentage of hospitals raising such issues were: insufficient interoperability and standards (53%), lack of flexibility and features to capture individual patient characteristics (48%), and a lack of human usability in the software layout (44%). Given this information, it is not surprising that a recent study of tertiary hospitals showed high EHR adoption rates (73%) but extremely low (1%) levels of data integration for informing clinical decision support [14].

The development of healthcare databases able to support secondary analysis is impeded by 2 skill shortages. The supply side issue is the lack of skilled information technology (IT) professionals needed to integrate EHR data to generate a research resource and that understand healthcare data. Just as important, is the lack of demand for such resources by potential clinical researchers in China due to a lack of familiarity with the field. According to an informal survey conducted with 37 staff members from the 20 departments of the PLAGH ahead of the datathon, clinicians admitted lacking knowledge and background about how to incorporate insights from digital health data into their clinical practice. Clinicians typically conduct clinical research independently or as the director of a group that includes a statistician. However, secondary analysis of EHRs requires a multidisciplinary team approach, with members appreciating each other's contribution and acknowledging the limitations of their expertise [15]. Clinicians must be willing to embrace the uncertainties and information gaps in the practice of medicine and their limited understanding of machine learning, while data scientists must defer to clinicians in formulating relevant projects that can lead to a change in practice and in contextually interpreting their findings.

Another issue in China regarding health analytics is the quality and accessibility of data. Despite the availability of massive data sources like EHRs, wireless sensors, and medical images, the aggregation of data to produce resources that facilitate clinical research is very limited. An ideal research database is the Medical Information Mart for Intensive Care (MIMIC), a well-curated open-access database developed and maintained by the Laboratory of Computational Physiology (LCP) at the Massachusetts Institute of Technology (MIT), and supported by a vibrant research community [16]. Databases drawn from EHRs in China are smaller and accessible only to investigators internally within a hospital or organization. One notable exception is the National Scientific Data Sharing Platform for Population and Health, which encompasses or connects to various databases that include biological data, clinical data, public health data, Chinese traditional medicine data, pharmacy data, and national population and reproductive health science data [17]. These sources are analogous to US sources such as the National Inpatient Sample, which do not provide such high-resolution data as that contained in MIMIC. The following are 2 central factors seen in all countries, including China, that impede the large-scale building and dissemination of healthcare databases: (1) the previously mentioned issue of the lack of interoperability, and (2) outdated government regulations and general attitudes about data sharing and management.

As one of the largest premier hospitals in the country, and the center at which the Military No.1 project was developed, the PLAGH can be used as a reference for the state of the art in China. Each of its departments has collected vast amounts of data and can access central patient EHRs. However, there is no unified database or any direct link between the individual departments' specialized information systems, which were designed by different vendors. Although the hospital implements several health information system standards, the individual software vendors do not consistently do so, choosing instead to focus on building software features for the various specialties. During purchase of information systems (eg, pharmacy, laboratory, clinical departments), hospital-wide integration has not been a consideration, as the recognition of EHRs as valuable research resource is only recent. As of October 2016, most of the PLAGH's departments are working to build databases or extend and port their existing ones. A key issue that has hindered the hospital from creating a unified health information system is the undersupply of healthcare IT professionals. The hospital is only recently expanding its dedicated hospital-wide IT department. This is an even bigger issue for smaller hospitals in the country that do not have the budget for an IT team.

The Clinical Data Conference and Health Datathon

Given these challenges, and the opportunities for growth in the utilization of EHRs for clinical research, we organized the first PLAGH-MIT clinical data conference and health datathon in Beijing, China on October 21-22, 2016.

The word datathon originates from hackathon, a short but high-energy event in which teams generate innovative, technological solutions to real-world problems. Deviating from hackathons, datathons bring together people from diverse backgrounds around data. In the context of health datathons, participants may include clinicians, data scientists, statisticians, and even patients [18].

We believe that gathering data scientists and clinicians and giving them the opportunity to explore an EHR-derived clinical database could demonstrate the value of such resources for knowledge creation and validation. As with the hackathon, the key objective of the datathon is to convince the stakeholders, who are typically holed up in their own silos, that they can accomplish so much more if they take advantage of each other's expertise.

The MIT team has organized dozens of health hackathons and datathons around the world (Table 1) [19-21]. For outcomes, the group has been focusing on measuring affective learning and teamwork skills gained by the participants instead of metrics such as publications produced and patents for and start-up companies. The ability to work across disciplines is considered an instrumental attribute in the design, implementation, and evaluation of technological solutions to address problems in healthcare.

Table 1. Healthcare hackathons and datathons hosted by the Massachusetts Institute of Technology Lab for Computational Physiology.

Date	Event	Location
July 2017	Health Datathon	Singapore
June 2017	Mobile Health Hackathon	Mexico City, Mexico
May 2017	Health Datathon	Sao Paulo, Brazil
April 2017	Hacking Discrimination Hackathon	MIT[a], United States
March 2017	Health Datathon	Melbourne, Australia
January 2017	Mobile Health Hackathon	Khon Kaen, Thailand
December 2016	Health Datathon	London, United Kingdom
October 2016	Health Datathon	Beijing, China
September 2016	Internet of Things Hackathon	Taipei, Taiwan
August 2016	Hacking Mobile Health Hackathons	MIT, United States
January 2016	Mobile Health Hackathon	Mexico City, Mexico
October 2015	Mobile Health Hackathon	Thessaloniki, Greece
September 2015	Health Datathon	MIT, United States and London, United Kingdom
July 2015	Mobile Health Hackathon	Kampala, Uganda
June 2015	Mobile Health Hackathon	Popayan, Colombia
September 2014	Health Datathon	MIT, United States, London, United Kingdom, and Paris, France
January 2014	Health Datathon	MIT, United States

[a]MIT: Massachusetts Institute of Technology.

Another notable initiative that addresses cross-disciplinary healthcare research in China is the Joint Institute for Translational and Clinical Research. Established between the University of Michigan Health System and the Peking University Health Science Center in 2010, this partnership aims to leverage the diverse expertise of researchers across both countries and universities. Members have established multiple clinician-led clinical data projects and programs such as the Biorepository and Bioinformatics Core that supports the acquisition, storage, and management of clinical information and bio-specimens [22].

Unsurprisingly, we received strong interest from many sectors as soon as the event was announced, with more than a thousand conference attendance requests on the first day of registration. Priority was given to clinicians who had previously collaborated with PLAGH. The final composition of the conference attendees was approximately one-third doctors, one-third data scientists, and one-third biomedical engineers. There were also several clinical directors, from both private and public hospitals across China. From the PLAGH, the directors of the emergency department, intensive care unit (ICU), and respiratory medicine department attended.

Disproportionately more data scientists registered for the datathon than clinicians did. This was expected and has been observed by the MIT team in events they organized across the globe. Clinicians are generally busy with patient care responsibilities and schedules that are not easy to rearrange. There is also no compelling incentive for healthcare providers to participate in research. To promote this event within the PLAGH and to recruit clinicians for the datathon, the organizers visited several departments including the emergency department, and the respiratory, cardiac and surgical ICUs. The biggest recruitment challenge within the hospital was that clinicians were not clear about their role in the datathon and had difficulty envisioning how to interface with data scientists with expertise in machine learning and signal processing. In addition, clinicians were uncertain about the types of research projects suitable for the datathon given their unfamiliarity with MIMIC-III.

The datathon centered around MIMIC-III, a de-identified database containing health related information associated with over 40,000 patients admitted into the ICUs of the Beth Israel Deaconess Medical Center between 2001 and 2012 [16]. In addition to administrative data such as transfers, discharges, and billing information, MIMIC-III contains high resolution medical data such as hourly physiological measurements, diagnoses, laboratory test results, death data collected from both the hospital and the government, and even clinical notes. A dedicated SQL server was created to allow participants to query the MIMIC-III database through a secure private network configured in the hospital. To ensure proper care was taken with the data, all participants were made to complete a training program in human research participant protection and Health Insurance Portability and Accountability Act (HIPAA) regulations beforehand. Those who are granted access agree to use it solely for population-based scientific research. They may not share it or search for specific individuals in the database.

Like prior health datathons hosted in other countries by the MIT team, the Beijing event consisted of 2 parts: a half-day conference on health data, and a full day of hands-on exploration and analysis of the clinical database.

The conference commenced with a welcome address from the PLAGH vice president Kunlun He. Dr Leo Anthony Celi discussed the opportunities from and challenges in the secondary analysis of EHRs and data sharing, based on his experiences at MIT as clinical research director of the LCP. Dr Kee Yuan Ngiam described how the National University of Singapore hospital has put health data into real-world use to establish a clinical decision support system. Dr Zhengbo Zhang described the use of numerical models applied to physiological data captured through wearables, in creating personalized diagnoses and treatments. Lastly, Dr Tom J Pollard and Dr Alistair Johnson described the evolution of the MIMIC-III database and shared their experiences in data analysis using MIMIC-III.

Research projects were proposed by PLAGH clinicians prior to the event and reviewed by both PLAGH and MIT teams in order to assess their suitability. The final approved projects were (1) total fluid balance and mortality in elderly critically ill patients; (2) serum N-terminal pro b-type natriuretic peptide (NT-proBNP) level and the duration of mechanical ventilation; (3) trends in the use of continuous renal replacement therapy in critically ill patients from 2001 through 2012; (4) variations in the treatment of hypotension according to time of day and/or day of the week; (5) the effect of age and clinical circumstances on the outcome of red blood cell transfusion in critically ill patients; and (6) the use of intra-aortic balloon pump and lactate clearance.

After the participants were split into 6 teams and assigned research questions, each team spent a full day to understand the context of the project assigned to them with input from their clinicians, extract the data from MIMIC-III, and develop data models to address their research question. Each team presented their study design and preliminary findings and shared their thoughts and experiences at the conclusion of the event. An expert panel that consisted of both computer scientists and clinicians judged the presentations and selected the winning team.

Results of the Datathon

The majority of the datathon participants had little or no experience performing research with EHR data. Key concepts were elucidated, including the variability of the data quality due to documentation methods, how database variables are stored and distributed across multiple tables to optimize data storage, and the lack of a graphical user interface (GUI) for viewing the time series nature of majority of the variables.

Data scientists and statisticians relied on their team's clinicians to understand the context of their assigned questions and interpret their analyses, and clinicians comfortably relied on the data scientists to perform the data extraction and analysis. Though it was the job of the data scientists to extract data, many of them had never worked with a high-resolution database, and hence extensively deferred to their team leaders from MIT. Data scientists were also aided by a collaborative code repository maintained by the LCP and hosted on GitHub. Once the data had been extracted and preprocessed, the teams were able to quickly progress through data modeling. As expected, the language barrier between some teams and the MIT team slowed down their investigations.

In the end, each of the 6 teams was able to extract their selected variables, including the outcome of interest, and perform a preliminary analysis. The expert panel gave feedback regarding each team's choice of covariates and data models and suggested follow-up methodologies to further their clinical projects. The winning team, which investigated red blood cell transfusion, presented their coding framework that allowed them to visualize quickly the effects of adding and removing variables in the data model. This framework was well received as a useful tool with which clinicians can explore the database. Visualization software has proven helpful for non-coding clinical researchers to explore databases [23].

Participants praised the datathon's model of providing practical experience in the design and implementation of research projects using electronic healthcare data, as opposed to previous events they had attended which only included presentations. The participants also stated that they enjoyed working with a diverse group of clinicians, data scientists, and biomedical engineers, and those from the PLAGH believed that this event would help foster interdepartmental collaboration.

Lessons Learned Towards a Data Driven Learning Healthcare System in China

Compared with previous datathons hosted by the MIT team in the United States, United Kingdom, and France, participants were less independent in carrying out their tasks, more reliant on the event organizers, and less familiar with the concept of sharing code and data [18]. Some attendees also left skeptical about the feasibility of performing research on EHR data given the state of the art in China. The experience of and the feedback from the datathon attendees, who represented some of the leading institutions of the country, suggest that the analysis of healthcare data in China is still in its infancy.

In the end, the event was still successful in that participants were shown the promise of EHR derived healthcare databases and the potential of using them to answer clinical questions, with every group producing preliminary findings and several groups choosing to continue their projects after the event. Projects originating from past datathons have led to publications and valuable code contributions to the MIMIC code repository, a shared open access storehouse of codes for querying and analyzing MIMIC [24]. Attendees also experienced firsthand the value of cross-disciplinary collaboration, an important take-home message that rings especially true in the convoluted field of health data analytics [2,18]. The clinicians expressed enthusiasm about the possibility of holding future datathons with databases constructed from the PLAGH. In conjunction with promoting the value of healthcare databases, the event also served its purpose of highlighting the challenges and limitations of building and learning from them, and the need to invest heavily in the development of such resources. Following the event, the PLAGH launched a series of hospital-wide data merging and data warehouse construction projects. In particular, it is building an ICU database with the intention of publishing

it and using it for the next datathon. In addition, the biomedical engineering department is preparing its new course: Secondary Analysis of EHRs for the hospital's graduate school. This course teaches skills in data extraction, processing, and modeling, training new researchers to leverage the value of big data in healthcare.

As the value of health data analytics becomes more apparent, it is likely that Chinese institutions will dedicate more funding and workers into building EHR systems and creating databases from them [25]. The tier 3 and the tier 2 and below hospitals that participated in the 2015 CHIMA survey plan to spend an average of RMB $15 million and RMB $4.4 million (US $2 million and $0.6 million, respectively) on their digital information systems in 2016 and 2017. In addition, 96% of them have already established a department dedicated to the digitization of their health records [13].

However, the speed and efficiency at which healthcare databases are built, and the volume and quality research that ultimately arises from them, will largely depend on institutions' understanding of health data—what it represents, its structure, the way it is captured, and the inherent biases as a result. The goals of digital health information systems are to improve the efficiency of operations (88%), to reduce clinical errors (85%), to reduce operating costs (67%), and to improve patient satisfaction (60%) [13]. Notably excluded from the list is the use of data by health organizations to learn continuously from the way care is delivered to maximize patient outcomes.

The MIT LCP and the PLAGH will continue to organize events to demonstrate the benefits of EHR databases to Chinese stakeholders: the government, industry, hospitals and clinics, but most of all, the patients. Only when these stakeholders appreciate their value, will there be a push to invest in building such resources. In addition, these events promote the need to train more technical specialists to build and manage healthcare databases and the clinicians and data scientists who understand how to use them.

Last, we cannot emphasize enough the importance of collaboration and data sharing across healthcare organizations. The lack of data standards and interoperability between EHRs has greatly impeded learning from routinely collected health data [26,27]. This, along with the culture of researchers hoarding data and working in silos, results in waste and inefficiencies in biomedical research. Hence members from government, industry, and academia must coordinate their efforts to develop and implement common data standards and adopt policies that promote data sharing.

The late emergence of EHRs in China relative to the United States provides China a unique opportunity to learn from past experiences, including failures in developing an efficient digital healthcare infrastructure. Obstacles and challenges, if addressed and circumvented, could facilitate country-wide EHR analysis and reduce institutional hoarding of data, clinicians and data scientists continuing to work in silos, and the lack of incentives for data systems interoperability. Experts should continue to push for a culture of data sharing and collaboration and expound the vast potential and practical limitations of the secondary analysis of EHRs to make data-driven learning in healthcare a reality in China.

Acknowledgments

This clinical data conference and health datathon was partially supported by a grant from the Chinese National Natural Science Foundation (Grant number: 61471398). MIMIC database is funded by the National Institute of Health through the R01 grant R01 EB017205-01A1. The authors would like to thank staff at the Chinese PLAGH, Beijing, for supporting database development and secondary usage of electronic medical records.

Conflicts of Interest

None declared.

References

1. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang G. Big data for health. IEEE J Biomed Health Inform 2015 Jul;19(4):1193-1208. [doi: 10.1109/JBHI.2015.2450362] [Medline: 26173222]
2. MIT Critical Data. Secondary Analysis of Electronic Health Records. New York, NY: Springer; 2016.
3. Frieden TR. Evidence for health decision making - beyond randomized, controlled trials. N Engl J Med 2017 Dec 03;377(5):465-475. [doi: 10.1056/NEJMra1614394] [Medline: 28767357]
4. Van Poucke S, Thomeer M, Heath J, Vukicevic M. Are randomized controlled trials the (g)old standard? From clinical intelligence to prescriptive analytics. J Med Internet Res 2016 Jul 06;18(7):e185 [FREE Full text] [doi: 10.2196/jmir.5549] [Medline: 27383622]
5. Schneeweiss S. Improving therapeutic effectiveness and safety through big healthcare data. Clin Pharmacol Ther 2016 Mar;99(3):262-265. [doi: 10.1002/cpt.316] [Medline: 26659268]
6. National Bureau of Statistics of China. Communiqué of the National Bureau of Statistics of People's Republic of China on Major Figures of the 2010 Population Census[1] (No. 1). 2011 Apr 28. URL: https://web.archive.org/web/20131108022004/http://www.stats.gov.cn/english/newsandcomingevents/t20110428_402722244.htm [accessed 2017-10-17] [WebCite Cache ID 6uHJfcprB]

7. Gao X, Xu J, Sorwar G, Croll P. Implementation of e-health record systems and e-medical record systems in China. ITMR 2013;3(2):127. [doi: 10.2991/itmr.2013.3.2.6]
8. Medical Institutions of China. Chinese National Health and Family Planning Commission. 2016. URL: http://www.nhfpc.gov.cn/mohwsbwstjxxzx/s7967/201608/87343c7d63ce41ca8d8fddf7a5db66b7.shtml [accessed 2017-01-20] [WebCite Cache ID 6nfU8WeMY]
9. Chinese National Health and Family Planning Commission. Usage of medical institutions of China, First Half of 2016. 2016. URL: http://www.nhfpc.gov.cn/mohwsbwstjxxzx/s7967/201608/33be8f3399614c11909dc4da0a940c95.shtml [accessed 2017-01-20] [WebCite Cache ID 6nfSuNfBS]
10. Yu G. Establishment and application of No. 1 Military Project. Hosp Admin J Chin PLA 2000.
11. China Hospital Information Management Association (CHIMA), National Institute of Hospital Administration (NIHA). White paper on China's hospital information systems 2008-2013. HIMSS 2013.
12. Chen Z. Launch of the health-care reform plan in China. Lancet 2009 Apr 18;373(9672):1322-1324. [doi: 10.1016/S0140-6736(09)60753-4] [Medline: 19376436]
13. China Hospital Information Management Association (CHIMA). Report on China's hospital information systems. CHIMA 2016.
14. Shu T, Liu H, Goss FR, Yang W, Zhou L, Bates DW, et al. EHR adoption across China's tertiary hospitals: a cross-sectional observational study. Int J Med Inform 2014 Feb;83(2):113-121. [doi: 10.1016/j.ijmedinf.2013.08.008] [Medline: 24262068]
15. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014 Jan 11;383(9912):166-175 [FREE Full text] [doi: 10.1016/S0140-6736(13)62227-8] [Medline: 24411645]
16. Johnson A, Pollard T, Shen L, Lehman L, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035. [doi: 10.1038/sdata.2016.35]
17. NCMI. URL: http://www.ncmi.cn/1 [accessed 2017-01-20] [WebCite Cache ID 6nfU1julb]
18. Aboab J, Celi LA, Charlton P, Feng M, Ghassemi M, Marshall DC, et al. A “datathon” model to support cross-disciplinary collaboration. Sci Transl Med 2016 Apr 06;8(333):333ps8. [doi: 10.1126/scitranslmed.aad9072] [Medline: 27053770]
19. Angelidis P, Berman L, Casas-Perez MD, Celi LA, Dafoulas GE, Dagan A, et al. The hackathon model to spur innovation around global mHealth. J Med Eng Technol 2016;40(7-8):392-399. [doi: 10.1080/03091902.2016.1213903] [Medline: 27538360]
20. Pathanasethpong A, Soomlek C, Morley K, Morley M, Polpinit P, Dagan A, et al. Tackling regional public health issues using mobile health technology: event report of an mHealth hackathon in Thailand. JMIR Mhealth Uhealth 2017 Oct 16;5(10):e155 [FREE Full text] [doi: 10.2196/mhealth.8259] [Medline: 29038098]
21. Neto A, Kugener G, Bulgarelli L, Rabello-Filho R, Armengol M, Alistair E, et al. White paper from the first critical care datathon in Brazil. Revista Brasileira de Terapia Intensiva (In Press) 2017.
22. The University of Michigan Health System and Peking University Health Science Center Joint Institute for Translational Clinical Research. URL: http://puuma.org/home [WebCite Cache ID 6tYKR6EEB]
23. Van PS, Zhang Z, Schmitz M, Vukicevic M, Laenen MV, Celi LA, et al. Scalable predictive analysis in critically ill patients using a visual open data analysis platform. PLoS One 2016;11(1):e0145791 [FREE Full text] [doi: 10.1371/journal.pone.0145791] [Medline: 26731286]
24. Johnson A, Stone D, Celi L, Pollard T. The MIMIC Code Repository: enabling reproducibility in critical care research. J Am Med Inform Assoc 2017 Sep 27. [doi: 10.1093/jamia/ocx084] [Medline: 29036464]
25. Reimers K, Guo X, Li M, Xie B, Zhang T. Innovating in a Learning Community: Emergence of an Open Information Infrastructure in China's Pharmaceutical Distribution Industry. New York, NY: Springer; 2014.
26. Brailer DJ. Interoperability: the key to the future health care system. Health Aff (Millwood) 2005;Suppl Web Exclusives:W5-19 [FREE Full text] [doi: 10.1377/hlthaff.w5.19] [Medline: 15659454]
27. Stewart BA, Fernandes S, Rodriguez-Huertas E, Landzberg M. A preliminary look at duplicate testing associated with lack of electronic health record interoperability for transferred patients. J Am Med Inform Assoc 2010;17(3):341-344 [FREE Full text] [doi: 10.1136/jamia.2009.001750] [Medline: 20442154]

Abbreviations

CHIMA: China Hospital Information Management Association
EHR: electronic health record
ICU: intensive care unit
IT: information technology
LCP: Lab for Computational Physiology
MIT: Massachusetts Institute of Technology
MIMIC-III: Medical Information Mart for Intensive Care
NHFPC: National Health and Family Planning Commission of the People’s Republic of China
PLAGH: People’s Liberation Army General Hospital

RCT: randomized control trial

Edited by G Eysenbach; submitted 22.01.17; peer-reviewed by C Kruse, M Feng, J Lei, J Calvert, S Van Poucke, T Naumann, R Goleva, R Mark; comments to author 12.07.17; revised version received 17.08.17; accepted 20.09.17; published 14.11.17.

Please cite as:
Li P, Xie C, Pollard T, Johnson AEW, Cao D, Kang H, Liang H, Zhang Y, Liu X, Fan Y, Zhang Y, Xue W, Xie L, Celi LA, Zhang Z
Promoting Secondary Analysis of Electronic Medical Records in China: Summary of the PLAGH-MIT Critical Data Conference and Health Datathon
JMIR Med Inform 2017;5(4):e43
URL: http://medinform.jmir.org/2017/4/e43/
doi:10.2196/medinform.7380
PMID:29138126

Original Paper

Standard Anatomic Terminologies: Comparison for Use in a Health Information Exchange–Based Prior Computed Tomography (CT) Alerting System

Anton Oscar Beitia[1], MD; Tina Lowry[1], MS; Daniel J Vreeman[2,3], PT, DPT, MSc; George T Loo[1], MPA, MPH, DrPH; Bradley N Delman[4], MD; Frederick L Thum[1], MD; Benjamin H Slovis[5], MD; Jason S Shapiro[1], MD

[1]Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[2]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States

[3]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

[4]Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[5]Department of Emergency Medicine, Thomas Jefferson University, Philadelphia, PA, United States

Corresponding Author:
Anton Oscar Beitia, MD
Department of Emergency Medicine
Icahn School of Medicine at Mount Sinai
Box 1620,
One Gustave L. Levy Pl
New York, NY, 10029
United States
Phone: 1 2128248047
Fax: 1 2124261946
Email: anton.beitia@mountsinai.org

Abstract

Background: A health information exchange (HIE)–based prior computed tomography (CT) alerting system may reduce avoidable CT imaging by notifying ordering clinicians of prior relevant studies when a study is ordered. For maximal effectiveness, a system would alert not only for prior same CTs (exams mapped to the same code from an exam name terminology) but also for similar CTs (exams mapped to different exam name terminology codes but in the same anatomic region) and anatomically proximate CTs (exams in adjacent anatomic regions). Notification of previous same studies across an HIE requires mapping of local site CT codes to a standard terminology for exam names (such as Logical Observation Identifiers Names and Codes [LOINC]) to show that two studies with different local codes and descriptions are equivalent. Notifying of prior similar or proximate CTs requires an additional mapping of exam codes to anatomic regions, ideally coded by an anatomic terminology. Several anatomic terminologies exist, but no prior studies have evaluated how well they would support an alerting use case.

Objective: The aim of this study was to evaluate the fitness of five existing standard anatomic terminologies to support similar or proximate alerts of an HIE-based prior CT alerting system.

Methods: We compared five standard anatomic terminologies (Foundational Model of Anatomy, Systematized Nomenclature of Medicine Clinical Terms, RadLex, LOINC, and LOINC/Radiological Society of North America [RSNA] Radiology Playbook) to an anatomic framework created specifically for our use case (Simple ANatomic Ontology for Proximity or Similarity [SANOPS]), to determine whether the existing terminologies could support our use case without modification. On the basis of an assessment of optimal terminology features for our purpose, we developed an ordinal anatomic terminology utility classification. We mapped samples of 100 random and the 100 most frequent LOINC CT codes to anatomic regions in each terminology, assigned utility classes for each mapping, and statistically compared each terminology's utility class rankings. We also constructed seven hypothetical alerting scenarios to illustrate the terminologies' differences.

Results: Both RadLex and the LOINC/RSNA Radiology Playbook anatomic terminologies ranked significantly better ($P<.001$) than the other standard terminologies for the 100 most frequent CTs, but no terminology ranked significantly better than any other for 100 random CTs. Hypothetical scenarios illustrated instances where no standard terminology would support appropriate proximate or similar alerts, without modification.

Conclusions: LOINC/RSNA Radiology Playbook and RadLex's anatomic terminologies appear well suited to support proximate or similar alerts for commonly ordered CTs, but for less commonly ordered tests, modification of the existing terminologies with concepts and relations from SANOPS would likely be required. Our findings suggest SANOPS may serve as a framework for enhancing anatomic terminologies in support of other similar use cases.

(JMIR Med Inform 2017;5(4):e49) doi:10.2196/medinform.8765

KEYWORDS

tomography, x-ray computed; health information exchange; radiation dosage; terminology; anatomy, regional

Introduction

Background

Computed tomography (CT) use has grown dramatically in recent years [1,2] and, because CT typically delivers higher radiation doses than conventional x-rays, there are concerns about appropriateness of utilization and the risks of cumulative radiation exposure [1,3-5]. Prior work by our group showed many patients underwent the same CT exam at more than one site within a health information exchange (HIE); some were likely duplicate studies and possibly avoidable [6]. Other authors have also shown that many CTs are duplicative and may be unnecessary [1,7-12].

An HIE-based prior CT alerting system may reduce avoidable CT imaging by notifying ordering clinicians of prior relevant studies when a repeat study is ordered [9,10,13,14]. For maximal effectiveness, a system would alert not only for prior *same* CTs (exams mapped to the exact same code from an exam name terminology) but also for *similar* CTs (exams in the same anatomic region but with different CT protocols and mapped to different exam name terminology codes) and anatomically *proximate* CTs (exams in adjacent anatomic regions). For example, in such an alerting system, a clinician ordering a head CT without intravenous contrast would be alerted not only to the same prior exam but also to a prior *similar* head CT with contrast study or a prior *proximate* neck CT performed anywhere within an HIE. Alerts for prior *proximate* CTs may be beneficial as scans often extend to tissues beyond the nominal scan range [15-17].

Alerting based on the existence of other prior imaging modalities (magnetic resonance imaging [MRI], ultrasound, plain films, and nuclear medicine) may also have utility in the decision to order a CT or other new imaging study. Although inclusion of all imaging modalities is the ultimate goal in our alerting system, we decided to start first with CT studies because of their frequency of use, cost, and potential impact on patient safety because of the relatively higher amount of radiation.

Notification of previous *same* studies across an HIE requires mapping of local site CT codes to a standard terminology for exam names (such as Logical Observation Identifiers Names and Codes [LOINC]) to show that two studies with different local codes and descriptions are equivalent (eg, "CT (-) head" at one site and "CT brain w/o contrast" at another) [18]. Notifying of prior *similar* or *proximate* CTs requires an additional mapping of LOINC CT codes to anatomic regions, ideally coded by an anatomic terminology. Several anatomic terminologies exist, including the anatomic hierarchy contained in LOINC's multiaxial hierarchy, but no prior studies have evaluated how well the concepts and relationships in these terminologies would support the alerting use case.

Objective

We sought to evaluate the fitness of five existing standard anatomic terminologies to support our alerting use case, including the (1) Foundational Model of Anatomy (FMA), (2) Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), (3) Radiological Society of North America's (RSNA) RadLex anatomic terminology, (4) anatomic hierarchy associated with LOINC, and (5) LOINC/RSNA Radiology Playbook's anatomic terminology by comparing them with an anatomic framework that we created specifically to meet the operational needs of our use case: "Simple ANatomic Ontology for Proximity or Similarity" (SANOPS).

We did not create SANOPS as a new anatomic framework for general use. Rather, we aimed to create a simple anatomic framework that could be implemented easily to support our alerting application and could also be used as a reference by which to compare the fitness for use of other existing terminologies. Our goals were to determine whether any of the existing terminologies could perform adequately in an unaltered state for our specific application and to characterize where they could be enriched, if necessary.

Significance

This study is a novel investigation of anatomic terminologies to support a prior CT alerting system. We previously described the pilot work to conceive of the SANOPS anatomic framework, which arose because we were designing an alerting system that accounted for similar and proximate CTs [19]. Other authors have compared the anatomic representations of SNOMED CT and FMA [20] and used anatomic terminologies to support radiology applications [21-23]. There are also reports of CT alerting systems implemented to notify ordering clinicians of an exam's appropriateness [24,25], as well as to notify technologists of possible excessive patient radiation dose [26]. These prior studies were not performed in the context of a prior CT alerting system and did not use an HIE of multiple organizations as the data source.

Methods

Overview

As an overview, to compare the existing standard terminologies, we previously developed an idealized anatomic framework (SANOPS) that we would use as a reference. In this study, we devised a terminology utility classification to provide a

quantitative assessment for the effort required to utilize and implement the standard anatomic terminologies for our use case compared with the SANOPS benchmark. We mapped a sampling of LOINC CT exam codes extracted from our regional HIE to anatomic regions in each terminology and assigned utility classes to each terminology for each mapping (described in detail below). We then statistically compared the utility classes of each terminology. We also constructed seven hypothetical alerting scenarios to further illustrate the terminologies' differences.

We performed our terminology comparison from December 2015 to February 2016. The versions of the terminologies were the latest available at the time and are indicated below.

Materials: Optimal Anatomic Terminology Features and SANOPS

Our development of SANOPS was guided by anatomic requirements and unique operational challenges required to support our specific use case of issuing prior *similar* or *proximate* alerts in an HIE-wide prior CT alerting system.

Anatomically, to support *similar* or *proximate* alerts, our ideal anatomic terminology would be organized by body regions rather than by organ systems and have information regarding containment of organs. For example, to issue a *similar* alert of a prior liver CT with a kidney CT order requires only that the terminology has information that the kidney and liver are contained in the same anatomic region, that is, the abdomen. To issue a *proximate* alert of a prior kidney CT with a pelvis CT order requires information that the kidney is contained in the abdomen region and that the abdomen and pelvis are specified as adjacent body regions. A terminology where kidneys are nested under the genitourinary system, without links to abdomen, would not be suitable. If a terminology is organized in an organ system hierarchy, then to be of any use it should also have information regarding the containment of organs within body regions. For the extremities, we would prefer a division into at least three relatively equivalent-sized proximate, mid, and distal anatomic regions. This would be done to help avoid clinically irrelevant proximate or similar alerts. For example, a prior right foot CT should not trigger an alert when a new right hip CT is ordered.

Operationally, issuing an order-time alert based on a prior exam performed at a different site within an HIE requires a complex series of steps. The local site where the order is placed must issue a Web-based communication with the HIE server that must match patient and exam records and determine whether same, similar, or proximate exams exist and then return the alert result and payload back to the local site. To be of any practical value, these steps must all be completed within fractions of a second. Any step that conserves computational resources and reduces query time would help ensure the successful firing of such an alert within the clinician's workflow. Given these special circumstances, it follows that body region organization and organ containment information within body regions should be expressed in the most direct and simple fashion as possible.

With these anatomic requirements and operational issues in mind, we designed the SANOPS anatomic framework with a relatively simplistic design. SANOPS divides the human body into 17 major regions (Figure 1). We used SANOPS by linking CT exam codes (which are LOINC codes in our application) to body regions that best subsume the region imaged. For example, kidney CT would be assigned to the abdominal region. Multiple anatomic identifiers can be assigned to exams that span more than one major body region. For example, an abdomen and pelvis CT exam as well as a lumbar spine CT exam would both be assigned to the abdomen and pelvic regions. Using SANOPS regions, similar alerts would be issued for a pair of different LOINC CT codes when they mapped to the same body regions. Proximate alerts would be issued for two CT codes that are assigned adjoining SANOPS regions, providing that they are not also assigned to the same region (in which case a similar alert would be issued).

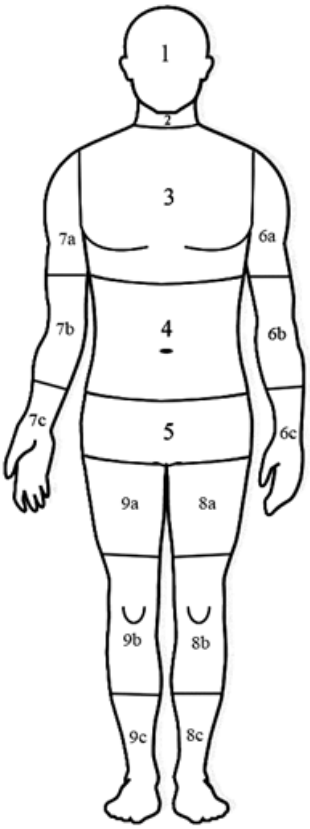
To avoid clinically irrelevant proximate or similar alerts in the extremities, SANOPS divides extremities into proximal, mid, and distal portions, with midshafts of long bones separating these regions. Therefore, extremity regions roughly correspond to the respective large joints plus adjacent portions of long bone shafts. This approach differs from the other anatomic terminologies that divide upper extremities into arm, forearm, and hand and wrist regions and the lower extremities into thigh, leg, and foot and ankle regions.

Our long-term goal is that SANOPS informs the use of standard terminologies to support HIE-wide prior exam alerting systems. More information regarding SANOPS and a translation table of SANOPS codes to other existing terminologies is available on the Internet [27].

Materials: Anatomic Terminologies

The FMA is an open source reference domain anatomic terminology of over 75,000 distinct anatomic concepts covering material objects from macroscopic to microscopic level, as well as nonmaterial entities (such as anatomic spaces) [28]. The FMA is both broader and more granular than extant anatomy texts or other terminologies. The FMA is an ontology in that it is "concerned with the representation of classes or types and relationships necessary for the symbolic representation of the phenotypic structure of the human body in a form that is understandable to humans and is also navigable, parseable, and interpretable by machine-based systems [29]." The FMA organizes its anatomic taxonomy in strict subsumption hierarchy. The FMA's anatomic structural abstraction also contains partonomy information that relates organ systems to constituent parts through *part_of*, *constitutional_part_of*, and *regional_part_of* links [29]. Many instances of FMA's *part_of* links relate organ systems to anatomic body regions. We used FMA version 4.4.1.

Figure 1. Seventeen major anatomic regions defined by our novel Simple ANatomic Ontology for Proximity or Similarity (SANOPS) terminology. The major anatomic regions and their respective codes are: 1-head, 2-neck, 3-chest, 4-abdomen, 5-pelvis, 6a-proximal left upper extremity (LUE), 6b-mid LUE, 6c-distal LUE, 7a-proximal right upper extremity (RUE), 7b-mid RUE, 7c-distal RUE, 8a-proximal left lower extremity (LLE), 8b-mid LLE, 8c-distal LLE, 9a-proximal right lower extremity (RLE), 9b-mid RLE, and 9c-distal RLE.



RadLex is a publicly available comprehensive clinical terminology providing a uniform standard for all radiology-related information [30]. The version used (3.13) contained over 68,000 terms organized in 15 main categories, including anatomic entity, clinical finding, and imaging modality. RadLex's anatomic terminology is derived from the FMA but employs simplified macroscopic terms relevant to radiology [31]. The RadLex Playbook (version 2.0 studied), comprising part of the comprehensive RadLex terminology, is a catalogue of radiology orderable exams, each given a unique "*RadLex* Playbook identifier (RPID)" defined by several attributes, including modality, body region, and anatomic focus. Body part(s) are indicated through body region and anatomic focus attributes that are represented by concepts in RadLex's anatomic terminology. Body region specifies the broad portion of the body that is imaged and anatomic focus indicates a more specific location (ie, liver CT has body region attribute "abdomen" and anatomic focus "liver") [32].

LOINC is a freely available international standard developed by the Regenstrief Institute Inc for tests, measurements, and documents. The version used (2.54) contained 78,959 terms with 798 CT exam codes [33]. Each radiology code has a system attribute (part) corresponding with the region or organ on which that exam was performed. LOINC anatomic regions are arranged hierarchically, although formal rules and relations between classes and subclasses are not currently defined.

The LOINC/RSNA Radiology Playbook was initially released in December 2015 as part of LOINC version 2.54 and was developed through collaboration between Regenstrief Institute and the RSNA and with support from the National Institute of Biomedical Imaging and Bioengineering [34]. This no-cost product combines and unifies useful aspects of LOINC Radiology and the RadLex Playbook. We used the initial release version that was limited to CT. Subsequent releases (most recent June 2017; part of LOINC version 2.61) included MRI, x-ray, ultrasound, nuclear medicine, mammography, and other imaging modalities [35]. Similar to RadLex, each CT exam in the LOINC/RSNA Radiology Playbook is defined by several attributes. Body part attributes are drawn from the RadLex anatomic hierarchy, with region imaged and imaging focus attributes in many instances also following RadLex's body region and anatomic focus model [35]. Where applicable, matching RadLex RPID codes are mapped to the LOINC codes.

SNOMED CT is a clinical health care terminology originally created by the College of American Pathologists and now maintained by SNOMED International (formerly, the International Health Terminology Standards Development Organisation) [36]. The version used (September 2015 release) is comprised of about 300,000 concepts, of which over 30,000 pertain to anatomic structures. SNOMED's anatomic hierarchy uses a Structure-Part-Entire (SEP) triplet to represent anatomic entities (eg, liver is represented through concepts of liver part, entire liver, and liver structure) [37]. This allows for anatomic relations to be expressed as subsumption (*is_a*) relations rather

than *part_of* relationships (eg, coronary artery structure *is_a* heart part *is_a* heart structure). SNOMED CT also contains direct partonomy information with anatomic structures related to constituent parts through *part_of* links that parallel the SEP relations [38]. By design, SNOMED CT's anatomic hierarchy is also a polyhierarchy, with many anatomic concepts having multiple parents or ancestors and children or descendants [38].

Methods: Mapping LOINC CT Exam Codes to Anatomic Regions in Anatomic Terminologies

Overview

The Office of the National Coordinator for Health IT has recommended LOINC as the best available standard for imaging procedures [39], and it has been used successfully in large HIEs [40]. Healthix, a large New York City area HIE currently working with our group, has chosen LOINC as the exam name terminology standard for CT exams and is mapping all local institution exam codes to LOINC. For these reasons, we chose LOINC as the exam name terminology to provide standardized CT exam codes and descriptions to which we assigned or mapped anatomic regions from our candidate anatomic terminologies, including LOINC's own core anatomic hierarchy. We extracted 100 random LOINC CT exam codes and the 100 most frequently performed LOINC CT exam codes among five sites in Healthix from March 1, 2009 to July 24, 2012. One of the authors (AOB), a board-certified radiologist, informatician, and domain expert manually mapped the anatomic regions from the LOINC name to the candidate anatomic terminologies. Our sampling approach enabled assessment of anatomic terminology performance over both a random sample and CTs that users would most frequently encounter.

Mapping Approach

LOINC CT exam codes were mapped to each anatomic terminology's region representing the nearest match to the region specified in the LOINC long common name. For example, "Head CT" and "Kidney CT" were respectively mapped to the "head" and "kidney" classes in FMA and were mapped to "head structure" and "kidney structure" concepts in SNOMED CT.

We made special considerations in mapping the LOINC CT exam codes to the LOINC, RadLex, and LOINC/RSNA Playbook anatomic terminologies because they are parts of larger comprehensive terminologies that include standardized codes and names or descriptions for CT exams and already have anatomic region attributes embedded with their CT exam names.

We mapped LOINC CT exam codes to regions in LOINC's core anatomic terminology by linking to the anatomic *system* attribute specified in the LOINC fully specified name [34]. Similarly, we linked LOINC CT exam codes to anatomic regions in the LOINC/RSNA Playbook anatomic terminology through the *region imaged* and *imaging focus* attributes contained in that terminology's distribution file. We also leveraged attribute relationships in the LOINC/RSNA Playbook to link LOINC CT exam codes to RadLex's core anatomic terminology through mappings between LOINC CT codes and RadLex Playbook RPID codes. If a LOINC code had no matching RPID, one of the authors (AOB) chose the closest matching RPID through manual review of the RadLex Playbook. Once we identified the corresponding RadLex Playbook RPID, we mapped LOINC Codes to each RPID's *body region* and *anatomic focus* attributes.

To further illustrate our method of anatomic mapping, consider the "Liver CT" exam (LOINC code 24815-3). In LOINC's anatomic terminology, we use "Abdomen>Liver" as the anatomic region because this composite element is the *system* attribute for this LOINC exam code. In both the LOINC/RSNA Radiology Playbook and RadLex anatomic terminologies, the anatomic mappings were to *body region* of "abdomen" and *anatomic focus* of "liver," as these were specified as attributes of the exam code in both terminologies. For FMA, the exam was mapped to *anatomic class* of "liver." For SNOMED CT, the exam code was mapped to *anatomic concept* of "liver structure."

Evaluation With Anatomic Terminology Utility Classification

Anatomic Terminology Utility Classification Features

Once a LOINC CT exam code was mapped to anatomic regions in each standard anatomic terminology, a terminology utility class was assigned to each mapping. We devised an anatomic terminology utility classification to provide a structured assessment of the effort required to utilize and implement the standard anatomic terminologies in an unmodified state for our use case. It is an ordinal sliding scale from 1 to 5. The criteria used to assign a terminology utility class is summarized in Table 1 and discussed in detail below. Using SANOPS as a benchmark, the class assignment is based on an approximation of terminology modifications or additional computing steps required to use the standard terminology to support an appropriate *similar* or *proximate* alert; a class of "1" is given if no additional computing steps above those used with SANOPS are necessary for a terminology to fully support an appropriate alert in its unaltered state; and a class of "5" is assigned if a terminology requires a large amount of computing resources or modifications. Utility class assignment was performed manually by AOB and then reviewed and validated by the other coauthors.

Table 1. Summary of terminology utility classification scale for mapping of each anatomic terminology to Logical Observation Identifiers Names and Codes (LOINC) coding.

Class	Criteria
1	The anatomic region specified is a body region (not an anatomic focus) that maps to a region in SANOPS[a] *or* if an anatomic focus, the corresponding body region attribute from the candidate terminology, maps to a SANOPS region
2	Anatomic focus nested under a SANOPS body region in uniform relation without other alternative path(s)
3	Polyhierarchy with uniform type ("is-a" or "part-of") edge links from anatomic focus concept to SANOPS body region but also with other paths to other superclasses bypassing the SANOPS body region
4	No link from anatomic focus to SANOPS body region through single type of edge relation
5	Anatomic focus concept not nested under SANOPS body region

[a]SANOPS: Simple ANatomic Ontology for Proximity or Similarity.

Anatomic Terminology Utility Classification: Class of "1"

A class of "1" was given when the anatomic region in the target terminology required no further computing steps above those if SANOPS were used to support a prior *similar* or *proximate* exam alert. This was satisfied in the following conditions:

- When the anatomic region specified by the LOINC CT exam name was equivalent to a SANOPS-specified major body region. For example, mappings of "Abdomen and Pelvis CT" in all five candidate terminologies were assigned "1" because all have concepts for abdomen and pelvis and there is a one-to-one correspondence of regions specified in the exam name with anatomic concepts in each terminology.
- For LOINC, RadLex, and LOINC/RSNA Playbook, when the *body region* or *imaged region* attribute specified by the exam code corresponded with one of SANOPS major body regions. For example, for a "Neck vessel CT angio" a class of "1" was given to RadLex and the LOINC/RSNA Radiology Playbook as "neck" is specified as a body region attribute in exam codes of both terminologies.
- In the extremities, when the anatomic region specified by the LOINC CT exam name was a major joint and the terminology had a matching concept for the specified joint. For example, mappings of "Right Knee CT" in all five candidate terminologies received a class rank of "1" because all have "right knee" anatomic concepts. The rationale for this assignment is that SANOPS concepts of proximal, mid, and distal regions of the extremities roughly corresponded with large extremity joints plus adjacent long bone shafts.
- If the CT exam description specified an entire extremity only as the anatomic region imaged and the terminology had a corresponding anatomic concept for the entire extremity.

Anatomic Terminology Utility Classification: Class of "2"

A terminology utility class of "2" was given to a target terminology if the anatomic focus specified by the CT exam code was nested directly under a major body region in SANOPS via uniform relationships and without other alternative path(s). For example, for a "CT angio abdomen," LOINC's anatomic terminology was assigned a class of "2" as the "abdominal vessels" anatomic region was nested directly under "abdomen." The rationale for this assignment is that this relation can be expressed in a "look-up" table and can be queried simply and quickly with only slightly more computing time and resources required over using SANOPS alone.

Anatomic Terminology Utility Classification: Class of "3"

A class of "3" was given to a target terminology in cases of a polyhierarchy where there were uniform type edge links from anatomic focus concept specified in the CT code to a body region corresponding with one of the SANOPS body regions, but also, there were paths to other superclasses bypassing the SANOPS body regions. For example, SNOMED CT ranked a "3" for "Temporal bone CT," as there were direct " *is_a* " links to the "head" through some paths, but there were also other " *is_a* " paths linking to the skeletal system, bypassing "head." The rationale for this assignment is that although this class-subclass or parent-child relation can also be expressed in a simple "look-up" table, the presence of multiple parents can lead to errors in classification, and description logic reasoners may need to be applied to ensure and verify that all anatomic class relations are expressed prior to implementation.

Anatomic Terminology Utility Classification: Class of "4"

A class of "4" was given to a target terminology when multiple different types of edge relations were necessary to link back to major anatomic regions. For example, FMA ranked a "4" for "Esophagus CT" because reaching the body region of chest requires first traveling down *has_part* link to thoracic esophagus and then back through *part_of* links to arrive at chest. The rationale for this class assignment is that to link back to major anatomic region through multiple different edge links would likely require a relatively complex algorithm and would require considerably more computer resources and processing time over using SANOPS alone.

Anatomic Terminology Utility Classification: Class of "5"

A class of "5" was given if there are no relationships to a major body region in the terminology from the anatomic focus specified by the exam code. For example, a class of "5" was assigned to LOINC and RadLex's terminologies for "Cervical spine CT" because each lacked links from "cervical spine" to "neck" and in neither was neck specified as a body region attribute by the exam code. The rationale for this class assignment is that the terminology could not be used in its unmodified state to support our use case. Special modifications would be needed to link the anatomic focus to the major SANOPS body region.

Analysis of Anatomic Terminology Utility Classes

Descriptive statistics were performed on the anatomic terminology utility classes for each terminology. As the data were ordinal, median and mode terminology utility classes for each of the five candidate terminologies were calculated. Mean classes and standard deviations for each of the five candidate terminologies were also calculated, although these are less informative for ordinal data. The Kruskal-Wallis test, a nonparametric analog of analysis of variance, was used to assess for a statistically significant difference in at least one utility class compared with remaining terminologies. The Wilcoxon Rank-Sum test, a nonparametric analog of the student *t* test, was used to assess for a significant difference between the classes of each pair of two candidate terminologies. Both tests used a level of significance (α) of .05 and were performed separately for the 100 random and the 100 most frequent LOINC codes.

Methods: Hypothetical Firing Scenarios

We devised seven hypothetical clinical cases that specified current and previous exams for which *proximate* or *similar* alerts should be fired based on anatomic location. These hypothetical scenarios were purposely selected to illustrate the differences between the anatomic terminologies in supporting our use case. We tested each anatomic terminology to see if, as presently constructed, *proximate* or *similar* alerts would appropriately fire. We defined a *proximate* region as the body region(s) adjacent to the body region hypothetically being ordered, as specified in SANOPS. For example, the "neck" region has *proximate* regions of "head" and "chest." If the current and prior exams in our hypothetical scenarios are in *proximate* regions, then a *proximate* alert should be fired. A *similar* alert would be fired if the current and previous exams had different LOINC codes but mapped to the same anatomic region.

Results

Anatomic Terminology Utility Scale Scoring: 100 Random LOINC CT Codes

Descriptive statistics for the anatomic terminology utility scale classes for the five terminologies mapping to anatomic regions for 100 random LOINC codes are given in Table 2. Mean anatomic terminology utility class ranks ranged from 1.82 for RadLex to 2.44 for LOINC, suggesting that moderate modifications and/or additional computing steps are required over SANOPS for these standard terminologies to support prior similar or proximate alerts. Kruskal-Wallis analysis showed that there was no statistically significant difference in rank of the five candidate terminologies (*P*=.30).

Table 2. Descriptive statistics of the anatomic terminology utility classes for each anatomic terminology's mapping to 100 random Logical Observation Identifiers Names and Codes (LOINC) codes. There was no statistical significance in mean terminology utility class rank.

Descriptive statistic	LOINC[a]	LOINC/RSNA[b]	RadLex	FMA[c]	SNOMED CT[d]
Median terminology utility class	1	1	1	1	1
Mode terminology utility class	1	1	1	1	1
Mean terminology utility class	2.44	2.18	1.82	2.22	2.03
Standard deviation of mean	1.83	1.75	1.37	1.56	1.32
Kruskal-Wallis mean rank class	267.53	245.75	229.66	258.16	251.40

[a]LOINC: Logical Observation Identifiers Names and Codes.

[b]RSNA: Radiological Society of North America.

[c]FMA: Foundational Model of Anatomy.

[d]SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

Table 3. Descriptive statistics of the anatomic terminology utility classes for each anatomic terminology's mapping to 100 most frequent Logical Observation Identifiers Names and Codes (LOINC) codes.

Descriptive statistic	LOINC[a]	LOINC/RSNA[b,c]	RadLex[c]	FMA[c,d]	SNOMED CT[c,e]
Median terminology utility class	2	1	1	2	2
Mode terminology utility class	2	1	1	2	2
Mean terminology utility class	2.6	1.5	1.38	2.29	2.19
Standard deviation of mean	1.85	1.22	1.13	1.44	1.31
Kruskal-Wallis mean rank class	293.16	203.66	188.18	285.65	281.88

[a]LOINC: Logical Observation Identifiers Names and Codes.

[b]RSNA: Radiological Society of North America.

[c]LOINC/RSNA Playbook and RadLex terminologies both had significantly lower class ranks compared with other terminologies by Wilcoxon Rank-Sum analysis.

[d]FMA: Foundational Model of Anatomy.

[e]SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

Anatomic Terminology Utility Scale Scoring: 100 Most Frequent LOINC CT Codes

Descriptive statistics for the anatomic terminology utility scale classes for the five terminologies mappings to anatomic regions for the 100 most frequent LOINC codes are given in Table 3. Kruskal-Wallis analysis shows a significant difference in at least one terminology mean class rank from another in the group. Wilcoxon Rank-Sum analysis shows that both RadLex and LOINC/RSNA terminologies ranked significantly lower (and therefore closer to our ideal SANOPS terminology) compared with the other terminologies (P<.001 for both), but there was no statistical difference in the ranking of the RadLex and LOINC/RSNA terminologies compared with each other (P=.15). Low mean terminology utility class ranks for RadLex and the LOINC/RSNA Playbook suggest that few, if any, modifications or additional computing steps above those used for SANOPS would be necessary to use these terminologies to support our use case in most instances where an alert would involve a frequently performed exam. Higher utility ranks for the LOINC, SNOMED, and FMA suggest that more computer resources and/or terminology modification would be necessary to employ these terminologies for our cases.

Hypothetical Alert Firing Scenarios

Table 4 shows seven illustrative scenarios where either *similar* or *proximate* alerts should be issued based on the anatomic location of current and previous exams. For each exam pair, the table indicates which anatomic terminologies would fire alerts based on the mappings and anatomic regions in each terminology.

In the first three scenarios, appropriate alerts would be fired using all terminologies (*similar* in the first example and *proximate* in the second and third). All terminologies have concepts for the regions specified by the exam codes, and these regions correspond with major anatomic SANOPS regions.

In the fourth scenario (prior "Head CT" and current "Paranasal sinus CT"), only LOINC/RSNA Playbook and RadLex would issue appropriate *similar* alerts. In the LOINC/RSNA Playbook and RadLex terminologies, "head" is specified as the *region imaged* and *body region* attribute, respectively. In LOINC, "paranasal sinuses" is nested under "skeletal system," bypassing "head." In FMA, "paranasal sinuses" is nested under "anatomic spaces," bypassing "head." In SNOMED CT, "paranasal sinuses" are nested under "head" in some hierarchies but bypass "head" in others; the "head" or "paranasal sinus" parent or child relation can be expressed in a "look-up" table, but this would require an additional computing step over using SANOPS alone.

In the fifth scenario (prior "Liver CT" and current "Kidney CT"), *similar* alerts are issued only with RadLex. "Abdomen" is the specified *body region* attribute of the both liver and kidney CT exam codes in RadLex. In the FMA, LOINC, and LOINC/RSNA Playbook anatomic terminologies, there are links from "liver" to "abdomen" but no links from "kidney" to "abdomen." SNOMED CT has links from both "liver" and "kidney" to abdomen but also divergent links bypassing "abdomen."

In the sixth scenario (prior "Cervical spine CT" and current "Neck CT"), no alert would be fired using any standard terminology in an unaltered state.

Table 4. Hypothetical alert firing examples. Exam pairs where either *similar* or *proximate* alerts should be issued based on anatomic locations. Check marks (✓) note where appropriate alerts would fire for each exam pair using the anatomic terminology specified in column heading in its unmodified state.

Alert scenario	Prior exam	Current exam	LOINC[a]	LOINC/RSNA[b]	RadLex	FMA[c]	SNOMED CT[d]
1	Head CT[e] with intravenous (IV) contrast	Head CT without IV contrast	✓	✓	✓	✓	✓
2	Head CT with IV contrast	Neck CT without IV contrast	✓	✓	✓	✓	✓
3	Elbow-bilateral CT without contrast	Shoulder-right CT with contrast IV	✓	✓	✓	✓	✓
4	Head CT with IV contrast	Paranasal sinuses CT without IV contrast		✓	✓		
5	Liver CT	Kidney CT without and with contrast IV			✓		
6	Cervical spine CT	Neck CT without IV contrast					
7	Esophagus CT	Chest CT					

[a]LOINC: Logical Observation Identifiers Names and Codes.

[b]RSNA: Radiological Society of North America.

[c]FMA: Foundational Model of Anatomy.

[d]SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

[e]CT: computed tomography.

In LOINC, LOINC/RSNA Playbook, and RadLex, the major *body region* is "cervical spine" rather than "neck," and there are no links from "cervical spine" to "neck." FMA and SNOMED CT both contain links from "cervical spine" to the "neck" region but also links bypassing neck.

In the seventh scenario (prior "Esophagus CT" and current "Chest CT"), no alert would be fired using any standard terminology in an unaltered state. "Esophagus" is not nested under any *body region* in LOINC, LOINC/RSNA Playbook, or RadLex terminologies. In SNOMED CT, there are links from "esophagus" to "chest," but there are also edge links bypassing "chest." In FMA, there are circuitous edge links from "esophagus" to "chest" involving a mix of *has_part* and *part_of* links, necessitating a custom algorithm to appropriately fire a *similar* alert.

Discussion

Principal Findings

Our analysis of anatomic terminology utility classes for the 100 most frequent exam codes shows that RadLex and the LOINC/RSNA Radiology Playbook terminologies outperformed the other terminologies for our use case; however, our analysis for the 100 random LOINC codes showed no statistically significant difference in the performance of candidate standard terminologies with a range of utility class ranks of 1.82 to 2.44. Our analysis suggests the LOINC/RSNA Radiology Playbook and RadLex's anatomic terminologies are suitable to support *proximate* or *similar* alerts for the most frequently performed CTs. The standard anatomic terminologies, as constructed at the time of this analysis, may have difficulties supporting our use case for uncommon CTs. Using the standard anatomic terminologies for issuing similar or proximate alerts would likely require the use of accessory look-up tables, modification of hierarchical relations, and/or application of SANOPS concepts and rules.

Our hypothetical test alerting scenarios illustrated how differences in the terminologies' modeling affect each terminology's fitness to support the alerting use case. In particular, the scenarios where no standard terminology supported appropriate alerts were selected to illustrate the difficulties of using existing terminologies, as is, for this use case.

The primary purpose of this study was to determine whether any of the standard anatomic terminologies in an unaltered state could approximate the performance of SANOPS and could potentially support a prior CT alerting system. We also wanted to assess and compare each of the standard terminologies to ascertain the effort required to adapt them for our use case. From our analysis, given the close approximation of utility classes for the LOINC/RSNA Playbook and RadLex anatomic frameworks to SANOPS for the most frequently performed CT exams, considerably less effort would be required to adapt these terminologies in their unmodified state for our use as compared with FMA or SNOMED CT.

Presently, we are collaborating with developers of the LOINC/RSNA terminology standard. In part influenced by our feedback, they are revising the modeling of the *region imaged* attribute such that all exams are assigned one of 11 discrete values (head, neck, chest, breast, abdomen, pelvis, extremity, upper extremity, lower extremity, whole body, and unspecified) [34]. This new modeling will enable us to leverage *region imaged* attribute to support our use case. As SANOPS's partitioning of the extremities still differs from LOINC/RSNA, SANOPS extremity concepts will still be used to augment LOINC/RSNA until such a time that LOINC/RSNA can fully support our use case.

It should be noted that none of the five standard anatomic terminologies contained adjacency information between major body regions that we could utilize to support *proximate* alerts. Therefore, to use a standard terminology, we would have to use SANOPS model of proximity and adjacency to support an alerting system. FMA does have adjacency information expressed as *adjacent_to* and *bounded_by* relationship links. However, these relations are very granular (eg, esophagus *adjacent_to* thoracic aorta) and are not scalable to major body region adjacency relations. Additionally, adjacency relations are not expressed uniformly for all FMA concepts.

It should also be noted that many instances of FMA's partonomy relations link organ systems to the body regions that contain them through homogeneous *part_of* links, but these relations are not expressed with the consistency needed to fully support our case (see esophagus to chest example above). The lack of comprehensive partonomy relations contributes to FMA's overall higher utility class. SNOMED CT, by contrast, through its SEP relations had direct *is_a* links from organ systems to the body regions containing them in all observed instances. However, SNOMED CT's polyhierarchy and alternate divergent pathways bypassing the body regions containing the organ systems may result in errors in linking organs back to body regions and contributed to its overall higher utility class.

Future Considerations

In the near future, we plan to build a pilot alerting system and to expand it to encompass other imaging modalities such as MRI. We plan to use similar alerting rules for other modalities, as we have for CTs, to notify users of prior *same*, *similar*, or *proximate* exams. We anticipate that SANOPS concepts and rules can also be used to guide the utilization of any standard anatomic terminology to support *proximate* or *similar* alerts for other modalities.

Limitations

The need for ongoing terminology maintenance would be a drawback to the long-term use of SANOPS as a stand-alone terminology. The standard terminologies are actively managed to support other use cases, and have active user communities that can enable more generalizable knowledge and sharing of resources. Using SANOPS in an operational system would require an ongoing effort to link any new standardized exam descriptions from LOINC used in our HIE to a SANOPS region. However, SANOPS is relatively simple with only 17 anatomic regions. On its own, SANOPS requires little maintenance. Also, SANOPS was never intended to be a stand-alone terminology. In our system, SANOPS extremity concepts are currently being

used to augment anatomic concepts that extend beyond the current LOINC/RSNA Playbook.

The structure we chose for partitioning of extremities in SANOPS is not congruent with the standard anatomic terminologies and currently prevents direct mapping of SANOPS extremity concepts to "flattened" versions of these terminologies. However, SANOPS alerting rules in the extremities can be applied to standard terminologies by either mapping SANOPS extremity concepts to modified large extremity joint concepts in the standard terminologies or by partitioning the extremities by concepts already included in the standard terminology and then devising rules similar to SANOPS for *similar* or *proximate* alerts. For example, if a terminology contains concepts for "hip," "thigh," "knee," "leg," and "foot or ankle," one could use these concepts to partition the lower extremity into categories for grouping exams by anatomic location and then set up alerting rules accordingly.

Conclusions

Our analysis of the fitness of five standard anatomic terminologies to support *proximate* or *similar* alerts in a prior CT alerting system suggests that modifications of these terminologies based on our novel SANOPS anatomic framework may be necessary to fully support the use case. With increased interoperability and exchange of information across health systems, we foresee the need for anatomic frameworks to support *similar* or *proximate* alerts based on anatomic location. Our work with SANOPS may serve as guidance on methodology for using any terminology to support a prior imaging exam alerting system. Our evaluation may also inform the future assessment and use of these anatomic terminologies in other clinical applications.

Acknowledgments

This project was supported by grant number 1R01LM012196-01 from the National Library of Medicine. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the US Department of Health and Human Services or any of its agencies.

Conflicts of Interest

One of the authors (DJV) reports grants from RSNA during the conduct of the study, personal fees from Icahn School of Medicine at Mt. Sinai, and grants from US National Library of Medicine outside the submitted work. The other coauthors have no conflicts of interest to disclose.

References

1. Griffey RT, Sodickson A. Cumulative radiation exposure and cancer risk estimates in emergency department patients undergoing repeat or multiple CT. AJR Am J Roentgenol 2009 Apr;192(4):887-892. [doi: 10.2214/AJR.08.1351] [Medline: 19304691]
2. Smith-Bindman R, Lipson J, Marcus R, Kim KP, Mahesh M, Gould R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med 2009 Dec 14;169(22):2078-2086 [FREE Full text] [doi: 10.1001/archinternmed.2009.427] [Medline: 20008690]
3. Amis Jr ES, Butler PF, Applegate KE, Birnbaum SB, Brateman LF, Hevezi JM, American College of Radiology. American College of Radiology white paper on radiation dose in medicine. J Am Coll Radiol 2007 May;4(5):272-284. [doi: 10.1016/j.jacr.2007.03.002] [Medline: 17467608]
4. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. N Engl J Med 2007 Nov 29;357(22):2277-2284. [doi: 10.1056/NEJMra072149] [Medline: 18046031]
5. Shah KH, Slovis BH, Runde D, Godbout B, Newman DH, Lee J. Radiation exposure among patients with the highest CT scan utilization in the emergency department. Emerg Radiol 2013 Dec;20(6):485-491. [doi: 10.1007/s10140-013-1142-8] [Medline: 23852432]
6. Slovis BH, Lowry T, Delman BN, Beitia AO, Kuperman G, DiMaggio C, et al. Patient crossover and potentially avoidable repeat computed tomography exams across a health information exchange. J Am Med Inform Assoc 2017 Jan;24(1):30-38. [doi: 10.1093/jamia/ocw035] [Medline: 27178985]
7. Wiest PW, Locken JA, Heintz PH, Mettler FA. CT scanning: a major source of radiation exposure. Semin Ultrasound CT MR 2002 Oct;23(5):402-410. [Medline: 12509110]
8. Broder J, Bowen J, Lohr J, Babcock A, Yoon J. Cumulative CT exposures in emergency department patients evaluated for suspected renal colic. J Emerg Med 2007 Aug;33(2):161-168. [doi: 10.1016/j.jemermed.2006.12.035] [Medline: 17692768]
9. Gupta R, Greer SE, Martin ED. Inefficiencies in a rural trauma system: the burden of repeat imaging in interfacility transfers. J Trauma 2010 Aug;69(2):253-255. [doi: 10.1097/TA.0b013e3181e4d579] [Medline: 20699732]
10. Bailey JE, Pope RA, Elliott EC, Wan JY, Waters TM, Frisse ME. Health information exchange reduces repeated diagnostic imaging for back pain. Ann Emerg Med 2013 Jul;62(1):16-24. [doi: 10.1016/j.annemergmed.2013.01.006] [Medline: 23465552]

11. Bailey JE, Wan JY, Mabry LM, Landy SH, Pope RA, Waters TM, et al. Does health information exchange reduce unnecessary neuroimaging and improve quality of headache care in the emergency department? J Gen Intern Med 2013 Feb;28(2):176-183 [FREE Full text] [doi: 10.1007/s11606-012-2092-7] [Medline: 22648609]
12. Vest JR, Grinspan ZM, Kern LM, Campion Jr TR, Kaushal R, HITEC Investigators. Using a health information exchange system for imaging information: patterns and predictors. AMIA Annu Symp Proc 2013 Nov 16;2013:1402-1411 [FREE Full text] [Medline: 24551416]
13. Lammers EJ, Adler-Milstein J, Kocher KE. Does health information exchange reduce redundant imaging? Evidence from emergency departments. Med Care 2014 Mar;52(3):227-234. [doi: 10.1097/MLR.0000000000000067] [Medline: 24374414]
14. Frisse ME, Johnson KB, Nian H, Davison CL, Gadd CS, Unertl KM, et al. The financial impact of health information exchange on emergency department care. J Am Med Inform Assoc 2012;19(3):328-333 [FREE Full text] [doi: 10.1136/amiajnl-2011-000394] [Medline: 22058169]
15. Boone JM, Cooper 3rd VN, Nemzek WR, McGahan JP, Seibert JA. Monte Carlo assessment of computed tomography dose to tissue adjacent to the scanned volume. Med Phys 2000 Oct;27(10):2393-2407. [doi: 10.1118/1.1312809] [Medline: 11099210]
16. Goldman LW. Principles of CT: radiation dose and image quality. J Nucl Med Technol 2007 Dec;35(4):213-25; quiz 226-8 [FREE Full text] [doi: 10.2967/jnmt.106.037846] [Medline: 18006597]
17. Campbell J, Kalra MK, Rizzo S, Maher MM, Shepard JA. Scanning beyond anatomic limits of the thorax in chest CT: findings, radiation dose, and automatic tube current modulation. AJR Am J Roentgenol 2005 Dec;185(6):1525-1530. [doi: 10.2214/AJR.04.1512] [Medline: 16304007]
18. Beitia AO, Kuperman G, Delman BN, Shapiro JS. Assessing the performance of LOINC® and RadLex for coverage of CT scans across three sites in a health information exchange. AMIA Annu Symp Proc 2013;2013:94-102 [FREE Full text] [Medline: 24551324]
19. Slovis BH, Lowry T, Delman B, Beitia AO, Kuperman G, DiMaggio C, et al. A novel anatomical semantic ontology for identification of anatomically proximate CTs using LOINC codes. In: AMIA 2015 Proceedings.: AMIA 2015 Annual Symposium; 2015 Presented at: AMIA 2015 Annual Symposium; November 14-18, 2015; San Francisco, CA p. 1693 URL: https://knowledge.amia.org/59310-amia-1.2741865/t005-1.2744350/f005-1.2744351/2229149-1.2744547/2229149-1.2744548
20. Bodenreider O, Zhang S. Comparing the representation of anatomy in the FMA and SNOMED CT. AMIA Annu Symp Proc 2006:46-50 [FREE Full text] [Medline: 17238300]
21. Wennerberg P, Schulz K, Buitelaar P. Ontology modularization to improve semantic medical image annotation. J Biomed Inform 2011 Feb;44(1):155-162 [FREE Full text] [doi: 10.1016/j.jbi.2010.12.005] [Medline: 21195207]
22. Kokciyan N, Turkay R, Uskudarli S, Yolum P, Bakir B, Acar B. Semantic description of liver CT images: an ontological approach. IEEE J Biomed Health Inform 2014 Jul;18(4):1363-1369. [doi: 10.1109/JBHI.2014.2298880] [Medline: 25014939]
23. Lacson R, Andriole KP, Prevedello LM, Khorasani R. Information from Searching Content with an Ontology-Utilizing Toolkit (iSCOUT). J Digit Imaging 2012 Aug;25(4):512-519 [FREE Full text] [doi: 10.1007/s10278-012-9463-9] [Medline: 22349993]
24. Ip IK, Raja AS, Gupta A, Andruchow J, Sodickson A, Khorasani R. Impact of clinical decision support on head computed tomography use in patients with mild traumatic brain injury in the ED. Am J Emerg Med 2015 Mar;33(3):320-325. [doi: 10.1016/j.ajem.2014.11.005] [Medline: 25572644]
25. Dunne RM, Ip IK, Abbett S, Gershanik EF, Raja AS, Hunsaker A, et al. Effect of evidence-based clinical decision support on the use and yield of CT pulmonary angiographic imaging in hospitalized patients. Radiology 2015 Jul;276(1):167-174 [FREE Full text] [doi: 10.1148/radiol.15141208] [Medline: 25686367]
26. Howard ME, McCollough CH, Leng S, Yu L, Bruesewitz MR. Use of CT dose notification and alert values in routine clinical practice. J Am Coll Radiol 2014 May;11(5):450-455. [doi: 10.1016/j.jacr.2013.12.017] [Medline: 24656790]
27. Icahn School of Medicine at Mount Sinai. New York, NY: Icahn School of Medicine at Mount Sinai; 2016. Shapiro Laboratory: SANOPS URL: http://labs.icahn.mssm.edu/shapirolab/research/sanops/ [WebCite Cache ID 6sdj8PS1C]
28. Structural Informatics Group. Sig.biostr.washington.edu. Seattle, WA: University of Washington School of Medicine Foundational model of anatomy URL: http://sig.biostr.washington.edu/projects/fm/AboutFM.html [accessed 2017-08-10] [WebCite Cache ID 6sca6AfdM]
29. Rosse C, Mejino Jr JL. A reference ontology for biomedical informatics: the foundational model of anatomy. J Biomed Inform 2003 Dec;36(6):478-500 [FREE Full text] [doi: 10.1016/j.jbi.2003.11.007] [Medline: 14759820]
30. RSNA. RadLex URL: http://www.rsna.org/RadLex.aspx [WebCite Cache ID 6scaQ2fdt]
31. Mejino JL, Rubin DL, Brinkley JF. FMA-RadLex: an application ontology of radiological anatomy derived from the foundational model of anatomy reference ontology. AMIA Annu Symp Proc 2008 Nov 6:465-469 [FREE Full text] [Medline: 18999035]
32. Radiological Society of North America. RSNA Informatics: RadLex. 2016. RadLex Playbook 2.2 User Guide URL: http://playbook.radlex.org/playbook-user-guide.pdf [WebCite Cache ID 6scampGAl]
33. LOINC. The international standard for identifying health measurements, observations, and documents URL: https://loinc.org/ [accessed 2017-08-11] [WebCite Cache ID 6sdgLDQJf]

34. LOINC. Regenstrief and the RSNA are working together to unify radiology procedures in LOINC and RadLex URL: https://loinc.org/collaboration/rsna/ [accessed 2017-08-11] [WebCite Cache ID 6sdhWWhYq]
35. McDonald C, Huff S, Deckard J, Armson S, Abhyankar S, Vreeman D. Logical Observation Identifiers Names and Codes (LOINC®) users' guide. Indianapolis (IN): Regenstrief Institute; 2017. URL: https://loinc.org/download/loinc-users-guide/ [accessed 2017-08-11] [WebCite Cache ID 6sdhgjgpj]
36. International Health Terminology Standards Development Organisation. 2014 Dec. SNOMED CT Starter Guide URL: http://doc.ihtsdo.org/download/doc_StarterGuide_Current-en-US_INT_20141202.pdf [WebCite Cache ID 6sdi6xe9b]
37. Schulz S, Romacker M, Hahn U. Part-whole reasoning in medical ontologies revisited--introducing SEP triplets into classification-based description logics. Proc AMIA Symp 1998:830-834 [FREE Full text] [Medline: 9929335]
38. SNOMED-CT style guide: body structures - anatomy. Copenhagen, Denmark: The International Health Terminology Standards Development Organisation (IHTSDO); 2008 Feb 07. URL: http://ihtsdo.org/fileadmin/user_upload/Docs_01/Copenhagen_Apr_2008/SNOMED_CT_Style_Guides/IHTSDO_Modeling_StyleGuide-Anatomy-20080207_v1-06.pdf [accessed 2017-11-10] [WebCite Cache ID 6ustiNZ9O]
39. Office of the National Coordinator for Health IT. Washington, DC: Office of the National Coordinator for Health IT; 2016. 2016 Interoperability Standards Advisory URL: https://www.healthit.gov/sites/default/files/2016_InteroperabilityStandardsAdvisoryFINAL.pdf [WebCite Cache ID 6sdj0tzr2]
40. Vreeman DJ, McDonald CJ. Automated mapping of local radiology terms to LOINC. AMIA Annu Symp Proc 2005:769-773 [FREE Full text] [Medline: 16779144]

Abbreviations

CT: computerized tomography
FMA: Foundational Model of Anatomy
HIE: health information exchange
LOINC: Logical Observations Identifiers Names and Codes
MRI: magnetic resonance imaging
RPID: RadLex Playbook identifier
RSNA: Radiological Society of North America
SANOPS: Simple ANatomic Ontology for Proximity or Similarity
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms

Edited by G Eysenbach; submitted 17.08.17; peer-reviewed by X Ling, D Newman, J Bailey; comments to author 20.09.17; revised version received 04.10.17; accepted 05.10.17; published 14.12.17.

Please cite as:
Beitia AO, Lowry T, Vreeman DJ, Loo GT, Delman BN, Thum FL, Slovis BH, Shapiro JS
Standard Anatomic Terminologies: Comparison for Use in a Health Information Exchange–Based Prior Computed Tomography (CT) Alerting System
JMIR Med Inform 2017;5(4):e49
URL: http://medinform.jmir.org/2017/4/e49/
doi:10.2196/medinform.8765
PMID:29242174

Review

Adopting Telemedicine for the Self-Management of Hypertension: Systematic Review

Michael Mileski[1*], MPH, MHA, MSHEd, DC; Clemens Scott Kruse[1*], MBA, MSIT, MHA, PhD; Justin Catalani[1*], BHA; Tara Haderer[1*], BHA

School of Health Administration, Texas State University, San Marcos, TX, United States

[*]all authors contributed equally

Corresponding Author:
Michael Mileski, MPH, MHA, MSHEd, DC
School of Health Administration
Texas State University
601 University Drive
San Marcos, TX, 78666
United States
Phone: 1 832 752 4419
Fax: 1 512 245 8712
Email: mileski@txstate.edu

Abstract

Background: Hypertension is a chronic condition that affects adults of all ages. In the United States, 1 in 3 adults has hypertension, and about half of the hypertensive population is adequately controlled. This costs the nation US $46 billion each year in health care services and medications required for treatment and missed workdays. Finding easier ways of managing this condition is key to successful treatment.

Objective: A solution to reduce visits to physicians for chronic conditions is to utilize telemedicine. Research is limited on the effects of utilizing telemedicine in health care facilities. There are potential benefits for implementing telemedicine programs with patients dealing with chronic conditions. The purpose of this review was to weigh the facilitators against the barriers for implementing telemedicine.

Methods: Searches were methodically conducted in the Cumulative Index to Nursing and Allied Health Literature Complete (CINAHL Complete) via Elton B Stephens Company (EBSCO) and PubMed (which queries MEDLINE) to collect information about self-management of hypertension through the use of telemedicine.

Results: Results identify facilitators and barriers corresponding to the implementation of self-management of hypertension using telemedicine. The most common facilitators include increased access, increase in health and quality, patient knowledge and involvement, technology growth with remote monitoring, cost-effectiveness, and increased convenience/ease. The most prevalent barriers include lack of evidence, self-management difficult to maintain, no long-term results/more areas to address, and long-term added workload commitment.

Conclusions: This review guides health care professionals in incorporating new practices and identifying the best methods to introduce telemedicine into their practices. Understanding the facilitators and barriers to implementation is important, as is understanding how these factors will impact a successful implementation of telemedicine in the area of self-management of hypertension.

(JMIR Med Inform 2017;5(4):e41) doi:10.2196/medinform.6603

KEYWORDS

hypertension; telemedicine; eHealth; mHealth; disease management

Introduction

Incentives

The Patient Protection and Affordable Care Act of 2010 (PPACA) has put a burden on health care facilities, forcing them to cut costs and focus on quality. We must find effective methods to keep patients with chronic diseases such as hypertension and diabetes from having to come to the facility directly, by allowing them the choice of being monitored from home. This is easily and effectively done through telemedicine, such as telephone-based medicine, electronic medicine, or videoconferencing [1]. Telemedicine has incentives for health care facilities and the patients who would participate. These incentives include increased access to rural areas, increased involvement of nurses, decreased involvement of doctors in menial tasks, potential cost-effectiveness, interactive behavior change, high patient satisfaction, and positive long-term health outcomes [1-5]. Although there are clear advantages and evidence showing that telemedicine improves outcomes of hypertension and other chronic illness not only in the short term but also in the long term, studies show that health care facilities and medical staff are skeptical regarding the adoption of telemedicine models because of the shift of responsibility to the provider to check up on the patient [6].

Incentives for cost cutting and improved outcomes exist through PPACA and Medicare reimbursement policies. By increasing the quality and patient satisfaction, higher reimbursement can be realized through federally funded programs. Telemedicine can be rapidly implemented and can easily negate some of the financial burdens in facilities throughout the United States by higher patient volumes and by more efficient use of patient-physician care time [1].

Identification and Definition of Key Terms

According to the World Health Organization, telemedicine is "the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of diseases and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities" [7]. The use of telemedicine has expanded vastly to include a "variety of applications and services using two-way video, email, smartphones, wireless tools and other forms of telecommunications technology" [2]. Telemedicine has the potential to impact the health care industry in profound ways with the constant creation and innovation of technology. Currently, the health care system relies on face-to-face communication to deliver care. Telemedicine offers a method to be utilized in conjunction with face-to-face communication with providers. It is not a separate specialty in the medical field because telemedicine is typically integrated in health care institutions within information technology or the delivery of care [2]. There is no distinction between the terms "telemedicine" and "telehealth," and they are considered interchangeable to encompass a variety of remote health care options [7]. Patient consultations via videoconferencing, transmission of still images, electronic health (eHealth) including patient portals, remote monitoring of vital signs, continuing medical education, consumer-focused wireless applications, and nursing call centers, among other applications, are all considered part of telemedicine and telehealth [7].

Hypertension Defined

Hypertension is defined as having abnormally high blood pressure [8]. Known as the "silent killer," hypertension is listed as one of the most important causes for premature death; it affects 1 billion people worldwide, with two-thirds found in developing countries [9]. These numbers are growing at an alarming rate, with an estimated 1.56 billion adults to be afflicted with hypertension by 2025 [9]. Hypertension is a chronic condition that occurs when blood is pumped through the arteries with excessive force [1]. Hypertension can lead to many health risks, including heart attacks, strokes, kidney failure, or other life-threatening health problems [1]. Causes of hypertension can include kidney fluid and salt balances, blood vessel structure, genetic causes, and environmental causes such as unhealthy lifestyles, obesity, and the use of certain medications [8]. In the United States, 1 of every 3 adults has hypertension and only about half of the population with hypertension has the condition under control [9]. This costs the nation US $46 billion each year in health care services and medications required for treatment and missed workdays [9].

Physician Dependence and Self-Management

Patients are highly dependent on their physicians for information on their health. Self-management refers to taking responsibility for one's own behavior and well-being. Implementing self-management in the health care setting is the start to educating patients of their current state of health and conditions affecting them. Educating not only improves the individual's knowledge of the condition but also allows for early detection of health problems and enables the individual to correctly self-titrate medications and allows for timely interventions [2,8].

Telemedicine Adoption Among Health Facilities

Facilities that have adopted models of telemedicine have shown better patient outcomes and satisfaction, higher patient volumes, and increased facility space to be used for other purposes. Although recent studies have not shown a direct short-term relationship to cost savings, they recognize that in the long-term, cost-effectiveness will be realized [5,10,11]. With the obesity rate in America at an all-time high, along with hypertension diagnoses on the rise, it is important to recognize the benefits that telemedicine offers to consumers. Some of these are sustainable intervention, long-term blood pressure control, improved patient knowledge and accountability, facilitated communication between the patient and the provider, and long-term cost-effectiveness [3,12]. The benefits of telemedicine are yet to be fully realized through research, and it could easily become a regulation to implement it as a tool, such as electronic health records or the International Classification of Diseases, Tenth Edition (ICD-10). It is in the best interests of health facilities to implement this technologically advanced medical care technique into their operations, allowing for a competitive advantage.

Telemedicine Impact

Although telemedicine offers numerous benefits, the benefits are not fully recognized at the beginning of implementation. With no governmental incentives existing in place for telemedicine, facilities are skeptical to take the leap into this new care technique. This review highlights the facilitators and barriers to implementing telemedicine techniques into health care facilities to identify the benefits that can be realized if adopted [2-4,6,11,12].

Rationale

The findings of this review will be useful to health administrators, physicians, nurses, and other stakeholders in facilities that are weighing the potential benefits and barriers of adopting a telemedicine policy into their organization. This review is also useful to patients. With the ability of technology to inform the public, patient awareness on the management of their own health has made them a major stakeholder in health care. By extending these findings to the public, increased awareness of the positive outcomes to the adoption of telemedicine might be the push to implementation that the health care facilities need.

Methods

Data Collection Process

Information for this review was collected through the use of two databases: Cumulative Index to Nursing and Allied Health Literature (CINAHL) Complete via Elton B Stephens Company (EBSCOhost) and PubMed (which queries MEDLINE). The search focused on self-management of hypertension through the use of telemedicine. The members of the research team reviewed the papers identified during the search and summarized data relative to this review. During successive independent reviews, members compared and discussed the papers and reasons for their inclusion in the study. Papers were included based on their discussion of facilitators and barriers to the self-management of hypertension via telemedicine. The members of the research team had full agreement on all papers included in this systematic review.

Sample

Research databases were queried from PubMed and CINAHL using search terms of ("telehealth" OR "Telemedicine" OR "eHealth" OR "mHealth" OR "information technology") AND ("self-management" OR "self management") AND ("hypertension" OR "high blood pressure" OR "elevated blood pressure"). Several exclusion criteria were also specified: duplicates were excluded; only academic journals were included; papers were in English only; human-based studies were only included; study protocols and designs were eliminated; and nongermane trials were excluded. Searches were limited to date of publication from 2010 to 2016 (n=14). This process is illustrated in Figure 1.

Figure 1. Preferred Reporting Items for Systematic Reviews flow diagram.

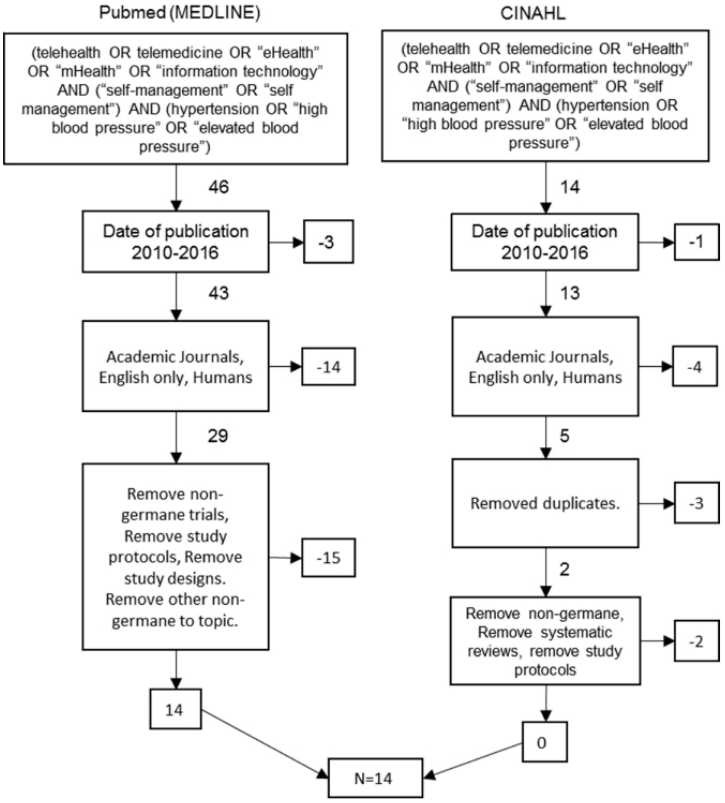


Table 1. Facilitators and barriers associated with the implementation of telehealth in the self-management of hypertension.

Authors	Facilitators	Barriers
McKoy et al [13]	Rapidly evolving technology	Self-management requires much support
	Decreased costs for providing care	Limited oversight, regulation, and guidelines
	Mobile phones can become electrocardiographs (EKG) or other diagnostic machines	Blurring of professional role of practitioners
	Increased access to care that patients otherwise may not receive	Potential for increased liability for practitioners
	Large application to medically underserved areas	Need for identification for ways to mitigate risk
	Ease in scheduling, communication, monitoring, and management for patients and practitioners	
Kumar et al [14]	Apps attached to practitioners have increased oversight of patients and ease of facilitation of care	Many available apps for mobile phones do not have a practitioner involved with them
		Apps available are not documented as valid ways to measure blood pressure
		No oversight to app production or effectiveness
		Lack of Food and Drug Administration approval
Kaambwa et al [3]	Cost-effective in the long run	Differences in results compared with other studies
	Adjusted life years gained	Varying results between men and women
	Markov model	
	Reduced blood pressure compared with usual care	
Maciejewski et al [12]	Sustainable intervention	No economic gains
	Long-term blood pressure control	Not many other studies to exhibit consistency
Wakefield et al [15]	Home-based	No significant difference between intervention and control group
	More timely changes	Information technology unlikely to lead to improved outcomes alone
	Targets remote treatment outcomes	Need for responsive clinical processes
	Effectiveness in the short term	
Shaw et al [16]	Improved patient outcome	Unclear long-term commitment
	Positive organizational culture	Added workload
	Evidence-based and nurse administered	Skeptical staff on positive outcomes
	Information technology infrastructure and support	
	Utilization of existing equipment and space	
Fitzner and Moss [4]	Increased access	Literacy level
	Interactive behavior change technology	Comfort level with technology
	Chronic care model	Security of personal health data
	Technological tools lead to improved patient health	Accuracy of information
	High access to telecommunication	Medicare's efforts to extend reimbursement to self-management training
	Ease and immediacy of communication	
	Convenience of home	
	Portable	
	Rapid growth of use in mobile phones	
	Effective, efficient, and affordable ways to reach and support minorities	

Authors	Facilitators	Barriers
Melnyk et al [10]	Patient Care Alignment Teams; stresses non-face-to-face interactions	Patients adherence to the guidelines/rules
	Facilitates individualized personal interaction	Self-management components to consider
	Remedies face-to-face intervention problems	
Piette, Marinec et al [5]	Increased access to health information between visits	Interactive voice response call completion rate declined
	Information for users	Technical challenges
	Delivered from long distances	
Jackson et al [17]	Medical and behavioral aspects addressed	
Piette, Datwani et al [18]	Access	Labor intensive and rarely available in low- and middle-income countries
	Cloud computing can make mobile health (mHealth) services more accessible	Lacks the resources to launch and maintain an mHealth service
		Relatively little collaborative work with patients' clinical teams
		Effort to educate providers
Wang et al [11]	Hypertension is a common reason for men to go to the doctor	Substantial time costs accumulated for nurses to prepare
	Different telephone interventions added with usual care	No long-term difference in results compared with usual care
	Costs may not significantly differ from that of usual care	Hypertension Intervention Nurse Telemedicine Study (HINTS) intervention was costly and time-consuming to deliver
		Unknown whether intervention generates other patient-centered outcomes or efficiencies in other aspects of medical care
Jones et al [6]	Participants valued additional information	Self-management hard to be maintained by participants
	Home blood pressure readings more natural	Borderline readings
	Greater control and more involvement	Self-titration
	Improvement of knowledge for the patient	Needs significant input from general practitioner

Data Analysis

Narrative summaries related to factors that influenced the adoption of telemedicine for the self-management of hypertension were extracted from each paper and included in Table 1. These, in turn, were grouped into larger recurring themes: either key determinants or impediments to success. The themes chosen were by consensus of the authors. Those chosen were agreed upon to be ones that provided overarching summary to the facilitators and barriers extracted. These themes were then divided into two affinity matrix tables for facilitators and barriers. Each table documents the themes, their citation occurrence, their frequency sum, and their frequency percentage.

Results

The findings were summarized into the facilitators and barriers table after the members of the research team chose papers to construct the systematic literature review. All identical papers were analyzed and unified before the table was generated. The members of the research team reevaluated the papers and determined facilitators and barriers in the self-management of hypertension via telemedicine. Results are summarized in Table 1. Papers are listed in the order of publication, with the most recent papers at the top.

During the course of the period chosen (2010-2016) for this systematic review, 112 authors published 14 works that specifically studied, analyzed, and discussed factors relating to the self-management of hypertension via telemedicine. Most of these works highlighted both facilitators and barriers, and only one highlighted only facilitators. Papers originated from multiple countries, and a total of 48 facilitators (55%) and 40 barriers (45%) were observed.

Discussion

Facilitators

About 17% more facilitators to implementation were noted than barriers (48:40). The authors of this review compared and grouped the facilitators and barriers into common themes. A total of 24 themes were noted between the two categories. A total of 13 facilitator theme categories and 11 barrier categories were identified. Table 2 illustrates and rank orders the themes from the facilitators based on their frequency of occurrence in the literature.

Table 2. Facilitating themes associated with the implementation of telemedicine in the self-management of hypertension.

Facilitators	Occurrences	Frequency (N=48) n (%)
Increased access	[4][a], [5][a], [13][a], [18]	7 (15)
Increase in health and quality	[3], [4], [12], [14], [15][a], [16]	7 (15)
Patient knowledge and involvement	[4], [5], [6][a], [17]	6 (13)
Technology growth with remote monitoring	[4][a], [13][a], [16], [18]	6 (13)
Cost-effectiveness	[3][a], [4], [11], [13]	5 (10)
Increased convenience and ease	[4][a], [13], [15]	4 (8)
Facilitates communication	[10][a], [16]	3 (6)
Natural readings	[3], [6], [12]	3 (6)
Personalized care	[10], [11][a]	3 (6)
Utilizing nurses	[16]	1 (2)
Portable	[4]	1 (2)
Timely	[15]	1 (2)
Utilizing space	[16]	1 (2)

[a]Denotes multiple occurrences in the same paper.

Facilitators mentioned most often in the literature were increased access [4,5,13,18] (multiple occurrences per article: 6,12,14) and increase in health and quality [3,4,12,14-16] (multiple occurrences per article: 10). These themes were identified 7 times each, out of 41 total occurrences (17% each). Regarding increased access, authors have noted facilitators where patients found themselves with access to care that they might otherwise not receive [13] and that technology could easily be adapted to use for medically underserved populations [13]. Furthermore, there was an overall increased access to care and telecommunication by the use of telemedicine [4]. The use of the technology also allowed for care to be delivered over long distances, which increased not only access to care but also access to health information [5,18].

Increases in health and quality facilitators were noted in areas that allowed practitioners to have increased oversight of their patients and an easing of the facilitation of care [14,16]. The telemedicine intervention was also noted to be sustainable [12] and could target treatment outcomes remotely [15], which allowed for increased effectiveness in treatment, a gain of adjusted life years, and improved patient outcomes [3,15,16].

The next two most identified themes were patient knowledge and involvement [4-6,17] (multiple occurrences per article: 18) and technology growth with remote monitoring [4,13,16,18] (multiple occurrences per article: 6,12). These themes were identified 6 times each, out of 41 total occurrences (15% each). Patient knowledge and involvement themes mostly focused on how patients could benefit from the use of this technology. Behavior changes [4,17] and additional information and greater control and involvement for patients [5,6] were noted. Increased knowledge for the patients empowered them to make better decisions and have more control of their own care [6,17]. There was also a "value-added" component to the use of telemedicine, as there was additional information available regarding the diagnosis to patients, which otherwise might not have been readily available [6].

Technology growth was also an important theme identified; the more this technology is used, the greater the applications that will be identified [13]. Everyday devices such as mobile phones can be utilized as an electrocardiograph (EKG) or other diagnostic machines [13]. Increased use also will lead to increased options in information technology infrastructure and support [16] and increased tools to use to promote health in patients [4]. This technology could also drive advances in mobile phones and cloud computing [4,18].

The cost-effective nature of the use of telemedicine in the treatment of hypertension was identified 5 times [3,4,11,13] (multiple occurrences per article: 8) (12%). Telemedicine has been shown to be effective, efficient, and affordable in reaching its target population [4]. Furthermore, it is shown to have decreased costs overall for providing care [13], despite the cost of the per-unit provision of care being the same [11]. The long-term use of this as an intervention is where most cost savings are seen [3]. Studies in the area of cost-effectiveness have been weak overall; however, it is believed that there has been a gross underestimation in this area, as the cost of technology has decreased over time [13]. Cost of self-management was shown to be higher overall in some cases, but as it was associated with an increase in quality of life for patients with a decrease in cardiovascular events, a net savings was noted [3]. These results are further verified in reduced overall costs for the care of patients with the use of telemedicine technologies, despite their increased front-end cost overall [4,11].

Increased convenience/ease of the use of telemedicine in the treatment of hypertension was identified 4 times [4,13,15] (multiple occurrences per article: 12) (10%). Patients were able to realize an ease in their ability to receive care from the convenience of their own homes [4,15]. This translated into not only easier access for patients and providers [18] but also more immediate communication between patients and providers [4]. Providers found significant benefits in the areas of scheduling ease, communications with patients, disease monitoring, and disease management [13].

Three theme areas were next tied with each other in incidence. Facilitates communication [10,16] (multiple occurrences per article: 13), natural readings [3,6,12], and personalized care [10,11] (multiple occurrences per article: 17) were all identified 3 times (7% each). An advantage to the use of telemedicine identified was that it alleviates issues regarding interventions in the face-to-face method [10]. Furthermore, communication is facilitated, as the use of telemedicine is able to create a positive organizational culture for patients and practitioners alike [16]. Also of note is that blood pressure readings in the home environment are more natural [6] and are not plagued by the stress of the medical office or "white coat syndrome." These more natural readings, which are facilitated because of the use of telemedicine, show patients to have decreased blood pressure readings versus normal in office care [13], which in turn assists in long-term blood pressure control [12]. This ties in with personalized care, as telemedicine facilitates individualized personalized interactions [10]. It also fosters care for those who might otherwise not go to the physician for treatment [11] and allows for different interventions outside the normal ones, which would come with usual care for hypertension [11].

The last four facilitators were each mentioned once in the identified papers. Utilizing nurses [16], portable [4], timely [15], and utilizing space [16] were all mentioned one time in the thematic review (2% each). Nurses were able to provide more timely care and intervention than would normally happen in the office environment [15,16]. The technology that is required is portable and can be carried with the patients anywhere they go [4], which allows for increased ability to give timely care [15]. An existing practice can additionally expand its reach and patient base, as the implementation of technology will not take more space resources, and it can utilize existing equipment in most cases [16]. This allows for an increased patient load and an increase in the amount of care given in a timely fashion.

Barriers

The barrier mentioned most often in the literature was lack of evidence [3,4,11,12,14,16] (multiple occurrences per article: 7,8), which was identified 9 times of 40 total incidences (23%). Lack of evidence was an area that showed many possible concerns. Telemedicine lacks oversight in the area of application production or effectiveness [14]. Additionally, it lacks Food and Drug Administration approval [14]. There is evidence to show that there are differences in effectiveness between men and women [3] and that there are other inconsistencies between the programs (and the research surrounding them) [12]. Due to these inconsistencies, staff seem to be skeptical about the use of technology in this fashion [16], as the accuracy of the information provided [4] and the validity of the results are in question [14]. Furthermore, inconsistencies also question the outcomes generated by the apps and the efficiency of the use of telemedicine to monitor hypertension [11].

The next most-cited barrier was self-management difficult to maintain [5,6,10,13] (multiple occurrences per article: 13,18), which was identified 7 times (18%). This barrier surrounds the ability of patients to support themselves using the technology. There is a large amount of support necessary for patients to be able to use and maintain a telemedicine program on their own [13]. There are also issues identified regarding patients adhering to rules and guidelines surrounding the use of the technology [10] and keeping up with these patients from the provider perspective [6]. The issues observed with the use of this platform include what to do with borderline readings [6] and patients self-titrating their treatments [6]. Getting patient buy-in and continual participation is in question [5]. Table 3 illustrates and rank orders the themes from the barriers based on their frequency of occurrence in the literature.

There are questions that remain regarding no long-term results and more areas to address [11,13,15] (multiple occurrences per article: 6,10) (15%). The lack of long-term studies on this topic has shown a difficulty in the fashioning of responsive clinical processes [15] and a lack of data to show comparison with normal management of hypertension [11]. In fact, one study showed no difference between a telemedicine group and a normally managed group [15]. Many practitioner issues with liability and the mitigation of risk have been raised with the use of telemedicine and its use in the self-management of hypertension [13], with potential for information technology issues causing the technology to not be reliable [15].

There is also a perception that the use of telemedicine will result in a long-term added workload commitment [11,16,18] (multiple occurrences per article: 11,16), as it was identified 5 times (12%). There are significant concerns with time to train staff [11] and increased efforts required to train providers on the new technology [11,18]. There are concerns over the perception that the technology is labor intensive and is only available in affluent countries [18]. Significant concerns were raised over added workload to already overburdened or time-strapped staff [10], increased educational efforts being required [18], and unclear time commitments to roll out and keep telemedicine programs implemented [10].

Table 3. Barrier themes associated with the implementation of telemedicine in the self-management of hypertension.

Barriers	Occurrences	Frequency (N=40) n (%)
Lack of evidence	[3][a], [4], [11], [12], [14][a], [16]	9 (23)
Self-management difficult to maintain	[5], [6][a], [10][a], [13]	7 (18)
No long-term results and more areas to address	[11], [13][a], [15][a]	6 (15)
Long-term added workload commitment	[11], [16][a], [18][a]	5 (13)
Costly	[4], [11], [12]	3 (8)
Technology challenges	[5], [18][a]	3 (8)
Significant input by general practitioner needed	[6], [13]	2 (5)
Variation with providers and systems	[13], [14]	2 (5)
Low health literacy level	[4]	1 (3)
Lack of comfort with technology	[4]	1 (3)
Security of data	[4]	1 (3)

[a]Denotes multiple occurrences in the same paper.

Another concern among providers is the cost of using such technology [4,11,12], as it was mentioned in 3 papers (8%). Payers for services (such as Medicare) have made little efforts to pay for services provided via telemedicine [4], and this is slowing implementation. Furthermore, practitioners are concerned over the cost of the delivery of such services possibly outstripping reimbursement [11] and that there might simply be no economic gain to providing services in such a fashion as via telemedicine [12]. Initial concerns that the cost for the use of the technology might be wasted also exist, as one study showed similar probabilities for inpatient admission in those who used technology versus those who did not [12]. An additional study also showed no significant gains in hypertension control with the addition of technology [11]. However, both studies [11,12] were specific to the Veterans Administration population, and findings could be specific to this particular population.

Technology challenges are also a significant barrier [5,18] (multiple occurrences per article: 16), as they are mentioned 3 times (8%). Practitioners are concerned that they lack the resources (technical and financial) to maintain an in-house mHealth service [18]. Simple technical challenges [5] and the perception that there will be little collaboration among clinical teams in the use of the technology [18] are also major concerns.

Several other barriers were found in the literature, which show concern for implementation on many levels. Significant input needed by general practitioner [6,13] was shown in two papers. The concerns here were that the implementation of such technology would blur professional roles in the treatment milieu [13] and that a fear of a significantly increased workload via input exists among these same practitioners [6]. A concern with variation with providers and systems [13,14] was also noted in two papers. There is limited oversight, regulation, and guidelines over telemedicine applications [13], and there is a fear that many applications exist with no oversight at all [14], medical or otherwise. The potential exists to put patients in an unsafe condition and/or without medical oversight and attention. Finally, there exists concern over low health literacy levels [4], lack of comfort with technology [4], and security of data [4]. Many patients already have a low health literacy, and adding technology to the mix simply complicates the already difficult to provide care [4]. These same patients have a lack of comfortability with technology [4], and there are concerns regarding the safety of personal health data [4].

Population

Hypertension is associated with multiple chronic conditions, including but not limited to diabetes and cardiovascular disease, and there is an ever-increasing need to enhance care to deliver to these individuals [4,13,15,17]. Telemedicine broadens access to people unable to reach health services easily [4,18]. Patients' involvement in health decisions will increase along with access to health information in between visits [6,18]. Remote data collection, monitoring, and cloud computing will allow for care to be delivered from a distance [3,4,14,18]. This allows doctors to attend to other patients' needs that cannot be addressed outside a health care facility. Using telemedicine to attend to the management of patients' health will vastly reduce problems occurring in face-to-face care [17].

Cost

Cost is often an argument against telemedicine use in health facilities; however, studies show that implementation such as a Tailored Case Management for Diabetes and Hypertension (TEACH-DM) system allows for rapid implementation at low cost [1]. Studies showing the cost-effectiveness of telemedicine point that there is a variation in savings within the short-term use of telemedicine and that this variation can be expected to be steady and to increase in cost-effectiveness over a 2-year period [3,4,14,15].

Although no direct government incentives exist for implementing a telemedicine system, the increase in quality and Medicare reimbursement shows steady increases in

reimbursement and patient volumes [10,11,15], and this should be ample incentive in itself to make the move toward this technologically advanced system of care. There are difficulties in understanding why telemedicine is not being heavily implemented throughout the United States despite its cost-effectiveness and research showing positive cost savings and patient outcomes in the long run [14]. However, this study can be used by facilities to raise awareness of the cost-saving potential that could result from the implementation of a telemedicine system.

Perceptions

Users' perceptions pose a great level of concern for facilities thinking of adopting a telemedicine program. Users need to understand the value in consistently using telemedicine to assist with self-management of hypertension. The benefits will not be evident immediately [13]. Having patients develop a positive perception will make administering a telemedicine program more desirable. Telemedicine has made it more convenient for users to keep track of their health; it allows users to take measurements from home and immediacy of communication [2,4,5,14]. By enabling the users to track and record their health status, they feel that they have an increase in knowledge with regard to understanding their health condition [5,6,13]. This helps to increase patient participation with their own health and modify behaviors that result in an unhealthy lifestyle [6,11].

Rejection of using telemedicine is not solely derived from the users' perceptions; it is heavily influenced by the organization deciding whether implementation is in its best interest. Cost plays a key role in conducting a cost-benefit analysis. Since there is little proven evidence suggesting success in gaining a financial return in using telemedicine for long-term conditions, organizations hesitate on its adoption [2,11,12]. Costs may pose an issue if organizations lack proper strategic planning initiatives and fail to adequately address the purpose for adopting telemedicine into their practice. The organization needs to align a telemedicine program in accordance with its mission, vision, and values. Upon alignment, there will be an increase in the quality of care provided to the patients, thus resulting in improved health. Telemedicine has been studied to show an improvement in the quality of life and health of patients who participated in the research studies [4,14]. Even with research that supports expanding the use of telemedicine, some clinical trials show no economic gains [12]; this poses a challenge for pushing the expansion of telemedicine in organizations.

Implementation

Implementation of a telemedicine program relies heavily on the vision being aligned throughout the organization. With cost-effectiveness and rapid implementation models such as TEACH-DM being available, the only barrier is uncertainty. With facilities moving toward implementation of telemedicine models, it is likely for this uncertainty to be demolished [2,9]. Poor planning has been an extreme issue because of the alignment of goals and vision required throughout the organization to be successful in a telemedicine implementation process [13,14]. When a health care facility implements telemedicine models, the benefits realized by the facility will quickly create a domino effect of implementation of telemedicine by other organizations. This is because of the high levels of competition in health care today, requiring facilities to consistently look for ways to reduce cost and increase quality.

Adopting a Telemedicine system

Practices may realize the benefits of using telemedicine in the self-management of hypertension after conducting a cost-benefit analysis. This will aid organizations in determining whether the venture is worth adoption. Qualifications for organizations to implement telemedicine programs, at a minimum, are alignment with the mission of the organization and the strategic plan. A common theme throughout the facilitators was an increase in patient health and the quality of care provided. This directly aligns with the goals and requirements established by the PPACA. Legislature is placing an increasing demand on health care organizations to increase the quality of care delivered to patients at a lower cost. Executing a telemedicine program may help alleviate the burden of the PPACA demands and may benefit the health of the population currently living with hypertension. Access is noted to increase with the use of telemedicine by reaching a broader population that may not have easy access to health care resources and transportation, and individuals who require more care such as the elderly.

These facilitators need to be kept in mind when determining a route to abide by PPACA guidelines. The long-term effects of increases in the health and quality of services that will facilitate financial benefits in the near future should override the initial start-up costs. All decisions are associated with costs and benefits; research indicates that the effectiveness of adopting telemedicine programs is growing, as time unveils the positive outcomes that result from the establishment of these programs.

Limitations

This review provides a collection of up-to-date data associated with using telemedicine and will assist organizations in weighing the costs and benefits associated with its adoption. A shortfall of this study is that all the papers focused on serving different populations with telemedicine. All the papers were related to using telemedicine to measure high blood pressure, but it was the target populations that varied. Some evaluated using telemedicine to reach people in developing countries, the poor and underserved in developed countries, individuals with diabetes, individuals who only speak Spanish, and individuals who are African American.

The extraordinarily vast number of uses for telemedicine made research difficult with regard to narrowing down the target population while still having enough data to conduct a systematic review. Telemedicine is relatively new to assist with self-management of chronic conditions, and there is limited research on the subject. Time is required for facilities to gather data information on the effectiveness of using telemedicine for this purpose, and this study was limited to papers published only within the past 5 years.

Conclusions

Weighing the costs versus the benefits in the creation of a telemedicine program for self-management of hypertension is essential to decide whether it fits the needs of the organization.

This review presented a myriad of facilitating factors and barriers through a meta-analysis and systematic review of up-to-date papers from two academic databases. The information presented is helpful in understanding the benefits of telemedicine and its function in an organization.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. This paper does not contain any studies with human participants or animals performed by any of the authors. For this type of study, formal consent is not required.

Conflicts of Interest

None declared.

References

1. Crowley MJ, Bosworth HB, Coffman CJ, Lindquist JH, Neary AM, Harris AC, et al. Tailored case management for diabetes and hypertension (TEACH-DM) in a community population: study design and baseline sample characteristics. Contemp Clin Trials 2013 Sep;36(1):298-306 [FREE Full text] [doi: 10.1016/j.cct.2013.07.010] [Medline: 23916915]
2. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015;9:CD002098. [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
3. Kaambwa B, Bryan S, Jowett S, Mant J, Bray EP, Hobbs FD, et al. Telemonitoring and self-management in the control of hypertension (TASMINH2): a cost-effectiveness analysis. Eur J Prev Cardiol 2014 Dec;21(12):1517-1530. [doi: 10.1177/2047487313501886] [Medline: 23990660]
4. Fitzner K, Moss G. Telehealth--an effective delivery method for diabetes self-management education? Popul Health Manag 2013 Jun;16(3):169-177. [doi: 10.1089/pop.2012.0054] [Medline: 23216062]
5. 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 Feb;19(2):89-94 [FREE Full text] [doi: 10.1177/1357633X13476234] [Medline: 23532005]
6. Jones MI, Greenfield SM, Bray EP, Baral-Grant S, Hobbs FD, Holder R, et al. Patients' experiences of self-monitoring blood pressure and self-titration of medication: the TASMINH2 trial qualitative study. Br J Gen Pract 2012 Feb;62(595):e135-e142 [FREE Full text] [doi: 10.3399/bjgp12X625201] [Medline: 22520791]
7. WHO. 2010 Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth URL: http://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2016-09-01] [WebCite Cache ID 6kCh7kpqZ]
8. Weber MA, Schiffrin EL, White WB, Mann S, Lindholm LH, Kenerson JG, et al. Clinical practice guidelines for the management of hypertension in the community a statement by the American society of hypertension and the international society of hypertension. J Hypertens 2014 Jan;32(1):3-15. [doi: 10.1097/HJH.0000000000000065] [Medline: 24270181]
9. WHO. 2013 A global brief on hypertension: silent killer, global public health crisis URL: http://apps.who.int/iris/bitstream/10665/79059/1/WHO_DCO_WHD_2013.2_eng.pdf [WebCite Cache ID 6kChSNAZ9]
10. Melnyk SD, Zullig LL, McCant F, Danus S, Oddone E, Bastian L, et al. Telemedicine cardiovascular risk reduction in veterans. Am Heart J 2013 Apr;165(4):501-508 [FREE Full text] [doi: 10.1016/j.ahj.2012.08.005] [Medline: 23537965]
11. Wang V, Smith VA, Bosworth HB, Oddone EZ, Olsen MK, McCant F, et al. Economic evaluation of telephone self-management interventions for blood pressure control. Am Heart J 2012 Jun;163(6):980-986. [doi: 10.1016/j.ahj.2012.03.016] [Medline: 22709750]
12. Maciejewski ML, Bosworth HB, Olsen MK, Smith VA, Edelman D, Powers BJ, et al. Do the benefits of participation in a hypertension self-management trial persist after patients resume usual care? Circ Cardiovasc Qual Outcomes 2014 Mar;7(2):269-275 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.113.000309] [Medline: 24619321]
13. McKoy J, Fitzner K, Margetts M, Heckinger E, Specker J, Roth L, et al. Are telehealth technologies for hypertension care and self-management effective or simply risky and costly? Popul Health Manag 2015 Jun;18(3):192-202. [doi: 10.1089/pop.2014.0073] [Medline: 25290443]
14. Kumar N, Khunger M, Gupta A, Garg N. A content analysis of smartphone-based applications for hypertension management. J Am Soc Hypertens 2015 Feb;9(2):130-136. [doi: 10.1016/j.jash.2014.12.001] [Medline: 25660364]
15. Maciejewski ML, Bosworth HB, Olsen MK, Smith VA, Edelman D, Powers BJ, et al. Do the benefits of participation in a hypertension self-management trial persist after patients resume usual care? Circ Cardiovasc Qual Outcomes 2014 Mar;7(2):269-275 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.113.000309] [Medline: 24619321]

16. Shaw RJ, Kaufman MA, Bosworth HB, Weiner BJ, Zullig LL, Lee SD, et al. Organizational factors associated with readiness to implement and translate a primary care based telemedicine behavioral program to improve blood pressure control: the HTN-IMPROVE study. Implement Sci 2013;8:106 [FREE Full text] [doi: 10.1186/1748-5908-8-106] [Medline: 24010683]
17. Jackson GL, Oddone EZ, Olsen MK, Powers BJ, Grubber JM, McCant F, et al. Racial differences in the effect of a telephone-delivered hypertension disease management program. J Gen Intern Med 2012 Dec;27(12):1682-1689 [FREE Full text] [doi: 10.1007/s11606-012-2138-x] [Medline: 22865016]
18. Piette JD, Datwani H, Gaudioso S, Foster SM, Westphal J, Perry W, et al. Hypertension management using mobile technology and home blood pressure monitoring: results of a randomized trial in two low/middle-income countries. Telemed J E Health 2012 Oct;18(8):613-620 [FREE Full text] [doi: 10.1089/tmj.2011.0271] [Medline: 23061642]

Abbreviations

CINAHL: Cumulative Index of Nursing and Allied Health Literature
EBSCO: Elton B Stephens Company
EKG: electrocardiograph
eHealth: electronic health
HINTS: Hypertension Intervention Nurse Telemedicine Study
ICD-10: International Classification of Diseases, Tenth Edition
mHealth: mobile health
PPACA: Patient Protection and Affordable Care Act of 2010
TEACH-DM: Tailored Case Management for Diabetes and Hypertension

Edited by G Eysenbach; submitted 01.09.16; peer-reviewed by D Cianflone, A Subasinghe; comments to author 25.10.16; revised version received 18.11.16; accepted 01.07.17; published 24.10.17.

Please cite as:
Mileski M, Kruse CS, Catalani J, Haderer T
Adopting Telemedicine for the Self-Management of Hypertension: Systematic Review
JMIR Med Inform 2017;5(4):e41
URL: http://medinform.jmir.org/2017/4/e41/
doi:10.2196/medinform.6603
PMID:29066424

Original Paper

A Data Model for Teleconsultation in Managing High-Risk Pregnancies: Design and Preliminary Evaluation

Kolsoum Deldar[1], MSc, PhD; Fatemeh Tara[2*], MD; Kambiz Bahaadinbeigy[3*], MD, PhD; Mohammad Khajedaluee[4*], MD, PhD; Mahmood Tara[5], MD, PhD

[1]Student Research Committee, Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Islamic Republic Of Iran

[2]Women Health Research Center, Department of Obstetrics and Gynecology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Islamic Republic Of Iran

[3]Medical Informatics Research Center, Institute of Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Islamic Republic Of Iran

[4]Department of Community Medicine, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Islamic Republic Of Iran

[5]Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Islamic Republic Of Iran

[*]these authors contributed equally

Corresponding Author:
Mahmood Tara, MD, PhD
Department of Medical Informatics
Faculty of Medicine
Mashhad University of Medical Sciences
Azadi Square
Mashhad,
Islamic Republic Of Iran
Phone: 98 5138002429
Email: taram@mums.ac.ir

Abstract

Background: Teleconsultation is a guarantor for virtual supervision of clinical professors on clinical decisions made by medical residents in teaching hospitals. Type, format, volume, and quality of exchanged information have a great influence on the quality of remote clinical decisions or tele-decisions. Thus, it is necessary to develop a reliable and standard model for these clinical relationships.

Objective: The goal of this study was to design and evaluate a data model for teleconsultation in the management of high-risk pregnancies.

Methods: This study was implemented in three phases. In the first phase, a systematic review, a qualitative study, and a Delphi approach were done in selected teaching hospitals. Systematic extraction and localization of diagnostic items to develop the tele-decision clinical archetypes were performed as the second phase. Finally, the developed model was evaluated using predefined consultation scenarios.

Results: Our review study has shown that present medical consultations have no specific structure or template for patient information exchange. Furthermore, there are many challenges in the remote medical decision-making process, and some of them are related to the lack of the mentioned structure. The evaluation phase of our research has shown that data quality (P<.001), adequacy (P<.001), organization (P<.001), confidence (P<.001), and convenience (P<.001) had more scores in archetype-based consultation scenarios compared with routine-based ones.

Conclusions: Our archetype-based model could acquire better and higher scores in the data quality, adequacy, organization, confidence, and convenience dimensions than ones with routine scenarios. It is probable that the suggested archetype-based teleconsultation model may improve the quality of physician-physician remote medical consultations.

(JMIR Med Inform 2017;5(4):e52) doi:10.2196/medinform.8393

KEYWORDS

remote consultation; clinical archetype; pregnancy; clinical decision-making

Introduction

Teleconsultation is an important application of telemedicine, and it could be done remotely between two or more health care providers, for diagnostic or therapeutic purposes, by the use of information and communication technology [1]. In fact, by using this technology, the physician's knowledge, experience, and services could be used despite the lack of his or her physical presence [2].

Teleconsultation Models Based on Time

A look at published articles in the field of telemedicine shows that medical consulting can be performed in different ways, especially in four models: real time, near real time, store and forward, and mixed or hybrid. In a real-time model, the physicians in the referral center and refer center are present at the same time for a teleconsultation session. In this model, data and information exchange could be done through methods such as live chat and Internet-based videoconferencing or procedures independent of the Internet, such as a telephone conversation, whereas in the store and forward model, there is no such synchronization between sending patient-related data and receiving the response of the consultant. It is also possible to use Internet-based methods such as email or Web-based forums and non-Internet-based methods such as fax or short message service (SMS) in this model. Some consultations also use methods that are between real-time and asynchronous methods or near real time. In this method, although physicians in both centers are simultaneously in place for giving consultations, there is usually a short delay between sending data and its delivery to the referring center. This method has been used where there is a high volume of data exchange but not enough bandwidth to communicate live on the Web. Examples of this approach include the ability to upload video files from the referral center and receive and view it after a short time in the referring center [3]. Some consultations are also carried out through a combination or mix of the aforementioned models.

Consultations in Educational Hospitals

In our educational hospitals and in the absence of clinical experts (eg, holidays), the senior and junior residents are responsible for patient care through telephone consulting with an on-call physician. Our previous study had shown that there is no formal education about the consultation process for requesting a resident, and it is learnt only by oral education or observing the performance of senior ones [4]. So, data and information content that is exchanged in these consultations has no standard format or template. On the other hand, type, format, quality, and volume of clinical data exchange is different. These differences can be caused by many factors such as the urgency of clinical conditions of the patient, the Internet speed, experience and background of prior consultations with the applicant physician, receiving or not receiving feedback from the consulting physician, the medium, personality traits of the applicant physician, and time of consultation. Usually clinical findings are summarized by resident, and only prominent points are transferred to an on-call physician. This issue can adversely affect the quality of clinical decisions made by an on-call physician. Additionally, it is important to consider that because of the lack of the consultant's physical presence beside the patient, changes in the type, quality, and the volume of clinical data provided by the applicant physician can make a significant impact on the quality of the distant physician's consultation. Physicians' discontent of poor quality consulting practices and the unstructured data being exchanged has been reported in several studies [5-8]. Moreover, the inadequacy of the data and submitting them as free text are among the problems in the teleconsultation process [9].

Data Model for Teleconsultation

It might improve the consultation process to use structured templates in the form of a data model for teleconsultation. This means that data items required for decision making of the physician in different conditions are identified. Hence, at the time of consultation, the applicant physician would have already completed and submitted all those necessary items to the consulting physician. On the other hand, to observe the minimum standards for the exchange of data and information, the above items such as templates, formats, and units of measure should be limited.

An effective step on the path to standardizing the exchange of data and information is the use of clinical archetypes. Archetype and template reviews show that some of their main goals of design and deployment are clinical data structuring, increasing interactivity between information systems, preserving data integrity, and simply enhancing quality [10]. However, it seems that such a model has not been used in teleconsultation. If the physicians on either side would use the special tele-decision-making clinical archetype for the data and information exchanging process, it is possible to create better outcomes for patients. As little research has been done to create a standard structure for exchanging data [11], this study was designed and conducted to develop a data model for teleconsultation based on the decision-making archetype for managing high-risk pregnancies.

Methods

This study was carried out in three phases; the methodology and results of phases 0 and 1 are described in detail in previous studies.

Phase 0

In this phase, a systematic review was done that aimed to make the physicians familiar with the concept of teleconsultation and finding potential problems. Additionally, a qualitative study was performed to determine the current status of telephone consultations among residents and on-call obstetricians or gynecologists, and a study based on Delphi was carried out to identify high-risk pregnancies in predetermined departments.

Phase 1

Extraction and Localization of Items for Tele-Decision-Making Clinical Archetypes

Main references in obstetrics and gynecology (eg, textbooks, clinical guidelines, and electronic databases) were used to extract the items for tele-decision-making clinical archetypes, and then

they were localized in several expert panel sessions. Details are described in a previous study.

Designing the Model

After finalizing the "extract and localize" process, the accepted items were divided into distinct subcategories (eg, "amniotic fluid index" was allocated to the "biometric ultrasonography" subcategory). These subcategories were also assigned into larger and more comprehensive groups (eg, the "biometric ultrasonography" subcategory was allocated to the "para-clinic" category). This process continued to achieve the main categories as needed for clinical teleconsultations. Required attributes for each item, such as format acceptable to respond to the items and accepted units of measurement for numeric domains, were defined based on the opinions of clinicians and informatics experts. The model was designed by using Microsoft Visio drawing software.

Phase 2

The model, which was designed by using pre-prepared consultation scenarios, was evaluated.

Designing of Consultation Scenarios to Evaluate the Model

In this phase, one of the qualified volunteers, who was a senior obstetrics and gynecology resident of a teaching hospital of the Mashhad University of Medical Sciences (MUMS), was asked to design five assumptive telephone consulting scenarios in which the physician consulted with an on-call professor based on the items of the tele-decision-making clinical archetypes. These scenarios were designed using five real, paper patient records that described a common high-risk pregnancy (one high-risk pregnancy record for each), so that on the one hand, a list of identified items for the physicians' tele-decision making and on the other hand, information of patients with hidden identities was placed at his or her disposal to use them to provide assumptive teleconsultation scenarios with professors.

Then, anonymous medical information of the same cases were given to other senior residents who were usually responsible for consulting with on-call professors. They were asked, assuming that their intention was to consult with an on-call physician about the patient by telephone, to provide a typical scenario for every case. A unique code was assigned to each scenario.

Designing Checklist

The design of the checklist used in this stage, took place in several phases (described in another study). In short, at the beginning, by using the results of qualitative research (interviews with experts), as well as a broad overview of the texts available in electronic resources and scientific articles, possible items suitable for the design of the checklist were identified. Then, all extracted items were examined and modified by clinicians and informatics experts in three stages to design a checklist tailored to the needs of research that was focused on the teleconsultations.

Comparing Routinely Designed Scenarios With an Archetype-Based Scenario

Scenario comparison was developed based on the archetype, with scenarios designed routinely, and these scenarios were randomly placed at the discretion of the clinicians. After reading each consultation scenario, experts commented on the content of consultations and registered them in the predesigned checklist. The allocation of archetype-based scenarios or routine-based scenarios to experts took place randomly. Wilcoxon test was used to compare the scores.

Results

Phase 0

The Result of Systematic Review

The most important finding of this study is that there is presently no structured format for data and information exchange in teleconsultations [11].

The Result of the Qualitative Study

The qualitative study has shown that specialists during teleconsultation with residents faced significant challenges at the time of the diagnosis and treatment of patients; the majority of these problems were caused by insufficient confidence in the judgment of the resident or disproportionate volume of information received concerning the patients [4].

The Result of the Delphi Study

Results of this study have revealed that the most common high-risk pregnancies in hospitals in Mashhad are as follows: pre-eclampsia or eclampsia, hemorrhage in the third trimester, postterm delivery, preterm birth, and premature rupture of membranes [12].

Phase 1

The Result of Extraction and Localization of Items for Tele-Decision-Making Clinical Archetypes

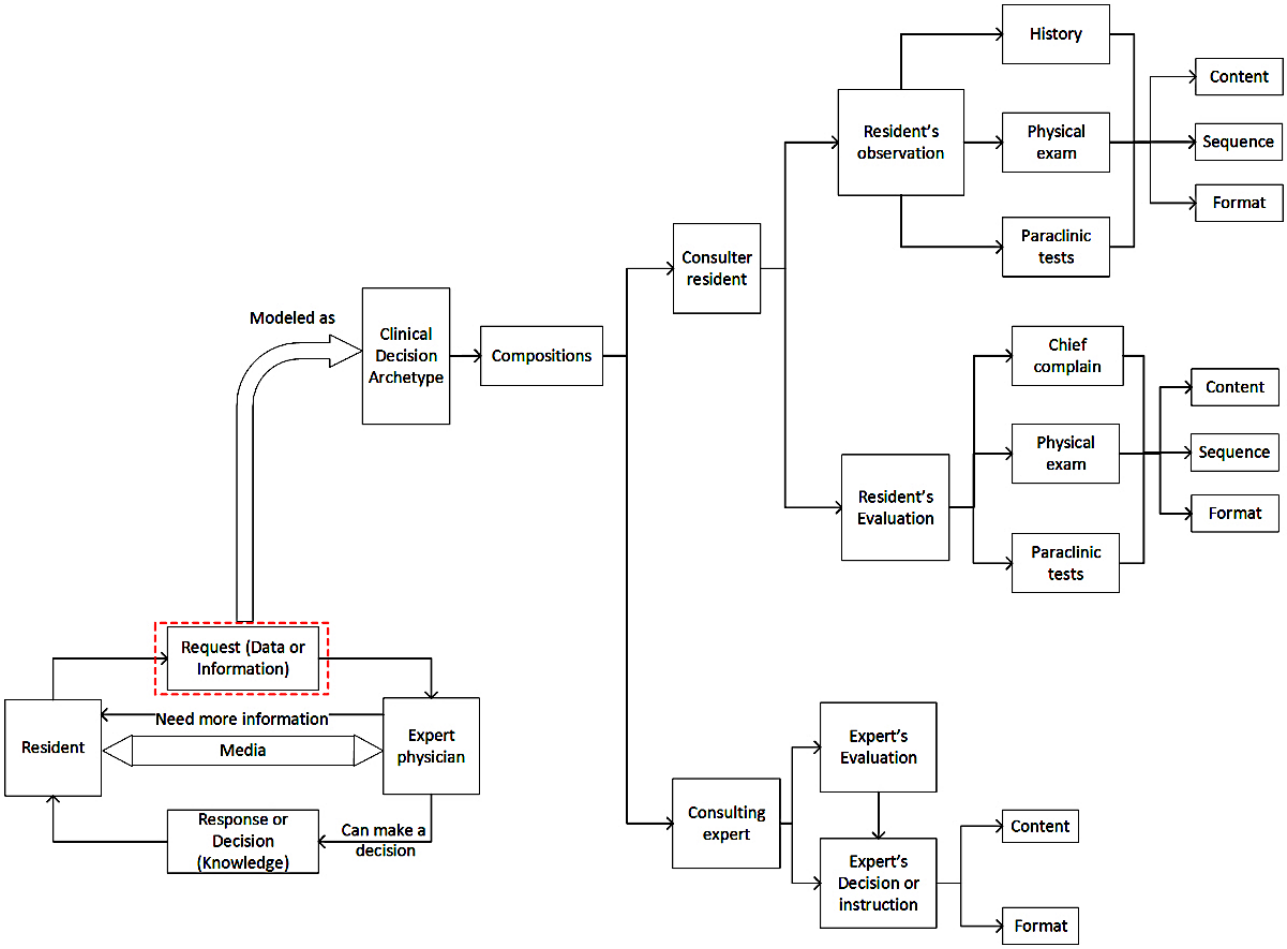
This step led to the formation of two groups: general and demographic information (16 items) and technical information (142 items) [12].

The Designed Model

The result of our previous review study has shown that overall the teleconsultation model is almost identical among physicians and follows a general pattern of requesting consulting and sending patient information by the first physician and responding to it by the consulting physician. In this study, this overall model and pattern was used as the basis. However, in the section related to the exchange of data and information, the clinical archetypes particular to decision making were used (Figure 1).

As seen in Figure 1, the designed clinical archetypes consisted of two compositions associated with the request of the resident and the physician's answer. Patient data and information collected by residents were grouped into two general categories of "observation" and "analysis." Information related to the patient's medical record, physical examination, and laboratory test results was inserted into the residents' observation group.

Figure 1. Designed model for teleconsultation, based on clinical decision-making archetype.



These observations, especially those related to physical examinations and laboratory tests, were analyzed by the residents; in consultations between the senior residents and specialists, usually the resident's analysis of such information was transferred to the specialist. For example, if the pregnant woman's platelet count was 127,000, instead of the aforementioned number, the resident transferred their own analysis and impressions of the number as "normal platelet count" to the specialist. It should be noted that the residents were able to properly analyze only some of these cases, and in the case of other data and information, they had to use the correct analysis based on clinical experience of the on-call physician. The most important issue that should be considered in this archetype is determining a template for type and amount of information to be transferred to the specialist. As was also identified in the qualitative part of this project (and similar studies), experts believe that factors such as the personality of residents, patient presentation, and the amount of oral information received during the consultation have a great impact on decisions made by physicians. Hence, in this model, we were trying to extract the most common, necessary items used to make decisions about high-risk pregnancy by the review of the literature and surveys from local experts. Then, with the help of clinicians and informatics experts, proper format and report priority was set for all of these items' components. In the next step, a structured format to provide a summary report of the completed items by the residents and to send it to the experts was designed. A subsidiary of one of the components of the designed model (the last available biometric ultrasound) is shown in more detail in Figure 2.

Phase 2

The Result of the Checklist Design Step

At this point, a checklist that contained seven items was used for evaluation of the model (Multimedia Appendix 1).

The Result of Comparison of Archetype-Based Scenario With Routine-Based Scenario Group

The following table shows quartiles scores of the self-assessment checklist of specialists, divided into the archetype-based scenario and routine-based scenario groups (Table 1).

Figure 2. One of the subsets of designed model for premature rupture of membrane (PROM; the last biometric ultrasound). AFI: amniotic fluid index; EFW: estimated fetal weight; AC: abdominal circumference; FL: femur length; BPD: biparietal diameter; HC: head circumference.

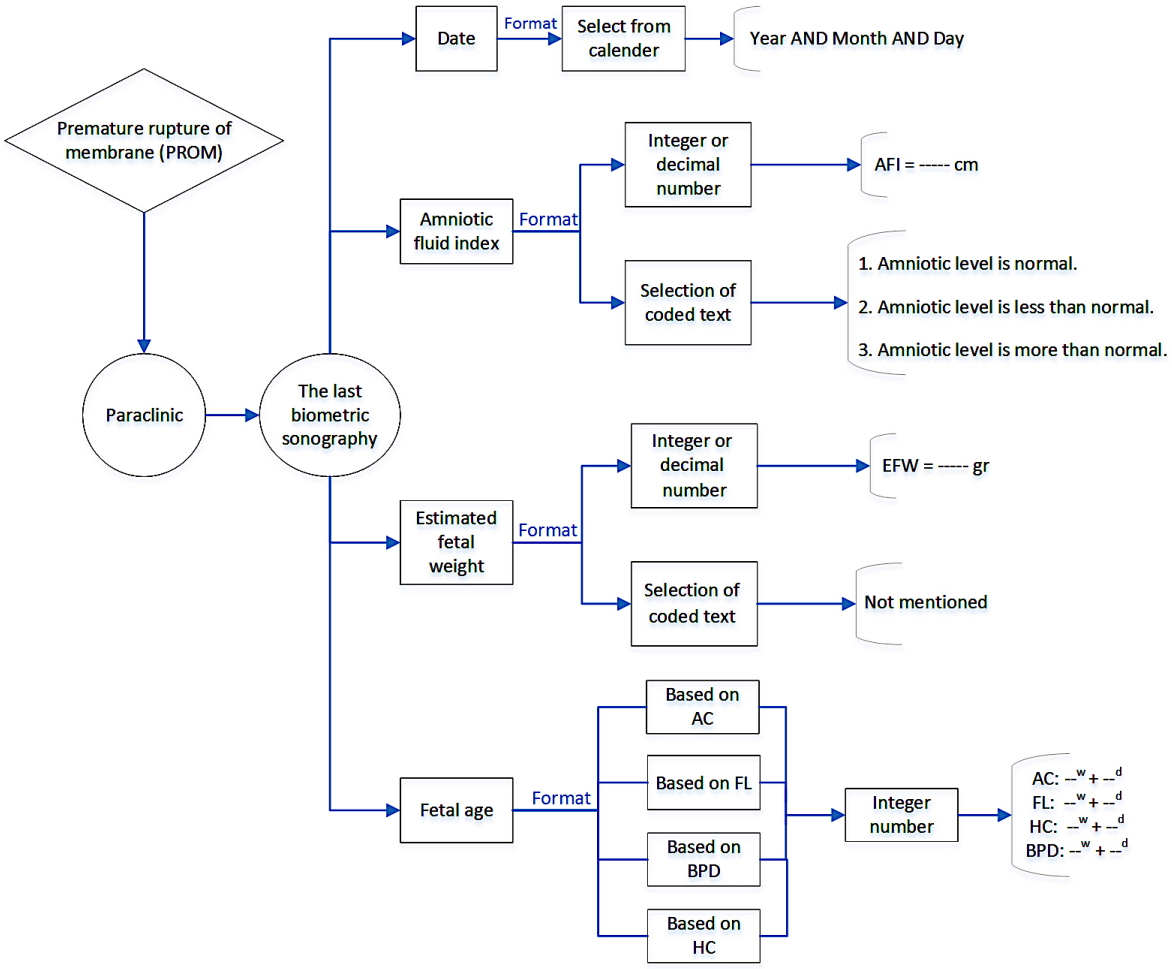


Table 1. Quartiles scores of self-assessment checklist of specialists in both groups.

Item number and description		Group	25th	Median	75th	*P* value
1	The quality of patient-related data and information provided on this consultation is acceptable.	AS[a]	4	4	5	<.001
		RS[b]	3	4	4	
2	The volume of information presented in this consultation, to decide about the patient, is enough.	AS	4	5	5	<.001
		RS	3	4	5	
3	There is additional nonuseful information in those provided about the patient.	AS	4	4	5	<.001
		RS	3	4	4	
4	Order and organization of the information in this consultation were acceptable.	AS	4	4	5	<.001
		RS	3	4	5	
5	According to the provided information, I'm sure about the decision for the patient.	AS	4	5	5	<.001
		RS	3	4	5	
6	The time required to make decisions with respect to the quality of information provided was acceptable.	AS	4	4	5	.122
		RS	3	4	5	
7	According to information provided, making the decision for the patient was easy for me.	AS	4	5	5	<.001
		RS	2.25	4	5	

[a]AS: archetype-based scenario.

[b]RS: routine-based scenario.

According to the results of the Wilcoxon test, in six items of the specialists' self-assessment checklist, there was a significant difference between the scores assigned to the archetype-based scenario and those of the routine-based scenario groups, and in all of them (except the third item), the archetype-based scenario group received better scores than the other group. In the third item related to the presence of additional information in the scenarios, scores of the archetype-based scenario group were higher but worse than the routine-based scenario group. The only item that did not have a significant difference was the one related to the time required for clinical decision making of specialists.

Discussion

Principal Findings

In this study, to evaluate the proposed data model for teleconsultation archetype-based clinical decision-making consultation scenarios, we compared them with those scenarios based on routine procedures prepared by senior residents using the self-assessment checklist of obstetrics and gynecology physicians to rank the quality of their clinical decisions. The results of this comparison indicated that in five items (out of seven items in the checklist), scenarios based on archetypes received more favorable scores. This means that according to specialists, scenarios based on archetypes were better than scenarios based on routine procedures in terms of "quality of data," "adequacy of information," "discipline and organization," "certainty of decision," and "ease of decision making." Although the average score of "acceptable time required for decision making" in the archetype-based scenario group was also higher than the routine-based scenario group, there was no statistically significant difference here. In the comparison between the two scenarios, the only item that experts evaluated unfavorable about scenarios based on archetypes was additional information in this scenario.

Results of previous studies have shown that increasing the quality of data and information exchange can improve the quality of decision making [13,14], although this increase depends on the quality of the individual who makes the decision. In other words, increasing the quality of data could increase the quality of decision making in the event that the decision-maker has knowledge about the relationship between the variables of the problem [14]. In this study, items related to acceptability of the quality of data and information in scenarios based on archetype received the higher score. The quality of sanitary data in some studies has been defined by two factors: accuracy and completeness [15]. It seems that evaluation of the accuracy of the data by specialists for any of the scenarios was not possible. Perhaps one of the reasons that specialists gave a higher score to this item in the archetype-based scenario compared with the routine-based scenario consultations was the completeness of this group's scenarios. Due to the scenarios prepared by routine methods being mostly short, and on the assumption that it is not necessary to provide the specialist with some of this information during telephone consultation, residents removed them from their consultations, whereas in the interviews with specialists, they narrated experiences when inadequate information submitted by residents had jeopardized the life of the mother and the baby. Normally, if specialists need more information about the patient, they receive this information through frequent questions and answers with residents. Due to the absence of this approach at the time of writing of the routine scenarios by the resident, such scenarios might be imperfect. It is probable that the use of archetypes needed for clinical decision making improves different data quality and information factors, especially when there is no possibility of frequent questions and answers.

Another self-assessment checklist item was the adequacy of the information for clinical decision making. This item is also one of the aspects needed to enhance the quality of the decision [16,17]. As most managers tend to get as much information as possible, providing the right amount of information to them is a challenging task [16]. Hence, it is better to have a mechanism for determining the volume of incoming information to specialists to prevent information overload and ensure the adequacy of the information required, after all the delivery of clinical information more than the amount required typically does not improve consultation outcomes [18]. By definition, clinical archetypes can be used as a model to determine the structure and content needed to obtain clinical information [19]. They are also an appropriate and good way to describe structured sanitary information [20]. Thus, it is possible to build up the content needed for medical consultation by using a certain type of archetype for clinical tele–decision making. This could be approved according to the higher scores given by clinicians to the adequacy of the information item in the clinical archetype-based scenario in comparison with the routine-based scenario.

Bergus (2006) showed a significant correlation between the way the questions of physicians requesting consultation was organized and that of responses of consulting physicians. In fact, physicians requesting consultations can affect the outcome of consultation with specialists by how they design the questions and structure them; and this impact is independent of the specialists' personal characteristics, the level of training, and the amount of information that is offered by the requesting consultations [18]. In a teledermatology study, a semistructured form was used to send information to physicians. However, the intended structure was only to determine the topic of the data entry fields, and text input was done freely. The most important declared finding of this study was the reduction in the number of patients referred to physicians, but it is not clear whether the use of this basic structure of the information sending form affected this reduction or not [21]. In another study, a pre-consultation structured questionnaire made students pay more attention to details of patient information [22]. Taken together, these studies demonstrate the positive results of using a specified structure and format in consultation. That is probably why the score associated with the acceptability of order and information structure item in archetype-based scenario (due to the specific structural for the display of data and their order of displaying) was higher than that of routine-based scenario.

In some studies, lack of request for conventional (in person) consultation is considered as an indicator to measure the physician's confidence in teleconsultation, and in some other

studies by using the 5-item Likert scale, the physician's confidence in clinical tele–decision making was measured. In the mentioned study, the confidence in decision making has been defined as a component consisting of data quality (such as quality of digital images) and the accuracy and details of the exchanged information. As a result, when digital image quality or resolution of radiological images was poor or amount of sent data was insufficient, the degree of confidence in the decision was reduced. The author of this study believes that sending more details, responding to the questions of consulting physicians, and increasing the number and quality of images could improve the confidence of surgeons to diagnose and treat patients [23]. It has been shown in other studies that the use of more advanced techniques of telemedicine has increased the confidence in the adequacy of the treatment and care provided by physicians and nurses [24]. In this study, specialists could ask more questions about the patients for consulting scenarios if more information is required. Information provided was inadequate, and thus, they felt a sense of uncertainty in the decision. However, only a small number of specialists have raised such questions at the end of scenarios (which could be because of the lack of questions for that scenario or because of the impatience of respondents), although the number of questions in routine-based scenario was more than those in the archetype-based scenario. Higher decision confidence scores may be explained by the above reason.

People usually tend to assess the quality of expectations and their feelings about the decisions that they have taken. One of the most basic and most important experiences after decision making is how comfortable the person feels about the decision taken. A comfortable decision is a decision that accompanies the sense of physical and mental ease and pleasure [25]. According to some studies, specialists believe that if clinical consultations and their related questions have a good structure, they could respond to them more easily than to unstructured questions [18]. In other words, it may be concluded that the structure and certain order in the archetype-based scenario can help the specialist to have a greater sense of comfort and ease when they are using such scenarios in comparison with the time they respond to the routine-based scenario.

The average scores of the item associated with the time needed to decide were the only average value that showed no significant difference between the two groups. In different studies, factors such as the low number of choices, fewer information inputs, and the limitation of the analysis, the small number of decision-makers and the low number of conflicting opinions are mentioned as factors that could increase the speed of decision making. Some other researchers believe that the greater the volume of information, the slower the decision making [26]. However, in our study, it seems that one of the factors that may result in higher average scores of “acceptability of the time required for decision making” item in the archetype-based scenario is sufficient volume of information needed for decision making, and of course, the order in presenting information to specialists.

In another study, in which the result were controversial, it was claimed that the accuracy of a diagnosis was directly associated with the processing speed of the information required to make decisions. In that study, it was concluded that increasing the speed of decision making (and reducing the time required for taking decisions) would increase the accuracy of diagnosis in physicians. In other words, spending a long time to diagnosis did not reduce the error rate in deciding [27]. Therefore, it is probably to expect if, with the improvement of various factors, the speed of physicians in decision making is accelerated, more accurate and less wrong decisions would be taken. Of course, the personality of the decision-maker should not be ignored. In interviews with specialists, it was also noted that the decision-making procedure and speed is different and this is because of the skills, previous experience in dealing with a variety of difficult clinical conditions, as well as the specific individual characteristics of physicians. Perhaps for this reason, there were no significant differences between the average scores of the time of decision making of the two groups. In other words, only modifications of external factors such as how to provide a consulting scenario cannot improve the speed of decision making because the internal and other important factors influence the process.

The only item where the archetype-based scenario received worse scores than the routine-based scenario was “the existence of additional unhelpful information” in these scenarios. As already mentioned, on the one hand, most decision-makers tend to gather as much information as possible to make better decisions [16], and on the other hand, it is better to improve the quality of decision making by a reduction in the volume of data inputs. Given that in the previous items medical specialists gave a higher score to the sufficient volume of archetype-based scenario in contrast to routine-based scenario, the existence of additional unhelpful information might be caused by items that normally are not mentioned because of their negative answer in routine consultation. For example, negative history for liver or kidney disease or absence of family history of hypertension are such information that in the specialists’ view, including written and electronics sources, are important to make decisions about diagnosis and treatment of pregnancy hypertension. Nonetheless, it seems these items along with their negative response, despite their importance in diagnosis and treatment, are interpreted as additional information. Most of our participants (specialists) believed that the lack of these items in the context of consultation meant that a particular item is negative, whereas in the world of information, the lack of these items, in addition to being a negative response, could be a sign that shows that the item is missed or left incomplete. So, ways to avoid misinterpretation in this regard should be looked for, and perhaps using techniques such as aggregation or summarization to reduce the amount of information available in the scenarios [16].

Conclusions

Improving the quality of clinical care is partly associated with the improvement of decisions and judgments of the medical staff [28]. Our study has demonstrated that archetype-based consultation scenarios for clinical decision making were superior to routine-based scenarios in terms of quality, volume, and structure, and specialists felt more confident and comfortable using the archetype-based scenario for decision making. In addition, in terms of time, speed of decision making in these

scenarios is estimated somewhat more acceptable than other scenarios. Nevertheless, there was futile and additional information in these scenarios, and by different techniques of aggregation and summarization, the substance of these scenarios can be improved. Along these lines, it appears that the proposed data model for archetype-based teleconsultation can enhance the quality and the nature of teleconsultation between physicians.

Acknowledgments

This study is a product of a PhD thesis of KD. SMT is guarantor. The authors are grateful to the research deputy of MUMS for financial support of this research. They would like to thank all the supervising physicians, residents, midwives, and nurses who participated in the study and gave their valuable time.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Self-assessment checklist of specialists for teleconsultations.

[PDF File (Adobe PDF File), 18KB - medinform_v5i4e52_app1.pdf]

References

1. Bove AA, Homko CJ, Santamore WP, Kashem M, Kerper M, Elliott DJ. Managing hypertension in urban underserved subjects using telemedicine--a clinical trial. Am Heart J 2013 Apr;165(4):615-621. [doi: 10.1016/j.ahj.2013.01.004] [Medline: 23537980]
2. Torani S, Khammarnia M, Delgoshaei B. The ability of specialized hospitals of Iran university of medical sciences in establishing remote medical advice. Health Inf Manag 2012;8(6):785-794 [FREE Full text]
3. Stahl JN, Zhang J, Zellner C, Pomerantsev EV, Chou TN, Huang HK. A new approach to teleconferencing with intravascular US and cardiac angiography in a low-bandwidth environment. Radiographics 2000;20(5):1495-1503. [doi: 10.1148/radiographics.20.5.g00se151495] [Medline: 10992038]
4. Deldar K, Tara F, Mirteimouri M, Tara SM. Experiences of obstetricians and gynecologists in teleconsultation with medical residents: a qualitative study. Glob J Health Sci 2016 Nov 29;9(6):1. [doi: 10.5539/gjhs.v9n6p1]
5. Bodenheimer T. Coordinating care--a perilous journey through the health care system. N Engl J Med 2008 Mar 06;358(10):1064-1071. [doi: 10.1056/NEJMhpr0706165] [Medline: 18322289]
6. Horner K, Wagner E, Tufano J. Electronic consultations between primary and specialty care clinicians: early insights. Issue Brief (Commonw Fund) 2011 Oct;23:1-14. [Medline: 22059281]
7. Mehrotra A, Forrest CB, Lin CY. Dropping the baton: specialty referrals in the United States. Milbank Q 2011 Mar;89(1):39-68 [FREE Full text] [doi: 10.1111/j.1468-0009.2011.00619.x] [Medline: 21418312]
8. O'Malley AS, Tynan A, Cohen GR, Kemper N, Davis MM. Coordination of care by primary care practices: strategies, lessons and implications. Res Brief 2009 Apr(12):1-16. [Medline: 19452679]
9. Patterson V, Wootton R. A web-based telemedicine system for low-resource settings 13 years on: insights from referrers and specialists. Glob Health Action 2013 Sep 23;6:21465 [FREE Full text] [Medline: 24063752]
10. Hoy D, Hardiker NR, McNicoll IT, Westwell P, Bryans A. Collaborative development of clinical templates as a national resource. Int J Med Inform 2009 Apr;78(Suppl 1):S95-100. [doi: 10.1016/j.ijmedinf.2008.06.003] [Medline: 18639486]
11. Deldar K, Bahaadinbeigy K, Tara SM. Teleconsultation and clinical decision making: a systematic review. Acta Inform Med 2016 Jul 16;24(4):286-292 [FREE Full text] [doi: 10.5455/aim.2016.24.286-292] [Medline: 27708494]
12. Deldar K, Tara F, Mirzaeian S, Pourhoseini S, Tara SM. Systematic extraction of diagnostic data items for common high-risk pregnancies using Delphi technique. IJOG 2017;20(3):18-28 [FREE Full text] [doi: 10.22038/IJOGI.2017.8869]
13. Frank AU. Analysis of dependence of decision quality on data quality. J Geograph Syst 2008 Feb 29;10(1):71-88. [doi: 10.1007/s10109-008-0059-3]
14. Raghunathan S. Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis. Decis Support Syst 1999 Oct;26(4):275-286. [doi: 10.1016/S0167-9236(99)00060-3]
15. Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. AMIA Annu Symp Proc 2006:324-328 [FREE Full text] [Medline: 17238356]
16. Hwang MI, Lin JW. Information dimension, information overload and decision quality. J Inf Sci 2016 Jul;25(3):213-218. [doi: 10.1177/016555159902500305]
17. Resnicow K, Abrahamse P, Tocco RS, Hawley S, Griggs J, Janz N, et al. Development and psychometric properties of a brief measure of subjective decision quality for breast cancer treatment. BMC Med Inform Decis Mak 2014 Dec 05;14:110 [FREE Full text] [doi: 10.1186/s12911-014-0110-x] [Medline: 25476986]

18. Bergus GR, Emerson M, Reed DA, Attaluri A. Email teleconsultations: well formulated clinical referrals reduce the need for clinic consultation. J Telemed Telecare 2006;12(1):33-38. [doi: 10.1258/135763306775321434] [Medline: 16438777]
19. Leslie H, Heard S. Archetypes 101. Brunswick East, Victoria: Health Informatics Society of Australia; 2006 Presented at: HIC 2006 and HINZ 2006: Proceedings; 2006; Brunswick East, Vic p. 18-23.
20. HISO. Health.gov. 2012. Health information exchange content model architecture building block, HISO 10040.2 URL: https://www.health.govt.nz/system/files/documents/publications/hiso-10040-2-hie-content-model-v1.pdf [accessed 2017-12-02] [WebCite Cache ID 6vPHVA9W7]
21. Eminović N, de Keizer NF, Wyatt JC, ter RG, Peek N, van Weert HC, et al. Teledermatologic consultation and reduction in referrals to dermatologists: a cluster randomized controlled trial. Arch Dermatol 2009 May;145(5):558-564. [doi: 10.1001/archdermatol.2009.44] [Medline: 19451500]
22. Shaikh N, Lehmann CU, Kaleida PH, Cohen BA. Efficacy and feasibility of teledermatology for paediatric medical education. J Telemed Telecare 2008;14(4):204-207. [doi: 10.1258/jtt.2008.071108] [Medline: 18534955]
23. Stutchfield BM, Jagilly R, Tulloh BR. Second opinions in remote surgical practice using email and digital photography. ANZ J Surg 2007 Nov;77(11):1009-1012. [doi: 10.1111/j.1445-2197.2007.04299.x] [Medline: 17931268]
24. Donnem T, Ervik B, Magnussen K, Andersen S, Pastow D, Andreassen S, et al. Bridging the distance: a prospective tele-oncology study in Northern Norway. Support Care Cancer 2012 Sep;20(9):2097-2103. [doi: 10.1007/s00520-011-1319-1] [Medline: 22076621]
25. Parker JR, Lehmann DR, Xie Y. Decision comfort. J Consum Res 2016 Feb 01;43(1):113-133. [doi: 10.1093/jcr/ucw010]
26. Eisenhardt KM. Making fast strategic decisions in high-velocity environments. Acad Manag J 1989 Sep 01;32(3):543-576. [doi: 10.2307/256434]
27. Sherbino J, Dore KL, Wood TJ, Young ME, Gaissmaier W, Kreuger S, et al. The relationship between response time and diagnostic accuracy. Acad Med 2012 Jun;87(6):785-791. [doi: 10.1097/ACM.0b013e318253acbd] [Medline: 22534592]
28. Thompson C, Stapley S. Do educational interventions improve nurses' clinical decision making and judgement? A systematic review. Int J Nurs Stud 2011 Jul;48(7):881-893. [doi: 10.1016/j.ijnurstu.2010.12.005] [Medline: 21241984]

Abbreviations

AS: archetype-based scenario
MUMS: Mashhad University of Medical Sciences
RS: routine-based scenario
SMS: short message service

Edited by G Eysenbach; submitted 11.07.17; peer-reviewed by L Grieten; comments to author 04.08.17; revised version received 15.10.17; accepted 29.10.17; published 14.12.17.

Please cite as:
Deldar K, Tara F, Bahaadinbeigy K, Khajedaluee M, Tara M
A Data Model for Teleconsultation in Managing High-Risk Pregnancies: Design and Preliminary Evaluation
JMIR Med Inform 2017;5(4):e52
URL: http://medinform.jmir.org/2017/4/e52/
doi:10.2196/medinform.8393
PMID:29242181

Original Paper

Expert Search Strategies: The Information Retrieval Practices of Healthcare Information Professionals

Tony Russell-Rose[1*], BEng, MSc, PhD; Jon Chamberlain[1,2*], BSc, PhD

[1]UXLabs Ltd, Guildford, United Kingdom

[2]School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Tony Russell-Rose, BEng, MSc, PhD
UXLabs Ltd
3000 Cathedral Hill
Guildford, GU2 7YB
United Kingdom
Phone: 44 7779 936191
Email: tgr@uxlabs.co.uk

Abstract

Background: Healthcare information professionals play a key role in closing the knowledge gap between medical research and clinical practice. Their work involves meticulous searching of literature databases using complex search strategies that can consist of hundreds of keywords, operators, and ontology terms. This process is prone to error and can lead to inefficiency and bias if performed incorrectly.

Objective: The aim of this study was to investigate the search behavior of healthcare information professionals, uncovering their needs, goals, and requirements for information retrieval systems.

Methods: A survey was distributed to healthcare information professionals via professional association email discussion lists. It investigated the search tasks they undertake, their techniques for search strategy formulation, their approaches to evaluating search results, and their preferred functionality for searching library-style databases. The popular literature search system PubMed was then evaluated to determine the extent to which their needs were met.

Results: The 107 respondents indicated that their information retrieval process relied on the use of complex, repeatable, and transparent search strategies. On average it took 60 minutes to formulate a search strategy, with a search task taking 4 hours and consisting of 15 strategy lines. Respondents reviewed a median of 175 results per search task, far more than they would ideally like (100). The most desired features of a search system were merging search queries and combining search results.

Conclusions: Healthcare information professionals routinely address some of the most challenging information retrieval problems of any profession. However, their needs are not fully supported by current literature search systems and there is demand for improved functionality, in particular regarding the development and management of search strategies.

(JMIR Med Inform 2017;5(4):e33) doi:10.2196/medinform.7680

KEYWORDS

review; surveys and questionnaires; search engine; information management; information systems

Introduction

Background

Medical knowledge is growing so rapidly that it is difficult for healthcare professionals to keep up. As the volume of published studies increases each year [1], the gap between research knowledge and professional practice grows [2]. Frontline healthcare providers (such as general practitioners [GPs]) responding to the immediate needs of patients may employ a Web-style search for diagnostic purposes, with Google being reported to be a useful diagnostic tool [3]; however, the credibility of results depends on the domain [4]. Medical staff may also perform more in-depth searches, such as rapid evidence reviews, where a concise summary of what is known about a topic or intervention is required [5].

Healthcare information professionals play the primary role in closing the gap between published research and medical practice, by synthesizing the complex, incomplete, and at times

conflicting findings of biomedical research into a form that can readily inform healthcare decision making [6]. The systematic literature review process relies on the painstaking and meticulous searching of multiple databases using complex Boolean search strategies that often consist of hundreds of keywords, operators, and ontology terms [7] (Textbox 1).

Textbox 1. An example of a multi-line search strategy.

1. Attention Deficit Disorder with Hyperactivity/
2. adhd
3. addh
4. adhs
5. hyperactiv$
6. hyperkin$
7. attention deficit$
8. brain dysfunction
9. or/1-8
10. Child/
11. Adolescent/
12. child$ or boy$ or girl$ or schoolchild$ or adolescen$ or teen$ or “young person$” or “young people$” or youth$
13. or/10-12
14. acupuncture therapy/or acupuncture, ear/or electroacupuncture/
15. accupunct$
16. or/14-15
17. 9 and 13 and 16

Performing a systematic review is a resource-intensive and time consuming undertaking, sometimes taking years to complete [8]. It involves a lengthy content production process whose output relies heavily on the quality of the initial search strategy, particularly in ensuring that the scope is sufficiently exhaustive and that the review is not biased by easily accessible studies [9].

Numerous studies have been performed to investigate the healthcare information retrieval process and to better understand the challenges involved in strategy development, as it has been noted that online health resources are not created by healthcare professionals [10]. For example, Grant [11] used a combination of a semi-structured questionnaire and interviews to study researchers’ experiences of searching the literature, with particular reference to the use of optimal search strategies. Sampson et al [12] used a combination of a Web-based survey and peer review forums to investigate what elements of the search process have the most impact on the overall quality of the resulting evidence base. Similarly, Gillies et al [13] used an online survey to investigate the review, with a view to identifying problems and barriers for authors of Cochrane reviews. Ciapponi and Glujovsky [14] also used an online survey to study the early stages of systematic review.

No single database can cover all the medical literature required for a systematic review, although some are considered to be a core element of any healthcare search strategy, such as MEDLINE [15], Embase [16], and the Cochrane Library [17]. Consequently, healthcare information professionals may consult these sources along with a number of other, more specialized databases to fit the precise scope area [18].

A survey [1] of online tools for searching literature databases using PubMed [19], the online literature search service primarily for MEDLINE, showed that most tools were developed for managing search results (such as ranking, clustering into topics and enriching with semantics). Very few tools improved on the standard PubMed search interface or offered advanced Boolean string editing methods in order to support complex literature searching.

Objective

To improve the accuracy and efficiency of the literature search process, it is essential that information retrieval applications (in this case, databases of medical literature and the interfaces through which they are accessed) are designed to support the tasks, needs, and expectations of their users. To do so they should consider the layers of context that influence the search task [20] and how this affects the various phases in the search process [21]. This study was designed to fill gaps in this knowledge by investigating the information retrieval practices of healthcare information professionals and contrasting their requirements to the level of support offered by a widely used literature search tool (PubMed).

The specific research questions addressed by this study were (1) How long do search tasks take when performed by healthcare information professionals? (2) How do they formulate search strategies and what kind of search functionality do they use? (3) How are search results evaluated? (4) What functionality

do they value in a literature search system? (5) To what extent are their requirements and aspirations met by the PubMed literature search system?

In answering these research questions we hope to provide direct comparisons within other professions (eg, in terms of the structure, complexity, and duration of their search tasks).

Methods

Online Survey

The survey instrument consisted of an online questionnaire of 58 questions divided into 5 sections. It was designed to align with the structure and content of Joho et al's [22] survey of patent searchers and wherever possible also with Geschwandtner et al's [23] survey of medical professionals to facilitate comparisons with other professions. The following were the 5 sections: (1) Demographics, the background and professional experience of the respondents; (2) Search tasks, the tasks that respondents perform when searching literature databases; (3) Query formulation, the techniques respondents used to formulate search strategies; (4) Evaluating search results, how respondents evaluate the results of their search tasks; and (5) Ideal functionality for searching databases, any other features that respondents value when searching literature databases.

The survey was designed to be completed in approximately 15 minutes and was pre-tested for face validity by 2 health sciences librarians.

Survey respondents were recruited by sending an email invitation with a link to the survey to 5 healthcare professional association mailing lists that deal with systematic reviews and medical librarianship: Lis-Medical [24], clinical librarians [25], evidence-based health [26], expert searching [27], and the Cochrane Information Retrieval Methods Group (IRMG) [28]. It was also sent directly to the members of the Chartered Institute of Library and Information Professionals (CILIP) Healthcare Libraries special interest group [29]. The recruitment message and start page of the survey described the eligibility criteria for survey participants, expected time to complete the survey, its purpose, and funding source.

The survey (Multimedia Appendix 1) was conducted using SurveyMonkey, a Web-based software application [30]. Data were collected from July to September 2015. A total of 218 responses were received, of which 107 (49.1%, 107/218) were complete (meaning all pages of the survey had been viewed and all compulsory questions responded to). Only complete surveys were examined. Since the number of unique individuals reached by the mailing list announcements is unknown, the participation rate cannot be determined.

Responses to numeric questions were not constrained to integers as a pilot survey had shown that respondents preferred to put in approximate and/or expressive values. Text responses corresponding to numerical questions (questions 14 to 22 and 32 to 38; 16 in total) were normalized as follows: (1) when the respondent specified a range (eg, 10 to 20 hours), the midpoint was entered (eg, 15 hours); (2) when the respondent indicated a minimum (eg, 10 years and greater), the minimum was entered (eg, 10 years); and (3) when the respondent entered an approximate number (eg, about 20), that number was entered (eg, 20).

After normalizing, 8.29% (142/1712) responses contained no numerical data and 21.61% (370/1712) responses were normalized.

Evaluation of PubMed

An evaluation of the PubMed search system was performed using online documentation [31], best practice advice [32], and direct testing of the interface using Boolean commands. In addition to the search portal, users can register to My NCBI which provides additional functionality for saving search queries, managing results sets, and customizing filters so this was included in the comparison. The mobile version of PubMed, PubMed Mobile [33], does not offer extended functionality so it was not considered in the evaluation. Although beyond the scope of this study, information seeking by healthcare practitioners on hand-held devices has been shown to save time and improve the early learning of new developments [34].

Results

Demographics

Of the respondents, 89.3% (92/103) were female. Their ages were distributed bi-modally, with peaks at 39 to 45 and 53 to 59, with a conflated average age of 46.0 (SD 10.9, N=104) (Figure 1).

The mean time for respondents' experience in their profession was 16.6 years (SD 10.0), greater than their 12.0 (SD 9.0) years of experience in the review of scientific literature (N=107, $P<.01$, paired t test). Most respondents worked full time (78.5%, 84/107) and the commissioning agents for their searches were predominantly internal (ie, within the same organization [72.9%, 78/107]).

The majority of respondents were either based in the UK (51.4%, 55/107), in the US (27.1%, 29/107), or in Canada (7.5%, 8/107). The remaining respondents were from Australia (2.8%, 3/107), Netherlands, Norway, and Germany (1.9% each, 2/107), and Denmark, Singapore, Uruguay, South Africa, Belgium, and Ireland (0.9% each, 1/107). All (100.0%, 107/107) respondents stated that the language they used most frequently for searching was English; however, 6.5% (7/107) stated that they did not use English most frequently for communication in their workplace.

The majority of respondents (81.3%, 87/107) worked in organizations that provide systematic reviews. These organizations also provided other services including reference management (72.0%, 77/107), rapid evidence reviews (63.6%, 68/107), background reviews (60.7%, 65/107), and critical appraisals (52.3%, 56/107).

Figure 1. Age of respondents.

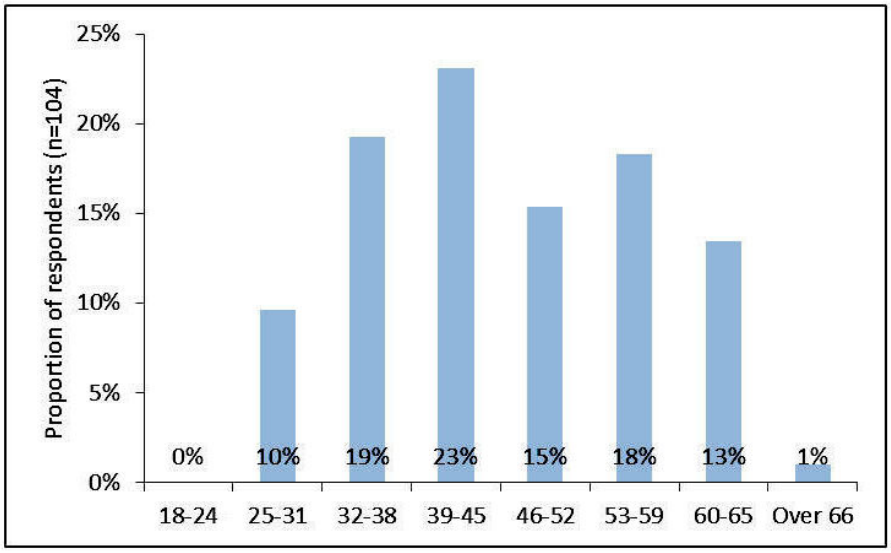


Table 1. Effort to complete search tasks and evaluate results.

Task	Minimum (IQR[a])	Average (IQR)	Maximum (IQR)
Search time per document collection/database, minutes	20 (10-30)	60 (27.5-150)	228 (86-480)
Search task completion time, hours	1 (0.5-2)	4 (2-6.5)	14 (7-30)
Strategy lines per search task, n	5 (2.8-10)	15 (9.1-30)	59 (30-105)
Results examined from a search task, n	10 (5-32)	175 (75-500)	850 (400-5250)
Time to assess relevance of a single result/document, minutes	1 (0.5-2)	3 (1-5)	10 (5-25)
Ideal number of search results per search task, n	0	100	10,000

[a]IQR: interquartile range.

Search Tasks

We considered a search task in this context to be the creation of one or more strategy lines to search a specific collection of documents or database, with task completion resulting in a set of search results that will be subject to further analysis. The output of this process is the search strategy, which is often published as part of the search documentation. This rationalization is in line with a healthcare information professionals' understanding but the complexity of search tasks in this domain is discussed in more detail later.

The time spent formulating search strategies, the amount of time respondents spend completing search tasks, and the number of strategy lines they use is shown in Table 1. Respondents were asked to estimate a minimum, average, and maximum for each of these measures, and the values reported here are the medians of each with the interquartile range (IQR) shown in brackets (in the form Q1 to Q3). The final row shows the minimum, average, and maximum answers to the question: "What would you consider to be the ideal number of results returned for a typical search task?" On average, it takes 60 minutes to formulate a search strategy for a document collection, with the search task taking 4 hours to complete, and the final strategy consisting of 15 lines.

The data sources most frequently searched were MEDLINE (96.3%, 103/107), the Cochrane Library (87.9%, 94/107), and Embase (80.4%, 86/107) (Figure 2).

The majority of respondents (86.9%, 93/107) used previous search strategies or templates at least sometimes, suggesting that the value embodied in them is recognized and should be re-used wherever possible. In addition, most respondents (89.7%, 96/107) routinely share their search strategies in some form, either with colleagues in their workgroup, more broadly within their organization, or in some other capacity (eg, with clients or as part of a published review).

Query Formulation

We examined the mechanics of the query formulation process by asking respondents to indicate a level of agreement to statements using a 5-point Likert scale ranging from 1 (strong disagreement) to 5 (strong agreement). The results are shown in Figure 3.

When asked which taxonomies are regularly used, 74.8% (80/107) of respondents indicated they used MeSH, 45.8% (49/107) Emtree, and 18.9% (20/107) CINAHL headings.

When asked which combination of techniques they used to create their search strategies, 44.9% (48/107) stated they used a form-based query builder, 41.1% (44/107) did so manually

on paper, and 40.2% (43/107) used a text editor. Only 9.3% (10/107) used some form of visual query builder.

Evaluating Search Results

Respondents indicated that the ideal number of results returned for a search task would be 100 documents, yet in practice they evaluate more than this (a median of 175 documents; Table 1). The ideal number of results and the actual number of results evaluated are strongly correlated (N=66, ρ=.661 [Spearman rank correlation]). The average time to assess relevance of a single document was 3 minutes.

Respondents were asked to indicate on a 5-point Likert scale how frequently they use search limits and restriction criteria to narrow down results. The results are shown in Figure 4.

We also examined respondents' strategies for examining the search results. The most popular approaches were to "start with the result that looked most relevant" (54.2%, 58/107) or simply "select the first result" (23.4%, 25/107). No respondent suggested selecting the "most trustworthy source."

Respondents were asked what types of activities [35] they typically engaged in whilst completing their search task (Figure 5). "Locating, verifying, and evaluating results" were the most common activities (see Multimedia Appendix 1 for the full description of each activity, as provided to the respondents).

Figure 2. Data sources most frequently searched.

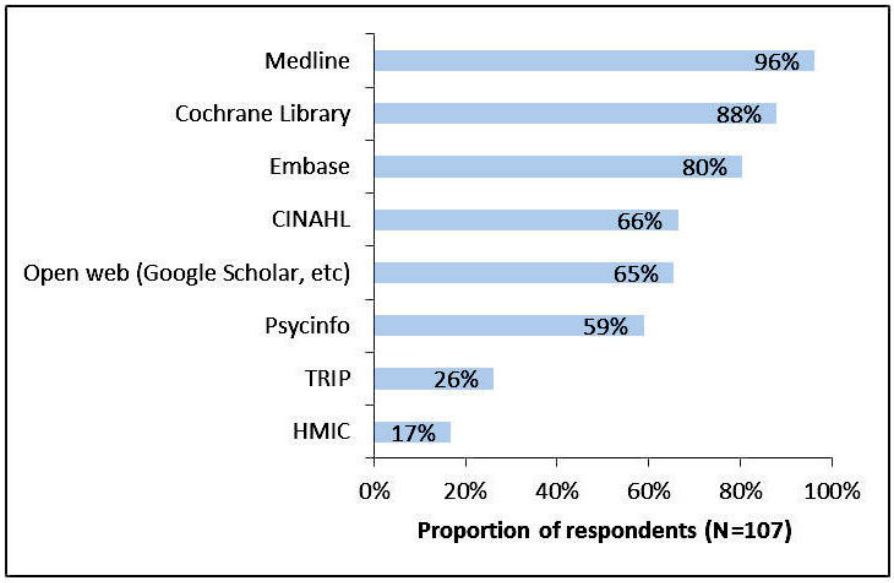


Figure 3. Importance of query formulation functionality.

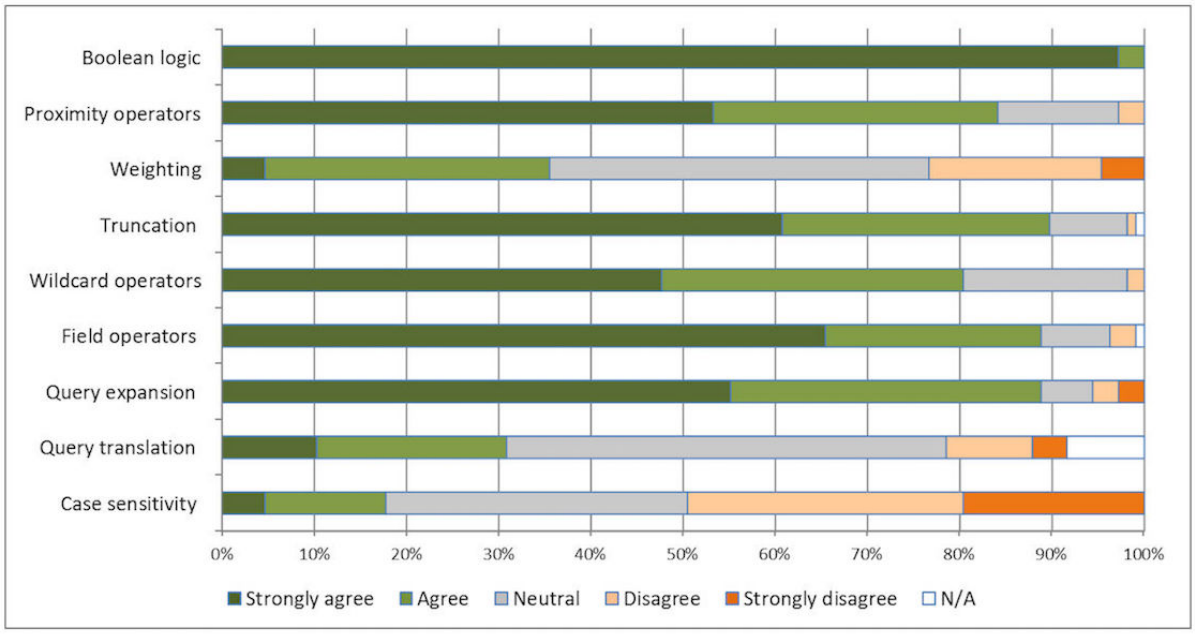


Figure 4. Usage of restriction criteria.

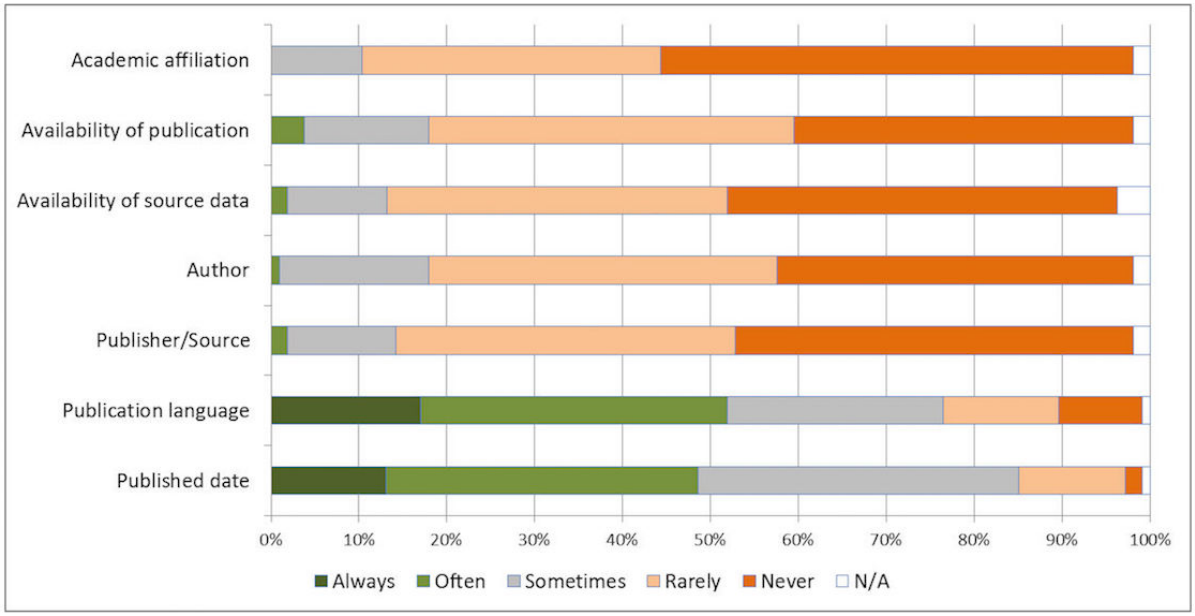


Figure 5. Activities that respondents engage in when completing a search task.

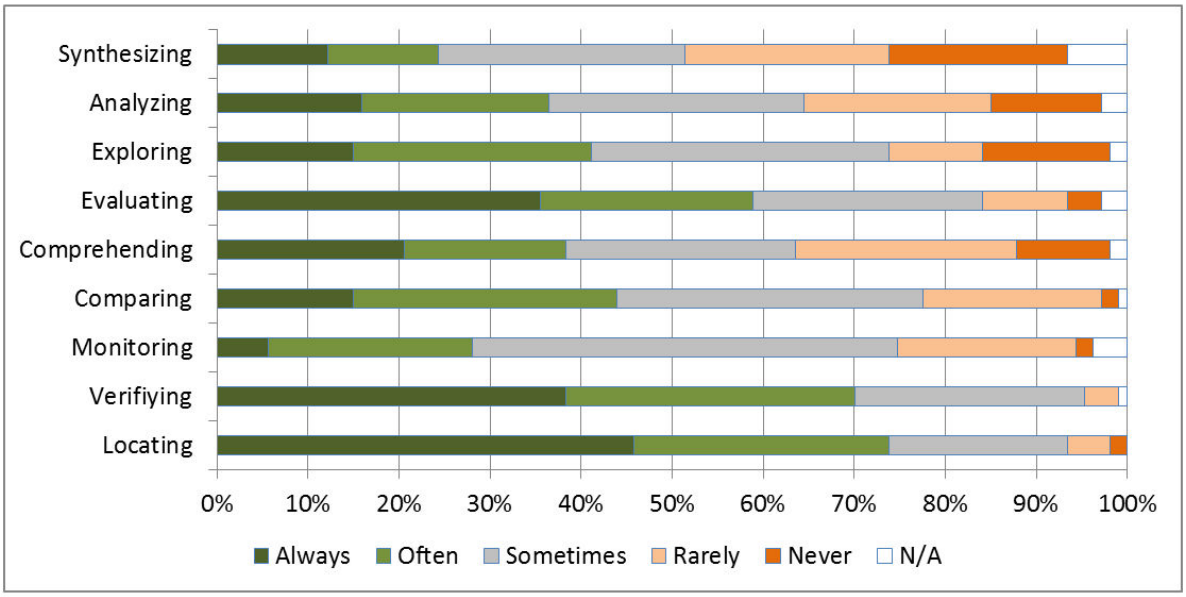
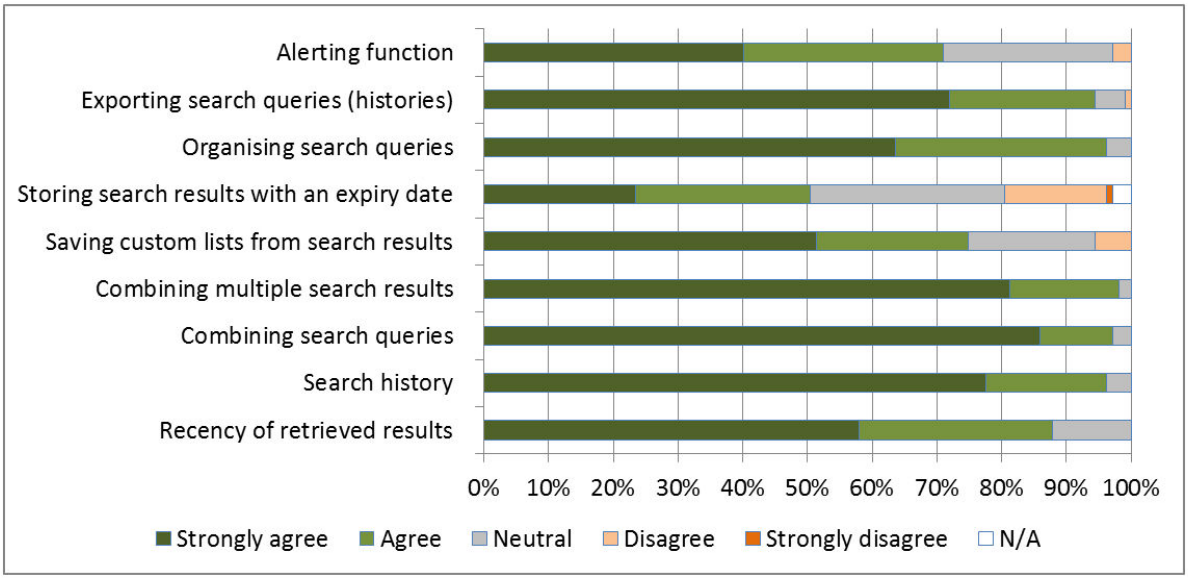


Figure 6. Ideal features of a literature search system.



Ideal Functionality for Searching Databases

We also examined other features related to search management, organization, and history that respondents value when performing search tasks. Respondents were asked to indicate a level of agreement to a statement using a 5-point Likert scale ranging from 1 (strong disagreement) to 5 (strong agreement). The results are shown in Figure 6.

Discussion

Here, the implications of the results with verbatim responses to the question "How could the process of creating and managing search strategies be improved for you?" are discussed and the findings are contextualized in relation to the PubMed literature search system.

Search Tasks

The respondents showed they invest considerable amounts of time performing search tasks and writing search strategies. The time to search a document collection (60 minutes) indicated that their search strategies were more complex to create than most literature search queries, given that 90% of individual queries on PubMed take less than 5 minutes [36]. It is also longer than diagnostic Web searches typical of front-line healthcare professionals (only 14% of medical practitioners reported spending more than 40 minutes on this search task) [23].

This search effort is often recycled and routinely shared indicating a need for facilities to manage and share strategies such as: "...being able to download, share, remix, transfer and translate search strategies." PubMed does not offer the ability to share search queries, only the results in the form of citation Collections.

Query Formulation

The results in Figure 3 suggest 2 observations regarding how healthcare information professionals formulate queries. Firstly, the scores suggest a willingness to adopt a wide range of search functionality to complete search tasks. This represents a marked contrast to the behavior of typical Web searchers who rarely, if ever, use any advanced search functionality [37]. Secondly, the use of Boolean logic was shown to be the most important feature, closely followed by the use of synonyms and related terms. A number of other syntactic features, notably proximity operators, truncation, and wildcarding, all scored highly, reflecting the need for fine control over search strategies. Field operators were also judged to be important, reflecting the structured nature of the document collections that are searched. Query expansion (ie, terms are expanded to include synonyms) scored highly, underlining the key role that controlled vocabularies such as MeSH play in forming effective search strategies (75% of respondents were familiar with using MeSH headings) and a requirement for, ideally, with "one universal thesaurus of medical terminology for all databases".

PubMed offers most of the query formulation functionality described in Figure 3, either through explicit Boolean queries or through related functionality. Simple keyword queries are converted into Boolean queries by using the AND operator, attempting to automatically align the keywords with MeSH terms (called Automatic Term Mapping) and expanding the query to match all search phrases. Boolean operators OR and NOT are also accepted. Users can search specific fields by using square brackets after the search term (such as for searching within abstract, author, title, etc). Spelling correction and phrase completion are offered as the user types into the textbox. Wildcard and truncation is partially supported by allowing right-truncation only (ie, child*) would return results for children and childhood. Proximity operators are not supported; however, PubMed offers a list of related articles derived from a word-weighted algorithm [38]. Search queries can also be made in multiple languages (although the only non-English data in PubMed is currently limited to the "transliterated title" field). The only functionality PubMed does not appear to offer is weighting search terms and case sensitivity, both of which were

rated as the least important functionality by respondents of the survey. This highlights the difference between comprehensive searches for literature as required for a review compared to more general Web searches where relevance ranking with semi-automatic methods would be considered more important.

A previous study has shown that as many as 90% of published strategies contained an error [39] and that reporting of strategies is commonly not in line with best practice [40]. A number of respondents suggested that healthcare information professionals need advanced query formulation support to help them with search tasks (Textbox 2).

Textbox 2. Examples of search functionality that require advanced query formulation support.

Search functionality

- Syntax checking: "…automate checking of parentheses, operators and field codes…"
- Truncation: "Wildcards at beginning of words; wildcard within a word (to replace a single or multiple letters eg, $sthetic or wom$n"
- Misspellings: "…account for misspellings…" and "UK/American spelling…"
- Proximity: "…interpreting proximity within sentence rather than crossing punctuation limits."
- Term frequency and location: "…terms in the first and/or last sentence of the abstract only"
- Negation: "…a negation that doesn't exclude articles where the negated concept is preceded by a negation. ex: NOT "palliative care" will exclude abstracts with sentences like this 'in this study we didn't take in account palliative care'"

PubMed allows users to build queries in stages using an HTML form to capture the query, then listing previous queries below in order for the user to make composite strategies of increasing complexity. Given that the average number of strategy lines required for a search task here was 15, this method of query construction can get increasingly complex and difficult for the user to understand and manipulate. Only 5.7% (10/176) of survey respondents reported using a visual query builder, an indication that there is very little support for healthcare information professionals in the intuitive construction of complex search queries. They also indicated a desire for advanced editing functionality, in particular:

> *move search lines up and down the history…*
>
> *being able to add tags or descriptions to search strategies, ability to sort by name, topic or date...*
>
> *take notes about why you added terms, syntax, etc.*

It is clear that respondents commonly work across multiple platforms, in particular MEDLINE, Cochrane, and Embase, and this is in line with findings from previous research [41]. There is therefore a need for standardization and consistency between suppliers: "A service that could map search strategy between databases would save a lot of time."

Evaluating Search Results

The figure of 100 average (median) ideal search results masks the non-parametric nature of the data; the number of search results obtained may vary considerably depending on the topic and body of literature available in that domain [41]. Healthcare information professionals may adjust their expectations of sensitivity (or recall) in relation to their searches, depending on the need for coverage and inclusiveness. Clearly the ideal response from a search is more nuanced than a single figure can convey; however, respondents indicated that they find more results from their searches than they would ideally like to evaluate. This may be the result of an abundance of published research or that the search parameters are not restrictive enough to return an appropriate number of results.

The time to assess each result (3 minutes) seems short when considering the length of some of the documents that will be analyzed. However, the search task is the first stage of a much longer process in which the retrieved documents are exposed to further phases of evaluation (Figure 5). In this context, the time to assess relevance may reflect the dynamics of the initial sift, which is a much smaller fraction of the overall attention given to a document.

Publication date was considered to be the most important results filtering criteria, followed by publication language (Figure 4); however, other criteria were not considered important in the restriction of results. Certain respondents mentioned other criteria they use including publication type, study scope (eg, human only), study design, age range, and gender. All filtering and restriction criteria mentioned here can be used to narrow down results in PubMed.

The fact that no respondent valued sorting by most trustworthy source contrasts sharply with the strategies used in another study of the healthcare profession [23]. This most likely reflects the difference between the largely curated (and to some extent implicitly trusted) databases referred to in our study (MEDLINE, Embase, and the Cochrane Library) and the relatively uncontrolled Web resources used in Geschwandtner's study.

Ideal Functionality

Respondents scored all options of ideal functionality highly, indicating a general desire for advanced search functionality. Combining search queries and combining search results were rated as the most important, reflecting the current paradigm for building search queries (ie, the line-by-line strategy building approach offered by most databases including PubMed). The participants rated the ability to export search queries (histories) highly, reflecting their need to publish completed search strategies as part of their professional practice.

All of the functionality that is described in Figure 6 as being desired by healthcare information professionals is available through PubMed, either directly or by registering as a free user of My NCBI. It is therefore surprising that the verbatim

responses from the respondents indicated that typical systems fall short in terms of their needs.

One reason may be that PubMed attempts to cater for a wide range of user knowledge (approximately one third of PubMed users are not domain experts [42]) and search expertise, from simple keyword queries to complex search strategies. Query log analysis has shown a difference between how users of different skills perform on PubMed [43] and PubMed attempts to accommodate all their needs in one interface. One example of this compromise is the lack of truncation and proximity operators, which may be exactly what is required by a healthcare information professional performing a systematic review for a topic with few articles.

Limitations

A limitation of this survey is the sample size compared to some surveys of healthcare information professionals [13,14]; however, engagement with professionals in this sector has been shown to be challenging, with lower participation rates reported elsewhere [11,12]. We believe the completion rate of the survey (49.1%, 107/218) is high for a survey of this length (approximately 15 minutes); however, greater participation could have been achieved with a shorter, more targeted survey. We acknowledge that the lack of control over distribution and that it was administered in English only may introduce selection bias. The demographics of this survey have a similar distribution to a larger survey of healthcare information professionals [44] (95% females compared to 86% reported here, with an average age of 47.2 compared to 46.0 here), an indication that the sampled population may be representative of the profession.

A further limitation of this study is whether respondents fully understood our distinction between search tasks and search strategies (which follows the precedent of previous survey designs and hence facilitates direct comparison with their results). An additional evaluation of other literature search tools (such as Ovid) would have provided a more extensive survey of functionality available to healthcare information professionals; however, as PubMed was the most frequently used by the respondents it is more representative of the tools they have at their disposal. A full survey of free and subscription search tools available in healthcare would be useful future work. Despite these limitations we believe the research provides valuable insight into the requirements of healthcare information professionals.

Conclusions

This paper summarizes the results of a survey of the information retrieval practices of healthcare information professionals, focusing in particular on the process of search strategy development. Our findings suggest that they routinely address some of the most challenging information retrieval problems of any profession, but current literature search systems offer only limited support for their requirements. The functionality offered by PubMed goes some way toward meeting those needs, but is compromised by the need to serve all types of users who may not require the same degree of fine control over their search strategies. In particular, there is a need for improved functionality regarding the management of search strategies and the ability to search across multiple databases.

The results of this study will be used to inform the development of future retrieval systems for healthcare information professionals and for others performing healthcare-related search tasks.

Acknowledgments

The authors would like to thank the participants for completing the survey, the professional associations that distributed the survey to their members, and to the healthcare information professionals who helped shape the survey instrument and provide context for the results.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey Instrument.

[PDF File (Adobe PDF File), 425KB - medinform_v5i4e33_app1.pdf]

References

1. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) 2011;2011:baq036 [FREE Full text] [doi: 10.1093/database/baq036] [Medline: 21245076]
2. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Sep 21;7(9):e1000326 [FREE Full text] [doi: 10.1371/journal.pmed.1000326] [Medline: 20877712]
3. Tang H, Ng JH. Googling for a diagnosis--use of Google as a diagnostic aid: internet based study. BMJ 2006 Dec 02;333(7579):1143-1145 [FREE Full text] [doi: 10.1136/bmj.39003.640567.AE] [Medline: 17098763]
4. Kitchens B, Harle CA, Li S. Quality of health-related online search results. Decis Support Syst 2014 Jan;57:454-462. [doi: 10.1016/j.dss.2012.10.050]

5. Hemingway P, Brereton N. Hayward Medical Communications, Second Edition. 2009. What is a systematic review? URL: http://docplayer.net/84335-What-is-a-systematic-review.html [accessed 2017-03-05] [WebCite Cache ID 6ojoyoe87]
6. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JP, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med 2014 Feb;11(2):e1001603 [FREE Full text] [doi: 10.1371/journal.pmed.1001603] [Medline: 24558353]
7. Karimi S, Pohl S, Scholer F, Cavedon L, Zobel J. Boolean versus ranked querying for biomedical systematic reviews. BMC Med Inform Decis Mak 2010 Oct 12;10:58 [FREE Full text] [doi: 10.1186/1472-6947-10-58] [Medline: 20937152]
8. The Cochrane Collaboration. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 URL: http://handbook.cochrane.org/chapter_6/6_searching_for_studies.htm [accessed 2017-03-05] [WebCite Cache ID 6ojmszrFw]
9. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev 2014 Jul 09;3:74 [FREE Full text] [doi: 10.1186/2046-4053-3-74] [Medline: 25005128]
10. Potts HW. Is e-health progressing faster than e-health researchers? J Med Internet Res 2006 Sep 29;8(3):e24 [FREE Full text] [doi: 10.2196/jmir.8.3.e24] [Medline: 17032640]
11. Grant MJ. How does your searching grow? A survey of search preferences and the use of optimal search strategies in the identification of qualitative research. Health Info Libr J 2004 Mar;21(1):21-32 [FREE Full text] [doi: 10.1111/j.1471-1842.2004.00483.x] [Medline: 15023206]
12. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
13. Gillies D, Maxwell H, New K. Cochrane Community. 2009. A collaboration-wide survey of Cochrane authors URL: https://community-archive.cochrane.org/sites/default/files/author_survey_report_20may2009.pdf [accessed 2017-03-05] [WebCite Cache ID 6ojn7jEkZ]
14. Ciapponi A, Glujovsky D. Survey among Cochrane authors about early stages of systematic reviews. 2012 Oct 03 Presented at: The 20th Cochrane Colloquium; 2012 Sep 30 - Oct 3; Auckland, New Zealand.
15. U.S. National Library of Medicine. MEDLINE/PubMed Resources Guide URL: https://www.nlm.nih.gov/bsd/pmresources.html [accessed 2017-03-05] [WebCite Cache ID 6ojnHKRuF]
16. Elsevier. Embase URL: https://www.elsevier.com/solutions/embase-biomedical-research [accessed 2017-08-31] [WebCite Cache ID 6t81eDe9M]
17. Cochrane Library. URL: http://www.cochranelibrary.com [accessed 2017-03-05] [WebCite Cache ID 6ojnKFy27]
18. Clarke J, Wentz R. Pragmatic approach is effective in evidence based health care. BMJ 2000 Sep 02;321(7260):566-567 [FREE Full text] [Medline: 10968827]
19. PubMed. URL: https://www.ncbi.nlm.nih.gov/pubmed [accessed 2017-03-05] [WebCite Cache ID 6ojnTtOsh]
20. Jarvelin K, Ingwersen P. Information Research. 2004. Information seeking research needs extension towards tasks and technology URL: http://files.eric.ed.gov/fulltext/EJ1082037.pdf [accessed 2017-03-05] [WebCite Cache ID 6ojncJyyf]
21. Kuhlthau C. Seeking Meaning: A Process Approach to Library and Information Services. Westport, CT: Libraries Unlimited; 2004.
22. Joho H, Azzopardi L, Vanderbauwhede W. A survey on patent users: an analysis of tasks, behavior, search functionality and system requirements. 2010 Presented at: Proceedings of the 3rd Symposium on Information Interaction in Context (IIiX); 2010 Aug 18-21; New Brunswick, NJ p. A. [doi: 10.1145/1840784.1840789]
23. Geschwandtner M, Kritz M, Boyer C. D8.1.2: Requirements of the Health Professional Search. URL: http://www.khresmoi.eu/assets/Deliverables/WP8/KhresmoiD812.pdf [accessed 2017-03-05] [WebCite Cache ID 6ojnmwTpX]
24. Email Discussion Lists for the UK Education and Research Communities. LIS-MEDICAL home page URL: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=lis-medical [accessed 2017-03-05] [WebCite Cache ID 6ojnq4QcF]
25. Email Discussion Lists for the UK Education and Research Communities. CLIN-LIB home page URL: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=CLIN-LIB [accessed 2017-03-05] [WebCite Cache ID 6ojnwLLC5]
26. Email Discussion Lists for the UK Education and Research Communities. EVIDENCE-BASED-HEALTH home page URL: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=EVIDENCE-BASED-HEALTH [accessed 2017-03-05] [WebCite Cache ID 6ojo09FMg]
27. Expertsearching. URL: http://pss.mlanet.org/mailman/listinfo/expertsearching_pss.mlanet.org [accessed 2017-03-05] [WebCite Cache ID 6ojo3opRM]
28. Cochrane Methods Information Retrieval. Information retrieval methods URL: http://methods.cochrane.org/irmg/welcome [accessed 2017-03-05] [WebCite Cache ID 6ojo5t3mP]
29. Health Libraries Group. URL: https://www.cilip.org.uk/about/special-interest-groups/health-libraries-group [accessed 2017-03-05] [WebCite Cache ID 6ojoA66I2]
30. SurveyMonkey. URL: https://www.surveymonkey.co.uk [accessed 2017-03-05] [WebCite Cache ID 6ojoDxC9r]
31. U.S. National Library of Medicine. PubMed tutorial URL: https://www.nlm.nih.gov/bsd/disted/pubmedtutorial/cover.html [accessed 2017-03-05] [WebCite Cache ID 6ojoHUpB3]
32. Chapman D. Advanced search features of PubMed. J Can Acad Child Adolesc Psychiatry 2009 Feb;18(1):58-59 [FREE Full text] [Medline: 19270851]

33. PubMed Mobile. URL: https://www.ncbi.nlm.nih.gov/m/pubmed [accessed 2017-03-05] [WebCite Cache ID 6ojoNKhk4]
34. Mickan S, Tilson JK, Atherton H, Roberts NW, Heneghan C. Evidence of effectiveness of health care professionals using handheld computers: a scoping review of systematic reviews. J Med Internet Res 2013 Oct 28;15(10):e212 [FREE Full text] [doi: 10.2196/jmir.2530] [Medline: 24165786]
35. Russell-Rose T, Lamantia J, Burrell M. A taxonomy of enterprise search discovery. 2011 Presented at: Proceedings of HCIR; 2011 Oct 19; California, US.
36. Islamaj DR, Murray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. Database (Oxford) 2009;2009:bap018 [FREE Full text] [doi: 10.1093/database/bap018] [Medline: 20157491]
37. Spink A, Wolfram D, Jansen M. Searching the web: the public and their queries. J Assoc Inf Sci Technol 2001;52(3):226-234. [doi: 10.1002/1097-4571]
38. PubMed Related Citations Algorithm. URL: https://ii.nlm.nih.gov/MTI/Details/related.shtml [accessed 2017-03-05] [WebCite Cache ID 6ojoaAmuq]
39. Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. J Clin Epidemiol 2006 Oct;59(10):1057-1063. [doi: 10.1016/j.jclinepi.2006.01.007] [Medline: 16980145]
40. Yoshii A, Plaut DA, McGraw KA, Anderson MJ, Wellik KE. Analysis of the reporting of search strategies in Cochrane systematic reviews. J Med Libr Assoc 2009 Jan;97(1):21-29 [FREE Full text] [doi: 10.3163/1536-5050.97.1.004] [Medline: 19158999]
41. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017 Feb 27;7(2):e012545 [FREE Full text] [doi: 10.1136/bmjopen-2016-012545] [Medline: 28242767]
42. Lacroix E, Mehnert R. The US National Library of Medicine in the 21st century: expanding collections, nontraditional formats, new audiences. Health Info Libr J 2002 Sep;19(3):126-132 [FREE Full text] [Medline: 12390234]
43. Yoo I, Mosa AS. Analysis of PubMed user sessions using a full-day PubMed query log: a comparison of experienced and nonexperienced PubMed users. JMIR Med Inform 2015 Jul 02;3(3):e25 [FREE Full text] [doi: 10.2196/medinform.3740] [Medline: 26139516]
44. Crook G, Newsham D, Abrams K. Health Informatics and Health Information Management. 2014. Human resources outlook 2014-2019 URL: https://www.echima.ca/uploaded/pdf/reports/HI-HIM-HR-Outlook-Report-Final-w-design.pdf [accessed 2017-03-05] [WebCite Cache ID 6ojotlaW5]

Edited by G Eysenbach; submitted 10.03.17; peer-reviewed by A Mosa, A Dunn; comments to author 04.05.17; revised version received 13.06.17; accepted 10.07.17; published 02.10.17.

Please cite as:
Russell-Rose T, Chamberlain J
Expert Search Strategies: The Information Retrieval Practices of Healthcare Information Professionals
JMIR Med Inform 2017;5(4):e33
URL: https://medinform.jmir.org/2017/4/e33/
doi:10.2196/medinform.7680
PMID:28970190

Original Paper

Search and Graph Database Technologies for Biomedical Semantic Indexing: Experimental Analysis

Isabel Segura Bedmar[1], PhD; Paloma Martínez[1], PhD; Adrián Carruana Martín[1], MBA, MD

LaBDA Group, Department of Computer Science, Universidad Carlos III de Madrid, Leganés, Spain

Corresponding Author:
Isabel Segura Bedmar, PhD
LaBDA Group
Department of Computer Science
Universidad Carlos III de Madrid
Avda. Universidad 30
Leganés, 28911
Spain
Phone: 34 916245961
Fax: 34 916245961
Email: isegura@inf.uc3m.es

Abstract

Background: Biomedical semantic indexing is a very useful support tool for human curators in their efforts for indexing and cataloging the biomedical literature.

Objective: The aim of this study was to describe a system to automatically assign Medical Subject Headings (MeSH) to biomedical articles from MEDLINE.

Methods: Our approach relies on the assumption that similar documents should be classified by similar MeSH terms. Although previous work has already exploited the document similarity by using a k-nearest neighbors algorithm, we represent documents as document vectors by search engine indexing and then compute the similarity between documents using cosine similarity. Once the most similar documents for a given input document are retrieved, we rank their MeSH terms to choose the most suitable set for the input document. To do this, we define a scoring function that takes into account the frequency of the term into the set of retrieved documents and the similarity between the input document and each retrieved document. In addition, we implement guidelines proposed by human curators to annotate MEDLINE articles; in particular, the heuristic that says if 3 MeSH terms are proposed to classify an article and they share the same ancestor, they should be replaced by this ancestor. The representation of the MeSH thesaurus as a graph database allows us to employ graph search algorithms to quickly and easily capture hierarchical relationships such as the lowest common ancestor between terms.

Results: Our experiments show promising results with an F1 of 69% on the test dataset.

Conclusions: To the best of our knowledge, this is the first work that combines search and graph database technologies for the task of biomedical semantic indexing. Due to its horizontal scalability, ElasticSearch becomes a real solution to index large collections of documents (such as the bibliographic database MEDLINE). Moreover, the use of graph search algorithms for accessing MeSH information could provide a support tool for cataloging MEDLINE abstracts in real time.

(JMIR Med Inform 2017;5(4):e48) doi:10.2196/medinform.7059

KEYWORDS

information storage and retrieval; semantic indexing; Medical Subject Headings

Introduction

Biomedical Semantic Indexing

The last two decades have witnessed tremendous advances in our knowledge of life sciences and medicine, leading to an exponential growth of the biomedical literature. There are several biomedical bibliographic databases such as EMBASE, OVID, Ebsco Host Research databases, Scielo, Cochrane, and the largest one, with 5600 journals and over 26 million articles, MEDLINE. In 2015, more than 806,000 citations were added to MEDLINE with a load of 2000 to 4000 documents per day. This quickly growing volume of articles is an overwhelming

challenge that requires a very specialized knowledge for organizing this bibliographic database.

To support the classification and indexing of the content of the MEDLINE database, the US National Library of Medicine (NLM) produces and maintains a thesaurus of medical concepts, MeSH (Medical Subject Headings), which is reviewed and updated continually (eg, 310 new headings were added to MeSH in 2015). Each document in MEDLINE is represented with a set of MeSH terms that describe its subject topic. This task, which is generally known as biomedical semantic indexing, is a crucial task to facilitate literature search because MeSH terms can be used in search queries to retrieve references that were annotated with these terms or with their hierarchically related terms in MeSH (ie, their synonyms, hypernyms, or hyponyms). The task of identifying the MeSH terms that best represent a MEDLINE article is manually performed by human experts (so-called curators). NLM also provides some basic principles [1] to assign MeSH terms that curators should follow when they catalog articles.

Biomedical semantic indexing is usually a costly, time-consuming, and laborious task [2]. Therefore, there is an urgent need to explore semiautomatic methods to support semantic indexing.

Several challenges such as Critical Assessment of Information Extraction in Biology (BioCreative) [3], Workshop on Biomedical Natural Language Processing (BioNLP) shared tasks [4,5], Informatics for Integrating Biology & the Bedside (i2b2) [6], and DDIExtraction [7,8] have significantly contributed to improve and advance the state of the art in Natural Language Processing for biomedicine, especially in the information extraction task. Similarly, the biomedical semantic indexing and question answering challenge (BioASQ) is being organized since 2013 to encourage and promote research in these fields and provide a common framework for assessment. The objective of the task is to tag an article with a set of terms (also known as headings or descriptors) from the MeSH thesaurus. In this task, the training data consist of a vast collection of MEDLINE abstracts. Each article includes the MeSH terms that the curators used to classify it. It also contains additional metadata such as its unique identifier number (PubMed unique identifier, PMID) used in PubMed (a free search engine for the MEDLINE database), title, journal name, and publication year (see Figure 1). The test data consist of recently published articles that have not been labeled by the curators yet. The participating systems have to find the best MeSH terms and report their answers for the test data.

Biomedical semantic indexing can be defined as a multilabel hierarchical classification problem because each document has to be classified with one or more concepts from a taxonomy. If the taxonomy has a significant number of concepts (more than hundreds), the main challenge is to work with this large number of classes in the classification problem. In the case of the BioASQ challenge, MeSH has a hierarchy with 16 main branches and contains more than 27,000 terms. Some works restrict the scope of MeSH hierarchy using only a particular branch in the MeSH tree (eg, heart diseases) [9] or a subset of terms (generally those appearing in the training collection) [10] to reduce the difficulty of the multilabel classification problem.

General Architecture

The general architecture of the most state-of-the-art systems comprises 2 differentiated phases: a first phase in which an initial set of MeSH terms is obtained and a second phase that ranks these terms to select the top K that better fit the input document. Several machine-learning techniques have been used such as Support Vector Machines (SVM) [11,12], logistic regression [13], k-nearest neighbors (k-NN) [11,13,14], or a combination of them.

Most previous systems employ either flat classifiers or cascades of classifiers [15]. Flat classifiers [11,16-18] do not take into account the hierarchical relations between the MeSH terms, whereas cascades approaches [19,20] apply a separate classifier top-down for each term. In each term, the method must decide whether to assign the current term to the article being classified or continue descending by the taxonomy and selecting which branches (children) to continue exploring. However, both approaches, flat and cascades, use the BoW (bag-of-words) model to represent the documents. One of the notorious disadvantages of BoW models is that they generate a large number of features (as many as the vocabulary size of the training set), which usually requires prohibitive computation time for practical applications. A possible solution could be the use of feature selection techniques to reduce the number of BoW features. However, these techniques have proved to be inefficient because of the large number of classes (as many as existing terms in MeSH) that must be represented. In other words, as mentioned above, this multilabel classification problem implies more than 20,000 classes (which are the terms stored in MeSH), and it would need to keep at least a few features to represent each class for the classification. Indeed, classifiers used in this problem usually obtain better performance without feature selection [15]. More recently, some works [21,22] use word embedding techniques as an attractive alternative of BoW-based approaches, leading to very large dimensionality reduction and promising results.

Some previous works have implemented different strategies based on the guidelines proposed by human curators to select the most appropriate set of MeSH terms for a given document. However, it is difficult to assess their real utility because human curators, paradoxically, do not always follow their own rules [23].

Table 1 summarizes some of the main systems for the task of biomedical semantic indexing. The underlying characteristics (such as the type of approach: flat vs hierarchical, if the system is based on a search engine, and a brief description of the main techniques used) of these works are presented.

Figure 1. JSON-based format for the training data in the biomedical semantic indexing and question answering challenge BioASQ task 4a.


```
"articles":[
  {
    "abstractText":"The addition of FMNH(2), to Vibrio harveyi luciferase at 2°C in the
    presence of tetradecanal results in the formation of a highly fluorescent transient species
    with a spectral distribution indistinguishable from that of the bioluminescence. The
    bioluminescence reaches maximum intensity in 1.5 s and decays in a complex manner with
    exponential components of 10(-1) s(-1) , 7 x 10(-3)S(-1). and 7 x10(4)s(-1). The
    fluorescent transient rises exponentially at 7 x 10(-2)s(-1) and decays at 3 x 10 (4)s(-1)
    . The slowest bioluminescence component. comprising the bulk of the bioluminescence. decays
    at twice the rate of the fluorescent transient under all variations of reaction conditions:
    concentration of reactants.temperature 2 - 20°C. and aldehyde chain length - decana1,
    dodecanal and tetradecanal. The activation energy for both the slowest bioluminescence
    decay and the transient fluorescence decay is 80 kJ-mol(-1). An energy transfer scheme is
    proposed to explain the results where two distinct chemically energized species utilize the
    fluorescent transient as emitter for the slower bioluminescences, and for the faster
    process a fluorophore present in the protein preparation. Kinetic observations suggest that
    typical preparations of V. harveyi luciferase comprise 15% active protein.",
    "journal":"Photochemistry and photobiology",
    "meshMajor":[
      "Flavin Mononucleotide",
      "Fluorescence",
      "Kinetics",
      "Luciferases",
      "Luminescence",
      "Time Factors",
      "Vibrio"
    ],
    "pmid":"23479819",
    "title":"Kinetics of bacterial bioluminescence and the fluorescent transient.",
    "year":"1983"
  },
```


Table 1. Main works for biomedical semantic indexing.

System	Type	Guidelines	Search engine	Approach	F1
MTI[a], Mork et al [14]	Hierarchical	Yes	PubMed	MetaMap, k-NN[b]	0.548
AUTH-Atypon, Papanikolaou et al [12]	Flat	No	No	SVM[c] with NLP[d] features	0.578
NCBI[e], Mao et al [11]	Flat	No	No	SVM + k-NN	0.605
Antinomyra, Liu et al [13]	Flat	No	No	k-NN + logistic regression	0.619
Ribadas et al [18]	Hierarchical	No	No	Bayesian network	0.615
Kosmopoulos et al [21]	Flat	No	No	k-NN + word embeddings	0.57
Peng et al [22]	Flat	No	No	k-NN + word embeddings	0.632

[a]MTI: Medical Text Indexer.

[b]k-NN: k-nearest neighbors.

[c]SVM: Support Vector Machine.

[d]NLP: Natural Language Processing.

[e]NCBI: National Center for Biotechnology Information.

This study is an extension of our earlier work [24] that described our participation on the BioASQ 2016 biomedical semantic indexing (Task 4a). Our main hypothesis is that similar documents should be classified by similar MeSH terms. Although this hypothesis is not new, and whereas most previous works [11,21,22,25] use document similarity by clustering methods such as k-NN algorithm, our approach exploits document similarity computed by an open source search engine, the ElasticSearch tool [26], one of the most efficient document store databases [27]. To the best of our knowledge, very few works have exploited search engines [14,18]. In particular, the work by Ribadas et al [18] used the search engine tool Indri [28], with the drawback of the high computational time needed for its searches.

Although some works [29,30] have applied the semantic similarity between concepts to the biomedical semantic indexing task, very few works have exploited the curators' guidelines defined by NLM to assign MeSH terms. Our work proposes the implementation of one of the most important annotation rules [1], named "Specific Headings vs Broader Headings," which had not been considered by any of the previous automatic systems. This rule claims that if 3 MeSH terms are proposed to classify an article and share the same ancestor, then the curator should replace these terms by their lowest common ancestor. To do this, the MeSH thesaurus is represented as a graph database. This model based on graph theory leads to query the thesaurus much faster than using a relation database. It enables to swiftly and effortlessly capture hierarchical relationships such as the shortest path between 2 terms or their lowest common ancestor, which are features very useful to decrease the unnecessary overlapping of MeSH terms when an abstract is classified.

The rest of the paper is organized as follows: first, in the Methods section, we give a description of the datasets used in this study and explain our approach. Then, we report and discuss the results of our method in the Results section. Finally, conclusions and future work are presented.

Methods

Objective

The goal of the task was to automatically predict the most descriptive MeSH terms for a given article. The predictions should be compared with MeSH terms that were assigned by human curators. This section describes the MeSH resource, the data, and approach used in this study.

MeSH

MeSH is a thesaurus of medical concepts, which was created to assist human curators in the task of cataloging the articles in the MEDLINE database. Thus, each MEDLINE document should be represented with a set of MeSH terms that describe its subject topic. MeSH is an annually updated document (eg, 310 new headings were added to MeSH in 2015). The MeSH 2016 version contains a total of 27,883 main terms (also known as headings or descriptors), 82 qualifiers (subheadings), and more than 232,000 supplementary concept records, which represent specific examples of chemicals, diseases, and drug protocols.

In MeSH, most terms contain a short definition, links to related descriptors, a list of synonyms or very similar terms, and a unique alphanumerical ID. Figure 2 shows the content for the term "Lymphoma." The terms are organized in a hierarchy in which each child can have more than one parent. Therefore, any MeSH term can appear at different branches of the hierarchical structure of MeSH. For example, the term "Lymphoma" belongs to 3 different branches: "Neoplasms [C04]," "Hemic and Lymphatic Diseases [C15]," and "Immunologic Diseases [C20]." The field "Tree Number" represents each possible location of a term in MeSH. Thus, the term "Lymphoma" has 3 tree numbers: C04.557.386, C15.604.515.569, and C20.683.515.761; C stands for Diseases, C04 for Neoplasms, and C04.557 for Neoplasms by Histologic Type; C15 for Hemic and Lymphatic Diseases, C15.604 for Lymphatic Diseases, and C15.604.515 for Lymphoproliferative Disorders; C20 for Immune System Diseases, C20.683 for Immunoproliferative Disorders, and C20.683.515 for Lymphoproliferative Disorders.

Data

The training data for the BioASQ Task 4a consisted of MEDLINE articles that were manually annotated with MeSH terms by human curators. During the BioASQ 2016 challenge, a test dataset was published each week for the assessment of the participating systems. A total of 15 test datasets were published, which were grouped into 3 different periods (batches). Although the BioASQ challenge ended last May 15, 2016, the test datasets with gold annotations were not released because many articles have not been manually annotated yet.

Figure 2. Medical Subject Headings (MeSH) descriptor data for the term "Lymphoma".

MeSH Heading	Lymphoma
Tree Number	C04.557.386
Tree Number	C15.604.515.569
Tree Number	C20.683.515.761
Annotation	GEN only or unspecified: prefer specific; do not confuse X ref LYMPHOMA, MALIGNANT with LYMPHOGRANULOMA, MALIGNANT see HODGKIN DISEASE; for lymphoma with AIDS, use LYMPHOMA, AIDS-RELATED
Scope Note	A general term for various neoplastic diseases of the lymphoid tissue.
Entry Term	Germinoblastoma
Entry Term	Lymphoma, Malignant
Entry Term	Reticulolymphosarcoma
Entry Term	Sarcoma, Germinoblastic
Allowable Qualifiers	BL CF CH CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SU TH UL UR US VE VI
Date of Entry	19990101
Unique ID	D008223

Two different versions of the training data were provided: (1) Training v.2016a with more than 12 million documents and (2) Training v.2016b with almost 5 million documents from the pool of journals that the BioASQ organizers used to select the articles for the test data. In both datasets, the average number of MeSH terms assigned to an article was 12 to 13.

In our previous work [24], we performed several experiments using each of the 2 training datasets, which led to the conclusion that they did not make a significant difference on the performance of our system. For this reason, we decided to only use the largest dataset (Training v.2016a) to perform all of the experiments described in this new work (see the Results section). Moreover, to optimize the best setting of our approach, we randomly chose 1099 documents from the training dataset and separated them for development set.

As mentioned before, no test datasets with gold standard annotations were released. However, to perform a transparent and consistent evaluation of our work, we developed a script that obtains the MeSH terms for all abstracts in the test batches of the 2016 BioASQ. For each test document, the script obtains its PMID and then generates a query for searching it in PubMed. If the PMID exists in MEDLINE, PubMed returns a structured document containing the metadata for this abstract, among them its MeSH labels (see Figure 3), collected by the script using a regular expression. Finally, the labels are also searched in the MeSH resource to obtain their corresponding MeSH identifiers. In this way, we obtained the same 15 test datasets used in the 2016 BioASQ edition. Table 2 shows the size of the different datasets used in this study.

Figure 3. MeSH terms for the abstract with Pubmed unique identifier (PMID)=26852276.

```
MH  - Adolescent
MH  - Adult
MH  - Aged
MH  - Aged, 80 and over
MH  - Antineoplastic Combined Chemotherapy Protocols/*therapeutic use
MH  - Combined Modality Therapy
MH  - Datasets as Topic
MH  - Female
MH  - Follow-Up Studies
MH  - Humans
MH  - Immunotherapy
MH  - Insurance Coverage/statistics & numerical data
MH  - Kaplan-Meier Estimate
MH  - Lymphoma, B-Cell/drug therapy/*radiotherapy
MH  - Male
MH  - Mediastinal Neoplasms/drug therapy/*radiotherapy
MH  - Middle Aged
MH  - Neoplasm Staging
MH  - Prognosis
MH  - Propensity Score
MH  - Proportional Hazards Models
MH  - Registries
MH  - Risk Factors
MH  - Rituximab/administration & dosage
MH  - Treatment Outcome
MH  - United States
MH  - Young Adult
```

Table 2. Size of datasets (number of documents).

Dataset	Documents, n
Training	10,099,281
Development	1099
Test	13,936

Indexing Documents and Query (Test Document) Using ElasticSearch

Our approach relies on the assumption that similar documents should be classified by similar MeSH terms. Previous research has generally used document clustering techniques, such as the k-NN algorithm, to obtain the similar documents for a given test document. Instead of using k-NN, we proposed the use of an open source search engine, ElasticSearch, to retrieve a set of similar documents for each test document.

Figure 4 shows the main steps of our approach. ElasticSearch was used to index all the documents of the training dataset (Training v.2016a). Each training document was stored along with its corresponding MeSH terms. Each test document was also represented as a query, which was fired against the index built from the training dataset. Then, ElasticSearch should return the most relevant (similar) documents to the query (the test document). Finally, our system initially assigns it all the MeSH terms of the similar documents retrieved by ElasticSearch for this document.

Below we explain in detail how the index was constructed and how a query (a test document) could be compared against this index to recover the most relevant (similar) documents.

The core of ElasticSearch is Apache Lucene, a free, open-source, and *de facto* standard retrieval software library (by The Apache Software Foundation). The efficiency of Lucene is because it searches on index instead of searching the text directly. Moreover, the index is stored in the main memory.

Lucene is based on the well-known and commonly used vector space model (VSM) for information retrieval. This model allows us to represent documents as vectors, where each position in the vector represents a specific term (typically terms are single words), and the value at that position denotes the weight of that term. There are several different ways of computing these values, being the most known term frequency-inverse document frequency (tf-idf) weighting. In this model, a given document d is represented as a vector $v_d=[w_{1,d}, w_{2,d},\ldots, w_{N,d}]$, where $w_{i,d}$ represents the frequency of the term i in the document d, D is the set of all documents, and $|\{d' \in D | I \in d'\}|$ is the number of documents containing the term I (see Figure 5).

In short, VSM represents documents and queries as weighted vectors, where each dimension refers to an index term and its value is its tf-idf value. To assess the relevance of a document d for a given query q, VSM calculates the cosine similarity of their vectors (see Figure 6). Therefore, the basic idea behind VMS is that the more frequent a term is in a document relative to its frequency in the whole collection of documents, the more relevant that document is to the query.

Another important advantage of ElasticSearch is its capacity to create distributed and scalable systems by specifying only the configuration of the hierarchy of nodes. Thus, ElasticSearch is self-managed to maintain better fault tolerance and load distribution. In 2014, an empirical evaluation study about the effectiveness of the current databases demonstrated that ElasticSearch achieved the best performance compared with other document store databases [27]. This is because ElasticSearch uses the main memory and compresses documents, thereby improving retrieval time. Moreover, another main challenge of the task is to manage the great amount of documents that have to be indexed. Thanks to its horizontal scalability (ie, the possibility of adding more storage and processing power), ElasticSearch is able to index large collections of documents such as the MEDLINE database.

In this study, ElasticSearch (version 5.0) was installed on an Ubuntu 16.04 server with 24 GB of RAM and 500 GB of disk space. It took 10,264.07 seconds to index all the training dataset (ie, an average of 1.02 milliseconds per document). The training dataset (Training v.2016a) consists of a total of 10,100,380 documents, with an average size of 2.1 KB per document.

MTI Processing

The Medical Text Indexer (MTI) [14] is a tool developed by NLM and is considered as a baseline system for the task, which provides a preliminary annotation of the articles. MTI is based on a combination of MetaMap- [31] and PubMed-related citations to recognize MeSH terms that are then clustered and ranked by a k-NN algorithm. Given a document, MTI uses MetaMap to find its concepts. The UMLS (Unified Medical Language System) concepts found by MetaMap are restricted to MeSH by a combination of synonym and interconcept relations, and mappings. MTI also obtains a second list of MeSH terms by obtaining similar documents for the input document. To do this, MTI uses the list of PubMed-related citations provided by the PubMed system. Then, the MeSH terms of these similar documents are also extracted. Finally, MTI clusters both lists of MeSH terms into a single list. Terms are clustered by a k-NN algorithm and ranked according to the product of the frequency and the MeSH tree depth of each term. MTI also includes a postprocessing phase that implements a set of filtering rules from the NLM guidelines. For instance, it contains a list of triggers that activate one or more MeSH tags and that comes mainly from the NLM guidelines, in the way of rules such as "if XXXX appears in the text then you should tag as AAAA."

As it was mentioned before, our system initially considered the set of MeSH terms from the relevant documents retrieved by ElasticSearch for a given test document. Then, that set was further extended with those terms provided by the MTI tool.

Figure 4. Architecture of our system.

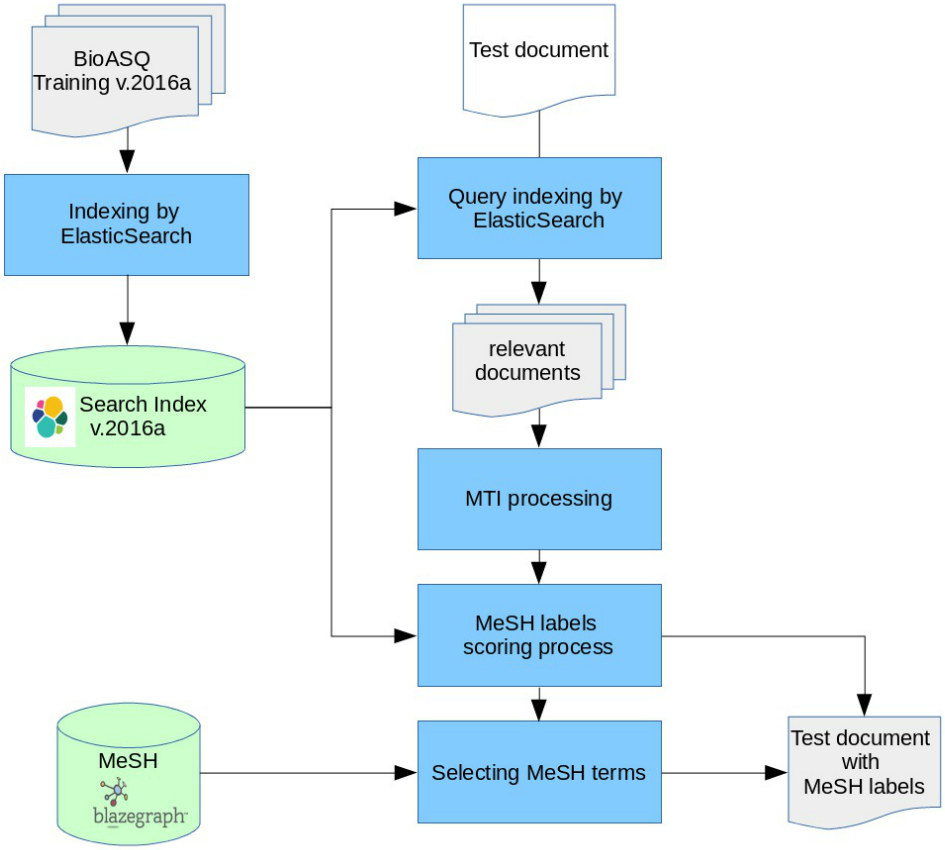


Figure 5. The element wi,d is the frequency of the term i in the document d.

$$w_{i,d} = tf_{i,d} * log \frac{|D|}{|\{d' \in D | i \in d'\}|} \text{ and } tf_{i,d}$$

Figure 6. Cosine similarity between a document d and a query w, where V(q).V(d) is the dot product of their vectors, and |V(q)| and |V(d)| are their Euclidean norms.

$$cosine - similarity(q,d) = \frac{V(q) * V(d)}{|V(q)||V(d)|}$$

MeSH Labels Scoring Process

In the previous two sections, we described how an initial set of MeSH terms is proposed by ElasticSearch and later extended by the MTI tool, for a given test document. In this section, we introduce a new scoring function to rank the MeSH terms for a given test document (represented as a query *q*). The basic idea behind this scoring function is the more number of times a MeSH term appears in the set of more relevant documents for a given test document (query), the more significant that term is to this test document. The scoring function (see Figure 7) for a MeSH term *l* and a test document *q* considers the following parameters:

tf(l): the frequency of the MeSH term *l* in the set of retrieved documents by ElasticSearch for the document *q (query)*.

$\Sigma_{d:l \in d}$ *score (d, q)* is the sum of all scores of the relevant documents to the query *q*, which also contain the MeSH term *l*. As mentioned before, ElasticSearch uses the cosine similarity function to obtain the score between a document and a query. We normalized the sum of all scores because some documents may present a large number of MeSH terms, whereas others very few. To do that, we divided it by the maximum score of the relevant documents containing the term *l*.

T is a real positive value that represents the minimum threshold for the scores of the MeSH terms. That is, only the MeSH terms whose scores are greater than *T* finally will be selected for cataloging the test document *q*.

Figure 7. Scoring function to rank Medical Subject Headings (MeSH) term.

$$rank\_score(l,q)_n = tf(l) * \frac{\sum_{d:l \in d} score(d,q)}{max\{score(a,q):l \in a\}} > T$$

Selecting MeSH Terms by Exploiting a Graph Database

In this point, we already have a set of ranked MeSH terms for a given test document.

In the last phase, we implemented a heuristic based on the guidelines of human annotators [1] to classify MEDLINE articles. In particular, the implemented rule claimed that if an abstract had 3 or more MeSH terms sharing some ancestor, then the curators should replace these 3 terms by their lowest common ancestor.

Our hypothesis here was that representing the MeSH thesaurus as a graph would let to query the MeSH thesaurus much faster than when using its original format. By using well-known graph search algorithms such as depth-first search, the model graph enabled to rapidly and easily capture hierarchical relationships such as the shortest path between 2 terms or their lowest common ancestor. Knowing these hierarchical relationships allowed us to find the most appropriate MeSH terms for a given abstract, decreasing the possible overlapping among them, as the NLM recommends.

BlazeGraph [32] is a graph database with support for Java APIs (Application Program Interface) and standardized query languages for graphs, such as SPARQL (Protocol and RDF Query Language). An important advantage of BlazeGraph is that it processes large graphs in near-real time by its GPU (Graphical Processor Unit) acceleration achieving better processing time than CPU (Central Processing Unit) technologies or other graph databases based on key values.

NLM provides a beta version of the MeSH thesaurus in RDF (Resource Description Framework), a standard format for linked open data. This RDF version of MeSH can be loaded into BlazeGraph using the dotNetRDF API, a free and open-source project for working with RDF, SPARQL, and the Semantic Web.

We also developed an algorithm that, given an input document, traverses each of the MeSH terms proposed in the previous step and searches its ancestors by querying the graph database of MeSH with the depth-first search algorithm. Finally, when our algorithm finds out that 3 or more of its MeSH terms share the same ancestor, it replaces them by their lowest common ancestor.

Initially, we restricted the search to a given depth of ancestors, that is, pruning the search subtree below to a given height. However, because the maximum depth is relatively small (consisting only of 9 levels, with an average depth of approximately 4.5 levels), we decided to explore the complete tree of ancestors for each term. Figure 8 shows the query used by ElasticSearch to retrieve all ancestors of the term “Lymphoma.” The output of this query is shown in Figure 9 where the term “Lymphoma” is in 3 different branches of the MeSH thesaurus: C04-Neoplasms, C15-Hemic and Lymphatic Diseases, and C20-Immune System Diseases. M

Table 3 shows the list of MeSH terms proposed by our system for the article with PMID=25676421. The first column contains the MeSH terms after applying our script to replace the terms (3 or more) sharing the same ancestor, whereas the second one contains the MeSH terms proposed by using only ElasticSearch and the score function. For example, the terms “Lymphoma, B-Cell,” “Ataxia Telangiectasia,” and “Lymphoma” were substituted by their lowest common ancestor “Immune System Diseases.”

Table 4 shows the comparison of search times for 3 different MeSH terms. The reader can see that the 3 searches on the MeSH thesaurus stored into a graph database are significantly faster than the same searches on the RDF format.

Figure 8. BlazeGraph query to obtain the ancestors of the term "Lymphoma".

```
PREFIX mesh2016: <http://id.nlm.nih.gov/mesh/2016/>
PREFIX meshv: <http://id.nlm.nih.gov/mesh/vocab#>

SELECT ?treeNum ?ancestorTreeNum ?ancestor ?alabel

WHERE {
  ?nodo rdfs:label "Lymphoma"@en .
  ?nodo meshv:treeNumber ?treeNum .
  ?treeNum meshv:parentTreeNumber+ ?ancestorTreeNum
  ?ancestor meshv:treeNumber ?ancestorTreeNum .
  ?ancestor rdfs:label ?alabel
}

ORDER BY ?treeNum ?ancestorTreeNum
```

Figure 9. List of ancestors for the term "Lymphoma" provided by BlazeGraph.

treeNum	ancestorTreeNum	ancestor	alabel
MeSH2016:C04.557.386	MeSH2016:C04	MeSH2016:D009369	Neoplasms
MeSH2016:C04.557.386	MeSH2016:C04.557	MeSH2016:D009370	Neoplasms by Histologic Type
MeSH2016:C15.604.515.569	MeSH2016:C15	MeSH2016:D006425	Hemic and Lymphatic Diseases
MeSH2016:C15.604.515.569	MeSH2016:C15.604	MeSH2016:D008206	Lymphatic Diseases
MeSH2016:C15.604.515.569	MeSH2016:C15.604.515	MeSH2016:D008232	Lymphoproliferative Disorders
MeSH2016:C20.683.515.761	MeSH2016:C20	MeSH2016:D007154	Immune System Diseases
MeSH2016:C20.683.515.761	MeSH2016:C20.683	MeSH2016:D007160	Immunoproliferative Disorders
MeSH2016:C20.683.515.761	MeSH2016:C20.683.515	MeSH2016:D008232	Lymphoproliferative Disorders

Table 3. MeSH (Medical Subject Headings) terms proposed by our system for the article with PMID (PubMed unique identifier)=25676421.

MeSH[a] exploiting the hierarchy of MeSH	MeSH terms
Ataxia Telangiectasia Mutated Proteins	Ataxia Telangiectasia Mutated Proteins
B-Lymphocytes	B-Lymphocytes
Cell Cycle Proteins	Cell Cycle Proteins
DNA-Binding Proteins	DNA-Binding Proteins
Humans	Humans
Protein-Serine-Threonine Kinases	Protein-Serine-Threonine Kinases
Animals	Animals
Genomic Instability	Genomic Instability
Mice, Knockout	Mice, Knockout
Cyclin D1 Mice In Situ Hybridization, Fluorescence	Cyclin D1 In Situ Hybridization, Fluorescence
Immune System Diseases	Lymphoma, B-Cell Ataxia Telangiectasia Lymphoma

[a]MeSH: Medical Subject Headings.

Table 4. Comparison of search times on the Resource Description Framework (RDF) format and the graph database of the MeSH (Medical Subject Headings) thesaurus.

MeSH[a] terms	RDF[b] in ms[c]	Graph database in ms
Lymphoma, B-Cell	193.39	112
Cyclin D1	210.44	100
Mice, Knockout	239.86	130

[a]MeSH: Medical Subject Headings.

[b]RDF: Resource Description Framework.

[c]ms: milliseconds.

Results

Design of the Experiments

This section conducts an exhaustive set of experiments, where different parameters and options are evaluated on the development dataset to determine the best setting for our system, which will finally be evaluated on the test datasets.

In BioASQ, the performance of the participating systems is evaluated using standard IR measures (eg, precision, recall, and F1), as well as hierarchical variants of them, such as the lowest common ancestor Precision (LCA_P), Recall (LCA_R) and F-measure (LCA-F). The reader can find a detailed explication of these measures in the article [33]. The HEMKit software [34], a tool that implements these measures and lets to easily evaluate the results of different experiments, was used to provide the scores.

Our experiments aimed to answer the following questions:

What is the effect of the number of relevant documents retrieved by ElasticSearch? It is expected that the more documents the search engine obtains, the higher the recall and the lower the precision of our system. We experimented with different number of relevant documents to obtain the best balance between precision and recall, that is, the best F1. In particular, we tried with 10, 20, 30, 40, and 50 documents.

What is the best threshold T that we should consider in our scoring function? Higher values of this threshold should provide a high precision but with a significant decrease of recall. Our objective was to determine the optimum value of this parameter T, that is, that value that obtains the highest F1.

Does the use of the hierarchical structure of MeSH improve the performance of our system? In particular, we assess whether the strategy of replacing terms sharing the same ancestor by their lowest common ancestor helped to improve the performance.

Experiment With/Without exploiting MeSH Hierarchical Structure

Tables 5 and 6 show the results exploiting the hierarchical structure of MeSH and without it, respectively. Each experiment is represented with the label *Elastic-X-T*, where X refers to the number of relevant documents retrieved by ElasticSearch and T to the threshold for our scoring function.

We tried with different number of retrieved relevant documents; in particular, the parameter X could take the following values: 10, 20, 30, 40, and 50. Although increasing the number of retrieved relevant documents achieves to improve the recall, it has a very negative effect on the precision of our system. Indeed, the best F1 (if we do not use the structure of MeSH, we obtain F1=0.70) is obtained with the lowest number of retrieved relevant documents regardless the value of the threshold T (see Tables 5 and 6). Therefore, we can conclude that the best value of X is 10. For values less than 10, the recall decreases significantly. In other words, the system achieves better performance if the search engine is set up to return at least 10 documents.

To assess the effect of the threshold T on the performance of our system, we tried with different values. Tables 5 and 6 show the results for values of T in range (0,9). The reader can see that, in general, the greater the value of the parameter T, the higher the precision, and also the maximum F1. However, the recall decreases when increasing the value of T. Any value lower than 1 achieves a very high recall but very low precision because the system would return all MeSH terms obtained by ElasticSearch along with those provided by the MTI tool, without applying any filter. That is, if the value of T is lower than 1, the scoring function does not rule out any term from the initial set of MeSH terms proposed by ElasticSearch and MTI. On the other hand, for values of T up to 5, the performance begins to drop. In general, best results are obtained for T equal to 5.

Table 5. Experimental results on our development dataset exploiting the hierarchical structure of Medical Subject Headings (MeSH).

Elastic-X-T	Precision	Recall	F1	LCA-P[a]	LCA-R[b]	LCA-F[c]
Elastic-10-0	0.3021	0.8784	0.4386	0.2061	0.6046	0.3006
Elastic-10-1.5	0.6290	0.6213	0.6039	0.4146	0.3979	0.3880
Elastic-10-2.5	0.6599	0.6214	0.6179	0.4376	0.3981	0.3982
Elastic-10-4	0.7371	0.6130	0.6466	0.4936	0.3927	0.4179
Elastic-10-5	0.7898	0.5987	0.6576	0.5316	0.3843	0.4256
Elastic-10-6	0.7434	0.6107	0.6475	0.4986	0.3914	0.4185
Elastic-10-7	0.7904	0.5980	0.6573	0.5321	0.3840	0.4255
Elastic-10-8	0.7968	0.5937	0.6566	0.5372	0.3415	0.4254
Elastic-10-9	0.7910	0.5976	0.6571	0.5325	0.3838	0.4255
Elastic-20-0	0.2174	0.9248	0.3441	0.1564	0.6530	0.2475
Elastic-20-1.5	0.5268	0.6303	0.5546	0.3396	0.4045	0.3542
Elastic-20-2.5	0.5723	0.6331	0.5803	0.3709	0.4044	0.3703
Elastic-20-4	0.6266	0.6332	0.6080	0.4108	0.4047	0.3901
Elastic-20-5	0.6879	0.6294	0.6350	0.4580	0.4037	0.4100
Elastic-20-6	0.6413	0.6333	0.6150	0.4221	0.4054	0.3953
Elastic-20-7	0.6914	0.6296	0.6367	0.4605	0.4039	0.4112
Elastic-20-8	0.7104	0.6279	0.6441	0.4755	0.4041	0.4173
Elastic-20-9	0.6945	0.6299	0.6383	0.4630	0.4043	0.4124
Elastic-30-0	0.1790	0.9434	0.2945	0.1331	0.6770	0.2185
Elastic-30-1.5	0.4776	0.6388	0.5290	0.3040	0.4108	0.3359
Elastic-30-2.5	0.5231	0.6341	0.5537	0.3374	0.4076	0.3537
Elastic-30-4	0.5652	0.6323	0.5766	0.3668	0.4046	0.3685
Elastic-30-5	0.6200	0.6354	0.6067	0.4063	0.4060	0.3888
Elastic-30-6	0.5831	0.6332	0.5864	0.3786	0.4050	0.3748
Elastic-30-7	0.6256	0.6359	0.6096	0.4104	0.4069	0.3911
Elastic-30-8	0.6506	0.6369	0.6217	0.4293	0.4070	0.3993
Elastic-30-9	0.6302	0.6364	0.6120	0.4139	0.4069	0.3928
Elastic-40-0	0.1555	0.9532	0.2621	0.1184	0.6924	0.1988
Elastic-40-1.5	0.4412	0.6473	0.5081	0.2801	0.4166	0.3227
Elastic-40-2.5	0.4915	0.6383	0.5366	0.3145	0.4106	0.3417
Elastic-40-4	0.5302	0.6343	0.5582	0.3404	0.4077	0.3556
Elastic-40-5	0.5755	0.6359	0.5840	0.3726	0.4069	0.3726
Elastic-40-6	0.5472	0.6356	0.5682	0.3521	0.4073	0.3618
Elastic-40-7	0.5819	0.6370	0.5878	0.3777	0.4083	0.3758
Elastic-40-8	0.6073	0.6374	0.6011	0.3959	0.4077	0.3847
Elastic-40-9	0.5870	0.6374	0.5908	0.3813	0.4082	0.3777
Elastic-50-0	0.1395	0.9603	0,239	0.1082	0.7045	0.1846
Elastic-50-1.5	0.4161	0.6542	0.4930	0.2628	0.4226	0.3127
Elastic-50-2.5	0.4669	0.6445	0.5239	0.2965	0.4151	0.3328
Elastic-50-4	0.5008	0.6392	0.5431	0.3192	0.4112	0.3452
Elastic-50-5	0.5447	0.6357	0.5670	0.3500	0.4081	0.3610
Elastic-50-6	0.5192	0.6390	0.5538	0.3324	0.4096	0.3518

Elastic-X-T	Precision	Recall	F1	LCA-P[a]	LCA-R[b]	LCA-F[c]
Elastic-50-7	0.5507	0.6361	0.5702	0.3548	0.4086	0.3637
Elastic-50-8	0.5767	0.6366	0.5849	0.3734	0.4074	0.3734
Elastic-50-9	0.5560	0.6360	0.5733	0.3585	0.4081	0.3656

[a]LCA-P: lowest common ancestor Precision.

[b]LCA-R: lowest common ancestor Recall.

[c]LCA-F: lowest common ancestor F-measure.

The exploitation of the hierarchical structure of MeSH does not improve the results; on the contrary, the recall is dropped almost by 5% (see Tables 5 and 6). Therefore, we can conclude that the strategy of replacing terms sharing the same ancestor by their lowest common ancestor does not increase the results. A possible explication for this fact could be that human curators do not to follow the annotation guidelines.

The pattern of the hierarchical scores (LCA-P, LCA-R, and LCA-F1) according to the different parameters is very similar to the behavior of the flat scores. That is, the best hierarchical scores are usually obtained using the lowest number of retrieved relevant documents and the threshold of the score function equal to 8. Likewise in the flat setting, the rule of replacing 3 or more MeSH terms by their lowest common ancestor does not seem to improve the results.

Experiments on BioASQ 2016 Test Dataset

Finally, we ran the best setting (X=10, T=5) on the test datasets. Tables 7 and 8 show the results of this setting exploiting the structure of MeSH and those without it, respectively. As in the development dataset, the performance is better if we do not use the structure of MeSH.

As mentioned above, the MTI system is considered the baseline for the task. Table 9 shows the results achieved by MTI on each test set published in the 2016 BioASQ. The top F1 is 0.5196 and top LCA-F is 0.4807.

Table 10 shows the temporary scores of the best systems in BioASQ Task 4a. The reader can see that the best F1 rates are between 58% and 65%, the best recall between 54% and 60%, and the best precision between 60% and 72%, depending on the batch. Our approach that does not exploit the hierarchical structure of MeSH seems to obtain better performance than the top systems (see Table 8). Our best F1 is 0.70 (batch 1, week 1). On the other hand, if our system uses the hierarchical relations of MeSH to select the best set of terms to label a given article, this obtains an F1 of 0.67, also better than the top F1 (0.61) of the best systems. Therefore, we can conclude that our approach achieves to overcome the top participating systems at the BioASQ 2016.

Table 6. Experimental results on our development dataset without using the hierarchical structure of Medical Subject Headings.

Systems	Precision	Recall	F1	LCA-P[a]	LCA-R[b]	LCA-F[c]
Elastic-10-0	0.4201	0.6273	0.4858	0.2678	0.4074	0.3104
Elastic-10-1.5	0.5737	0.7755	0.6439	0.3749	0.5260	0.4258
Elastic-10-2.5	0.6128	0.7598	0.6602	0.4017	0.5151	0.4374
Elastic-10-4	0.7102	0.7125	0.6927	0.4701	0.4812	0.4599
Elastic-10-5	0.7724	0.6755	0.7010	0.5141	0.4515	0.4636
Elastic-10-6	0.7178	0.7074	0.6935	0.4761	0.4773	0.4605
Elastic-10-7	0.7731	0.6746	0.7007	0.5149	0.4508	0.4634
Elastic-10-8	0.7803	0.6684	0.6997	0.5204	0.4456	0.4624
Elastic-10-9	0.7738	0.6740	0.7005	0.5154	0.4505	0.4634
Elastic-20-0	0.3498	0.6548	0.4413	0.2229	0.4274	0.2829
Elastic-20-1.5	0.4263	0.8527	0.5559	0.2821	0.5856	0.3723
Elastic-20-2.5	0.4859	0.8324	0.5982	0.3191	0.5702	0.3982
Elastic-20-4	0.5631	0.8008	0.6458	0.3678	0.5459	0.4280
Elastic-20-5	0.6433	0.7643	0.6820	0.4230	0.5213	0.4534
Elastic-20-6	0.5822	0.7926	0.6547	0.3811	0.5406	0.4345
Elastic-20-7	0.6479	0.7627	0.6838	0.4265	0.5203	0.4549
Elastic-20-8	0.6713	0.7515	0.6917	0.4434	0.5131	0.4609
Elastic-20-9	0.6518	0.7608	0.6852	0.4296	0.5195	0.4563
Elastic-30-0	0.3141	0.6747	0.4152	0.2023	0.4444	0.2690
Elastic-30-1.5	0.3538	0.8876	0.4950	0.2380	0.6146	0.3362
Elastic-30-2.5	0.4165	0.8668	0.5492	0.2760	0.5972	0.3686
Elastic-30-4	0.4769	0.8429	0.5956	0.3124	0.5773	0.3959
Elastic-30-5	0.5528	0.8115	0.6428	0.3602	0.5541	0.4254
Elastic-30-6	0.5016	0.8336	0.6113	0.3281	0.5705	0.4059
Elastic-30-7	0.5602	0.8087	0.6466	0.3652	0.5524	0.4281
Elastic-30-8	0.5913	0.7952	0.6623	0.3860	0.5430	0.4388
Elastic-30-9	0.5657	0.8067	0.6494	0.3690	0.5508	0.4300
Elastic-40-0	0.2905	0.6895	0.3962	0.1881	0.4562	0.2581
Elastic-40-1.5	0.3086	0.9071	0.4508	0.2112	0.6319	0.3106
Elastic-40-2.5	0.3710	0.8862	0.5110	0.2484	0.6135	0.3460
Elastic-40-4	0.4200	0.8675	0.5534	0.2777	0.5979	0.3710
Elastic-40-5	0.4895	0.8416	0.6054	0.3200	0.5770	0.4020
Elastic-40-6	0.4469	0.8591	0.5740	0.2942	0.5909	0.3834
Elastic-40-7	0.4980	0.8383	0.6106	0.3254	0.5752	0.4054
Elastic-40-8	0.5327	0.8242	0.6321	0.3471	0.5639	0.4187
Elastic-40-9	0.5052	0.8359	0.6152	0.3300	0.5733	0.4083
Elastic-50-0	0.2719	0.7006	0.3803	0.1776	0.4651	0.2496
Elastic-50-1.5	0.2769	0.9204	0.4168	0.1925	0.6458	0.2911
Elastic-50-2.5	0.3379	0.9003	0.4805	0.2287	0.6261	0.3282
Elastic-50-4	0.3791	0.8845	0.5192	0.2527	0.6118	0.3504
Elastic-50-5	0.4420	0.8615	0.5718	0.2904	0.5926	0.3813
Elastic-50-6	0.4065	0.8757	0.5425	0.2694	0.6042	0.3643

Systems	Precision	Recall	F1	LCA-P[a]	LCA-R[b]	LCA-F[c]
Elastic-50-7	0.4513	0.8578	0.5783	0.2961	0.5901	0.3854
Elastic-50-8	0.4876	0.8445	0.6043	0.3184	0.5793	0.4010
Elastic-50-9	0.4594	0.8555	0.5841	0.3012	0.5883	0.3890

[a]LCA-P: lowest common ancestor Precision.

[b]LCA-R: lowest common ancestor Recall.

[c]LCA-F: lowest common ancestor F-measure.

Table 7. Results on the biomedical semantic indexing and question answering 2016 test datasets (exploiting the Medical Subject Headings hierarchy.

Test	Precision	Recall	F1	LCA-P[a]	LCA-R[b]	LCA-F1[c]
Batch1						
Week1	0.705	0.619	0.635	0.470	0.389	0.406
Week2	0.717	0.627	0.646	0.476	0.397	0.413
Week3	0.701	0.625	0.635	0.467	0.395	0.407
Week4	0.725	0.613	0.643	0.486	0.385	0.410
Week5	0.707	0.624	0.638	0.474	0.398	0.410
Batch2						
Week1	0.695	0.633	0.637	0.457	0.398	0.405
Week2	0.713	0.637	0.649	0.467	0.410	0.412
Week3	0.691	0.637	0.673	0.464	0.402	0.410
Week4	0.676	0.659	0.641	0.446	0.420	0.4120
Week5	0.686	0.660	0.648	0.448	0.414	0.409
Batch3						
Week1	0.701	0.625	0.639	0.461	0.403	0.410
Week2	0.698	0.652	0.648	0.457	0.408	0.407
Week3	0.694	0.641	0.641	0.447	0.406	0.405
Week4	0.429	0.513	0.399	0.284	0.264	0.258
Week5	0.674	0.660	0.640	0.447	0.419	0.409

[a]LCA-P: lowest common ancestor Precision.

[b]LCA-R: lowest common ancestor Recall.

[c]LCA-F: lowest common ancestor F-measure.

Table 8. Results on the biomedical semantic indexing and question answering 2016 test datasets (without exploiting the Medical Subject Headings hierarchy).

Test	Precision	Recall	F1	LCA-P[a]	LCA-R[b]	LCA-F1[c]
Batch1						
Week1	0.665	0.753	0.687	0.438	0.503	0.452
Week2	0.674	0.767	0.700	0.441	0.513	0.460
Week3	0.661	0.755	0.684	0.437	0.509	0.453
Week4	0.683	0.749	0.697	0.451	0.502	0.460
Week5	0.667	0.757	0.690	0.438	0.509	0.455
Batch2						
Week1	0.655	0.755	0.681	0.427	0.501	0.445
Week2	0.669	0.758	0.692	0.427	0.508	0.454
Week3	0.653	0.757	0.681	0.433	0.509	0.452
Week4	0.639	0.764	0.674	0.420	0.516	0.445
Week5	0.643	0.797	0.692	0.417	0.531	0.451
Batch3						
Week1	0.666	0.746	0.684	0.437	0.512	0456
Week2	0.654	0.774	0.690	0.421	0.517	0.448
Week3	0.655	0.754	0.680	0.426	0.507	0.446
Week4	0.390	0.475	0.410	0.254	0.311	0.268
Week5	0.663	0.770	0.672	0.416	0.516	0.442

[a]LCA-P: lowest common ancestor Precision.

[b]LCA-R: lowest common ancestor Recall.

[c]LCA-F: lowest common ancestor F-measure.

Table 9. Baseline results provided by the Medical Text Indexer (MTI) tool. These results were taken from the biomedical semantic indexing and question answering website.

Test	Precision	Recall	F1	LCA-P[a]	LCA-R[b]	LCA-F1[c]
Batch1						
Week1	0.558	0.516	0.493	0.498	0.462	0.463
Week2	0.550	0.514	0.487	0.516	0.478	0.480
Week3	0.553	0.537	0.507	0.499	0.467	0.465
Week4	0.568	0.505	0.482	0.507	0.455	0.464
Week5	0.558	0.508	0.484	0.504	0.474	0.473
Batch2						
Week1	0.546	0.520	0.493	0.495	0.473	0.467
Week2	0.544	0.520	0.492	0.497	0.471	0.469
Week3	0.558	0.526	0.500	0.503	0.470	0.470
Week4	0.549	0.516	0.491	0.487	0.452	0.449
Week5	0.532	0.551	0.519	0.480	0.487	0.467
Batch3						
Week1	0.515	0.459	0.444	0.492	0.441	0.449
Week2	0.543	0.484	0.466	0.493	0.455	0.455
Week3	0.580	0.502	0.486	0.512	0.457	0.466
Week4	0.545	0.522	0.494	0.496	0.481	0.469
Week5	0.536	0.517	0.496	0.499	0.473	0.466

[a]LCA-P: lowest common ancestor Precision.

[b]LCA-R: lowest common ancestor Recall.

[c]LCA-F: lowest common ancestor F-measure.

Table 10. Results of the top systems in biomedical semantic indexing and question answering (BioASQ) task 4a. These scores were taken on December 5 from the BioASQ website.

Batch	System	Week	Number of annotated articles	Total of articles	Precision	Recall	F1
1	MeSHLabeler	1	1853	3740	0.626	0.521	0.513
	MeSHLabeler	2	1578	2872	0.625	0.515	0.506
	MeSHLabeler	3	1115	2599	0.602	0.519	0.515
	MeSHLabeler-1	4	1436	3294	0.649	0.496	0.495
	MTI	5	1181	3210	0.558	0.508	0.484
2	MTI	1	1080	3212	0.546	0.520	0.493
	MeSHLabeler-2	2	901	3213	0.630	0.505	0.499
	MeSHLabeler-2	3	850	2831	0.642	0.521	0.516
	MTI	4	800	3111	0.549	0.516	0.491
	MeSHLabeler	5	688	2470	0.615	0.538	0.526
3	MeSHLabeler	1	305	2994	0.637	0.462	0.462
	MeSHLabeler	2	507	3044	0.6449	0.4851	0.4825
	MeSHLabeler	3	501	3351	0.6544	0.4991	0.4966
	MeSHLabeler	4	514	2630	0.6312	0.5098	0.5012
	MeSHLabeler	5	627	3130	0.5017	0.5119	0.6135

Discussion

Principal Findings

Our approach relies on the assumption that similar documents should be classified by similar MeSH terms. Previous works have already applied a k-NN approach for obtaining the set of similar document for a given test document. Our previous work [24] and this study are the first efforts to explore the document similarity using the search engine ElasticSearch instead of k-NN. ElasticSearch is one of the most efficient document-based database. Given a test document, this is represented as a query, which is executed in the search engine, returning the documents more relevant (similar) to the query. Then, our system proposes the MeSH of all these documents as the initial set of MeSH terms for the test document and extends this set with the MeSH terms proposed by the MTI tool. Finally, the system uses a scoring function to determine the best set of MeSH terms for a given article. Those MeSH terms that achieve a higher score than a given threshold are finally selected. The experiments show that the best results are obtained when the number of retrieved relevant documents by ElasticSearch is small (10) and the threshold for the scoring function is equal to 5.

Comparison With Prior Work

Our approach seems to provide better results than the top systems in BioASQ 2016. We note that our results are not immediately comparable with those reported by the BioASQ challenge because we have used a different test dataset. However, we think that it is a reasonable evaluation while no official test datasets are available. Moreover, our development test datasets are available at our webpage [35] to facilitate reproducible research, objective assessment, and further analysis.

In addition, we implement one of the guidelines established by human curators to classify MEDLINE abstracts. To do this, we store the MeSH thesaurus into a graph-based database by using the BlazeGraph tool. The main advantage of using a graph structure is the possibility to use algorithms well known in graph theory (such as depth-first search) to extract subgraphs satisfying a given query. In particular, the graph is visited with the objective to determine whether 3 or more MeSH terms assigned to a given article share the same ancestor. In this case, this lowest common ancestor should substitute them. Contrary to expectations, the system produces worse results if this rule is applied. This may be because human curators do not always follow the recommendations to catalog MEDLINE abstracts.

Limitations

Although the results are better when we do no exploit the hierarchy of MeSH, we think that the graph database version of MeSH is a promising resource that will allow us to implement other guidelines or strategies to select the most appropriate MeSH terms for representing a given article.

Conclusions

Semantic indexing of MEDLINE articles is a manual, laborious task, which could be helped by information technology.

As future steps, we also plan to determine semantic similarity between documents using word embeddings [36] instead of the well-known and commonly used VSM for information retrieval. This approach has already been exploited by Liu et al [21] and Kosmopoulos et al [22]. Unlike these works, based on the use of k-NN for obtaining the set of similar documents, our approach will continue using ElasticSearch as search engine and our graph database format of MeSH. We also plan to explore deep learning methods (such as Convolutional Neural Networks) for supporting the automatic classification of MEDLINE abstracts.

Acknowledgments

This work was supported by eGovernAbility-Access project (TIN2014-52665-C2-2-R).

Authors' Contributions

All three authors designed the study. AC developed the system and performed the experiments. This document was prepared by ISB and PM.

Conflicts of Interest

None declared.

References

1. Nlm.nih. Use of medical subject headings for cataloging URL: https://www.nlm.nih.gov/mesh/catpractices.html [accessed 2017-11-11] [WebCite Cache ID 6mPBxudvZ]
2. Huang M, Névéol A, Lu Z. Recommending MeSH terms for annotating biomedical articles. J Am Med Inform Assoc 2011;18(5):660-667 [FREE Full text] [doi: 10.1136/amiajnl-2010-000055] [Medline: 21613640]
3. Krallinger M, Leitner F, Valencia A. The BioCreative II.5 challenge overview. 2009 Presented at: Proceedings of the BioCreative II.5 Workshop on Digital Annotations; October 7-9, 2009; Madrid, Spain URL: http://www.biocreative.org/media/store/files/2009/Proceedings.pdf
4. Kim JD, Pyysalo S, Ohta T, Bossy R, Nguyen N, Tsujii J. Overview of BioNLP shared task 2011. 2011 Presented at: Proceedings of the BioNLP Shared Task Workshop; June 24, 2011; Portland, Oregon, USA p. 1-6 URL: http://www.aclweb.org/anthology/W11-1801

5. Nédellec C, Bossy R, Kim JD, Ohta T, Pyysalo S, Zweigenbaum P. Overview of BioNLP shared task 2013. 2013 Presented at: Proceedings of the BioNLP Shared Task 2013 Workshop; August 9, 2013; Sofia, Bulgaria p. 1-7 URL: https://aclweb.org/anthology/W/W13/W13-2000.pdf
6. Stubbs A, Kotfila C, Xu H, Uzuner Ö. Identifying risk factors for heart disease over time: overview of 2014 i2b2/UTHealth shared task Track 2. J Biomed Inform 2015 Dec;58 Suppl:S67-S77 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.001] [Medline: 26210362]
7. Segura-Bedmar I, Martínez P, Sánchez-Cisneros D. The 1st DDIExtraction-2011 Challenge Task: Extraction of Drug-Drug Interactions from Biomedical Texts. In: Proceedings of the 1st Challenge task on Drug-Drug Interaction Extraction. 2011 Presented at: First Challenge Task on Drug-Drug Interaction Extraction 2011; September 2011; Huelva, Spain p. 1-9.
8. Segura-Bedmar I, Martínez P, Herrero-Zazo M. SemEval-2013 Task 9 : Extraction of Drug-Drug Interactions from Biomedical Texts (DDIExtraction 2013). Atlanta, Georgia: Association for Computational Linguistics; 2013 Presented at: Proceedings of Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013); June 14-15, 2013; Atlanta, Georgia, USA p. 341-350 URL: https://pdfs.semanticscholar.org/a9a5/b40e179ed22e3d59b7823bffc9d3aaa65474.pdf
9. Ruiz ME, Srinivasan P. Hierarchical text categorization using neural networks. Inf Retr 2002;5:87-118. [doi: 10.1023/A:1012782908347]
10. Jimeno-Yepes AJ, Plaza L, Carrillo-de-Albornoz J, Mork JG, Aronson AR. Feature engineering for MEDLINE citation categorization with MeSH. BMC Bioinformatics 2015 Apr 8;16:113 [FREE Full text] [doi: 10.1186/s12859-015-0539-7] [Medline: 25887792]
11. Mao Y, Wei CH, Lu Z. NCBI at the 2014 BioASQ Challenge Task: Large-scale Biomedical Semantic Indexing and Question Answering. 2014 Presented at: Proceedings of Conference and Labs of the Evaluation Forum (CLEF) 2014 (Working Notes); 2014; Sheffield, UK. [doi: 10.13140/2.1.4288.6405]
12. Papanikolaou Y, Tsoumakas G, Laliotis M, Markantonatos N, Vlahavas IP. AUTH-Atypon at BioASQ 3: Large-Scale Semantic Indexing in Biomedicine. 2015 Presented at: Proceedings Conference and Labs of the Evaluation Forum (CLEF) 2015 (Working Notes); 2015; Toulouse, France URL: http://ceur-ws.org/Vol-1391/29-CR.pdf
13. Liu K, Wu J, Peng S, Zhai C, Zhu S. The FUDAN-UIUC participation in the BioASQ challenge Task 2a: The antinomyra system. 2014 Presented at: Proceedings of Conference and Labs of the Evaluation Forum (CLEF)2014; September 15-18, 2015; Sheffield, UK URL: https://illinois-staging.pure.elsevier.com/en/publications/the-fudan-uiuc-participation-in-the-bioasq-challenge-task-2a-the-
14. Mork JG, Jimeno-Yepes AJ, Aronson AR. The NLM Medical Text Indexer System for Indexing Biomedical Literature. 2013 Presented at: Proceedings of the first Workshop on Bio-Medical Semantic Indexing and Question Answering, a Post-Conference Workshop of Conference and Labs of the Evaluation Forum 2013 (CLEF 2013); September 27, 2013; Valencia, Spain URL: https://ii.nlm.nih.gov/Publications/Papers/MTI_System_Description_Expanded_2013_Accessible.pdf
15. Tsoumakas G, Laliotis M, Markantonatos N, Vlahavas I. Large-Scale Semantic Indexing of Biomedical Publications at BioASQ. 2013 Presented at: Proceedings of the First Workshop on Bio-Medical Semantic Indexing Question Answering, a Post-Conference Workshop of Conference Labs of the Evaluation Forum (CLEF); 2013; Valencia, Spain URL: http://ceur-ws.org/Vol-1094/bioasq2013_submission_6.pdf
16. Madani O, Huang J. Large-scale many-class prediction via flat techniques. 2010 Presented at: Proceedings of Large-Scale Hierarchical Classification Workshop; March 28, 2010; Milton Keynes, UK URL: http://lshtc.iit.demokritos.gr/system/files/OmidMadani.pdf
17. Cissé M, Artieres T, Gallinari P. Learning efficient error correcting output codes for large hierarchical multi-class problems. 2011 Presented at: Proceedings of Workshop on Large-Scale Hierarchical Classification ECML/PKDD; September 5, 2011; Athens, Greece URL: http://lib.iit.demokritos.gr/system/files/CisseMouhamadou.pdf
18. Ribadas FJ, de Campos LM, Bilbao VM, Romero AE. Two hierarchical text categorization approaches for BioASQ semantic indexing challenge. 2013 Presented at: Proceedings of the first Workshop on Bio-Medical Semantic Indexing Question Answering, a Post-Conference Workshop of Conference and Labs of the Evaluation Forum (CLEF); September 2013; Valencia, Spain URL: http://bioasq.org/sites/default/files/workshop1/Ribadas_presentation.pdf
19. Xue G, Xing D, Yang Q, Yu Y. Deep classification in large-scale text hierarchies. 2008 Presented at: Proceedings of the 31st Annual international ACM SIGIR Conference on Research and Development in Information Retrieval; July 20-24, 2008; Singapore p. 619-626 URL: http://www.cs.ust.hk/~qyang/Docs/2008/fp350-xue.pdf
20. Liu TY, Yang Y, Wan H, Zeng HJ, Chen Z, Ma WY. Support vector machines classification with a very large-scale taxonomy. SIGKDD Explor 2005;7(1):36-43 [FREE Full text]
21. Kosmopoulos A, Androutsopoulos I, Paliouras G. Biomedical semantic indexing using dense word vectors in BioASQ. J BioMed Semant Suppl BioMedl Inf Retr 2015:3410 [FREE Full text]
22. Peng S, You R, Wang H, Zhai C, Mamitsuka H, Zhu S. DeepMeSH: deep semantic representation for improving large-scale MeSH indexing. Bioinformatics 2016 Jun 15;32(12):i70-i79. [Medline: 27307646]

23. Tsatsaronis G, Balikas G, Malakasiotis P, Partalas I, Zschunke M, Alvers MR, et al. An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition. BMC Bioinformatics 2015 Apr 30;16:138 [FREE Full text] [doi: 10.1186/s12859-015-0564-6] [Medline: 25925131]
24. Segura-Bedmar I, Carruana A, Martínez P. LABDA at the 2016 BioASQ challenge task 4a: Semantic Indexing by using ElasticSearch. 2016 Presented at: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics; August 7-12, 2016; Berlin, Germany p. 16-22 URL: http://www.aclweb.org/anthology/W16-3103
25. Trieschnigg D, Pezik P, Lee V, de Jong F, Kraaij W, Rebholz-Schuhmann D. MeSH Up: effective MeSH text classification for improved document retrieval. Bioinformatics 2009 Jun 1;25(11):1412-1418 [FREE Full text] [doi: 10.1093/bioinformatics/btp249] [Medline: 19376821]
26. Elastic. Hosted Elasticsearch & Kibana from the Source URL: https://www.elastic.co/ [accessed 2016-11-30] [WebCite Cache ID 6mPB5LREO]
27. Abramova V, Bernardino J, Furtado P. Experimental evaluation of NoSQL databases. Int J Database Manage Syst 2014 Jun;6(3):1-16. [doi: 10.5121/ijdms.2014.6301]
28. Lemurproject. INDRI language modeling meets inference networks URL: http://www.lemurproject.org/indri/ [accessed 2016-12-01] [WebCite Cache ID 6mQihXhgh]
29. Xiaonan J, Ritter A, Yen PY. Using ontology-based semantic similarity to facilitate the article screening process for systematic reviews. J Biomed Inform 2017 May;69:33-42. [Medline: 28302519]
30. Gu J, Qian L, Zhou G. Chemical-induced disease relation extraction with various linguistic features. Database (Oxford) 2016 May 6;2016:pii: baw042. [Medline: 27052618]
31. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
32. Blazegraph. Blazegraph feature overview (Full Feature Matrix) URL: https://www.blazegraph.com/ [accessed 2016-11-30] [WebCite Cache ID 6mPBceVUG]
33. Kosmopoulos A, Partalas I, Gaussier E, Paliouras G, Androutsopoulos I. Evaluation measures for hierarchical classification: a unified view and novel approaches. Data Min Knowl Discov 2015;29(3):820-865. [doi: 10.1007/s10618-014-0382-x]
34. Kosmopoulos A, Partalas I, Gaussier E, Paliouras G, Androutsopoulos I. Bioasq. URL: http://bioasq.org/resources/software/HEMKit.zip [accessed 2016-11-30] [WebCite Cache ID 6mPA2z3GW]
35. Segura-Bedmar I, Martínez P. Labda.inf. Laboratorio de Bases de Datos Avanzadas URL: http://labda.inf.uc3m.es [accessed 2017-11-11] [WebCite Cache ID 6mPA2z3GW]
36. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems; December 5-10, 2013; Nevada, USA p. 3111-3119.

Abbreviations

BioASQ: biomedical semantic indexing and question answering challenge
BoW: bag-of-words
F1: F-measure
k-NN: k-nearest neighbors
LCA: lowest common ancestor
LCA-P: lowest common ancestor Precision
LCA-R: lowest common ancestor Recall
LCA-F: lowest common ancestor F-measure
MeSH: Medical Subject Headings
MTI: Medical Text Indexer
NCBI: National Center for Biotechnology Information
NLM: National Library of Medicine
NLP: Natural Language Processing
PMID: PubMed unique identifier
RDF: Resource Description Framework.
SVM: Support Vector Machine
UMLS: Unified Medical Language System
VSM: vector space model

Edited by G Eysenbach; submitted 12.06.17; peer-reviewed by M Neves, N Alvaro, A Abbe; comments to author 21.07.17; revised version received 08.09.17; accepted 27.09.17; published 01.12.17.

Please cite as:
Segura Bedmar I, Martínez P, Carruana Martín A
Search and Graph Database Technologies for Biomedical Semantic Indexing: Experimental Analysis
JMIR Med Inform 2017;5(4):e48
URL: http://medinform.jmir.org/2017/4/e48/
doi:10.2196/medinform.7059
PMID:29196280

Original Paper

Predicting Unplanned Transfers to the Intensive Care Unit: A Machine Learning Approach Leveraging Diverse Clinical Elements

Ben Wellner[1], MSc, PhD; Joan Grand[1], RN; Elizabeth Canzone[1], PhD; Matt Coarr[1], MSc; Patrick W Brady[2], MSc, MD; Jeffrey Simmons[2], MSc, MD; Eric Kirkendall[2], MBI, MD, FAAP; Nathan Dean[3], MD; Monica Kleinman[4], MD; Peter Sylvester[1], BS

[1]The MITRE Corporation, Bedford, MA, United States

[2]Cincinnati Children's Hospital, Cincinnati, OH, United States

[3]Children's National Health System, Washington, DC, United States

[4]Boston Children's Hospital, Boston, MA, United States

Corresponding Author:
Ben Wellner, MSc, PhD
The MITRE Corporation
202 Burlington Rd
Bedford, MA, 01730
United States
Phone: +1 781 271 7191
Fax: +1 781 271 2352
Email: wellner@mitre.org

Abstract

Background: Early warning scores aid in the detection of pediatric clinical deteriorations but include limited data inputs, rarely include data trends over time, and have limited validation.

Objective: Machine learning methods that make use of large numbers of predictor variables are now commonplace. This work examines how different types of predictor variables derived from the electronic health record affect the performance of predicting unplanned transfers to the intensive care unit (ICU) at three large children's hospitals.

Methods: We trained separate models with data from three different institutions from 2011 through 2013 and evaluated models with 2014 data. Cases consisted of patients who transferred from the floor to the ICU and met one or more of 5 different priori defined criteria for suspected unplanned transfers. Controls were patients who were never transferred to the ICU. Predictor variables for the models were derived from vitals, labs, acuity scores, and nursing assessments. Classification models consisted of L1 and L2 regularized logistic regression and neural network models. We evaluated model performance over prediction horizons ranging from 1 to 16 hours.

Results: Across the three institutions, the c-statistic values for our best models were 0.892 (95% CI 0.875-0.904), 0.902 (95% CI 0.880-0.923), and 0.899 (95% CI 0.879-0.919) for the task of identifying unplanned ICU transfer 6 hours before its occurrence and achieved 0.871 (95% CI 0.855-0.888), 0.872 (95% CI 0.850-0.895), and 0.850 (95% CI 0.825-0.875) for a prediction horizon of 16 hours. For our first model at 80% sensitivity, this resulted in a specificity of 80.5% (95% CI 77.4-83.7) and a positive predictive value of 5.2% (95% CI 4.5-6.2).

Conclusions: Feature-rich models with many predictor variables allow for patient deterioration to be predicted accurately, even up to 16 hours in advance.

(JMIR Med Inform 2017;5(4):e45) doi:10.2196/medinform.8680

KEYWORDS

clinical deterioration; machine learning; data mining; electronic health record; patient acuity; vital signs; nursing assessment; clinical laboratory techniques

Introduction

Better prediction of clinical deterioration is a priority as many patients today get harmed when precursors go unrecognized, leading to potentially preventable morbidity, mortality, and cost. Over the last two decades, it has become increasingly clear that precursors to clinical deterioration commonly exist, and rapid response systems that detect and respond to early deterioration can improve outcomes [1-5].

Increased mortality and morbidity is associated with deterioration in patients who require an unplanned transfer from the nursing floor to the ICU (Intensive Care Unit). The mortality rate associated with unrecognized deterioration that results in a delay of ICU transfer and the need for resuscitation can be as high as 67% [6,7]. Missing precursors to deterioration reduces the window of opportunity and margin of error for effective intervention and increases the intensity and complexity of the required care.

Clinical EHR (electronic health record) systems and their rich, heterogeneous data provide opportunities for impactful secondary use [8,9]. Yet fully taking advantage of such large repositories of data is a challenge because of sheer complexity of the data [10]. Machine learning methods offer a promising set of techniques to address such challenges by providing statistically sound data-driven methods able to identify subtle patterns in data while remaining robust to problems in data quality and completeness [11].

Most machine learning methods for predicting deterioration have focused on logistic regression models preceded by careful variable selection [12,13]. Recently, more advanced machine learning approaches including nonlinear and nonparametric methods have been used [14]. These more powerful methods can accommodate larger feature sets and also identify implicit or explicit feature interactions. In many cases, however, model interpretability can suffer [15,16].

The purpose of this study was to develop highly accurate predictive models able to identify unplanned transfers to the ICU at least 6 hours before transfer. Critically, we leverage thousands of predictor variables, rather than dozens as is common in predicting adverse health events. We hypothesized that such complex models provide better accuracy at longer prediction horizons, providing more time and opportunity for clinicians to act to reverse deterioration.

Methods

Research Team

The MITRE Corporation together with three pediatric hospitals, Boston Children's Hospital (BCH), Children's National Health System (CNHS), and Cincinnati Children's Hospital Medical Center (CCHMC) formed a partnership for the purpose of sharing data to uncover issues impacting patient safety. Each hospital contributed EHR data from 2011 to 2014 totaling >1 million patients and >8 million patient encounters, forming >7.2 TB of data across all three hospitals. Clinical data available from the three hospitals using 2 different EHR vendors was used to in our study to predict deterioration.

Case Identification

Cases in our study involved instances of unplanned transfers from an inpatient ward to the ICU. The unit of analysis was the ICU transfer and not the patient, as each patient could experience more than one ICU transfer within the same hospital admission. The case identification proceeded in two phases. First, a set of *candidate cases* were identified from admission-discharge-transfer (ADT) data by selecting patient encounters that involved a stay on the nursing floor followed by a transfer to the ICU. Specifically, the candidate cases included ICU transfers originating from all nursing floors, excluding any transfers from the emergency department (ED), operating room (OR), postanesthesia care unit (PACU) or ICU and excluding any transfers to the neonatal intensive care unit (NICU).

From the candidate cases, we then developed a method for establishing whether a transfer was likely unplanned or not. Ideally, cases would be identified carefully by clinician review, as no variable or flag exists in the EHR to designate an unplanned transfer. To address this challenge, our team, which included clinicians at three hospitals, identified a set of five criteria to establish our case cohorts through objective, heuristic means. Unplanned transfers were identified as transfers to the ICU meeting one or more of the criteria (see Figure 1). This working definition of unplanned ICU transfer is the result of prior work in the literature [17] combined with knowledge gained from each institution's experience.

We further subdivided the list of cases into those patients who experienced a critical deterioration event (CDE) along with an unplanned transfer to the ICU. CDE was defined as an unplanned floor to ICU transfer with invasive or noninvasive positive pressure ventilation, vasopressors, fluid resuscitation, or other emergent procedures 2 hours before and 12 hours post transfer [5]. The prediction model was aimed at predicting unplanned transfers; however, the CDE subgroup was important in understanding the connection between unplanned transfers and critical deteriorations.

Identification of the Control Group

Controls were sampled from the set of patient visits where the patient spent at least 24 hours on an inpatient floor and was never transferred to the ICU. Sampling was done by ensuring that the ratios of ages and diagnoses were similar between the case and control population. Diagnoses were determined by discharge diagnosis according to ICD-9 (International Classification of Diseases-9). This sampling scheme was designed to balance the need for controls to be representative of the inpatient floor population, yet also to ensure that the control population did not differ from the case population in systematic ways. We removed patients from cases and controls that spent less than 8 hours on the floor. Table 1 provides the counts for the cases and controls across the three institutions.

Figure 1. The five criteria involved in determining an "unplanned" intensive care unit (ICU) transfer. CPAP: continuous positive airway pressure; BiPAP: bilevel positive airway pressure; NS: normal saline; LR: lactated ringer; MAR: medication administration record.

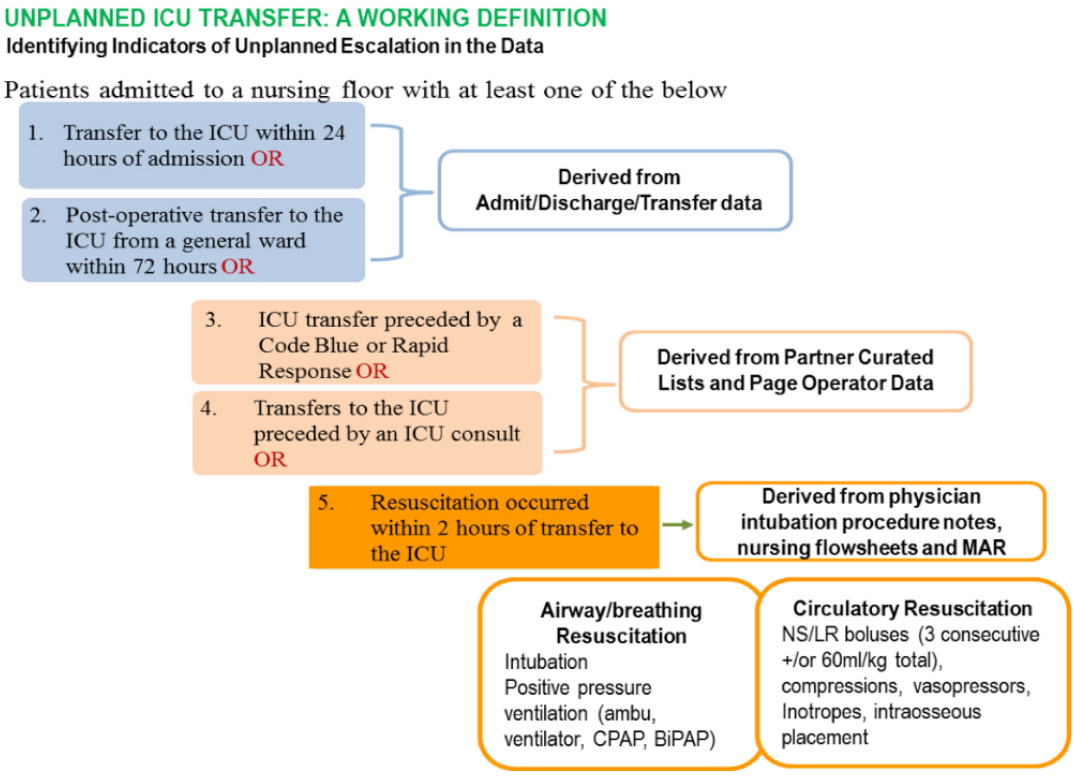


Table 1. Counts for cases and controls across three institutions.

Dataset	BCH[a]	CCHMC[b]	CNHS[c]
Training			
Cases	1163	1090	546
Controls	6448	6170	3893
Evaluation			
Cases	326	478	324
Controls	1878	1353	1339

[a]Boston Children's Hosptial.

[b]Cincinnati Children's Hospital and Medical Center.

[c]Children's National Health System.

Clinical Element and Feature Extraction

Data preparation involved two primary stages before creating data instances for training and evaluating predictive models. The first stage involved pulling clinical element data out of underlying vendor database tables with a complex schema into a simplified set of database tables through a set of Structured Query Language (SQL) queries. The clinical element categories included vitals, laboratory results, acuity scores (eg, existing early warning score or nurse acuity calculations) and nursing assessments. An overview of the clinical elements used in our study, specific to data from Cincinnati Children's are summarized in Table 2. Clinical elements based on patients' vitals were standardized across the three hospitals. The other types of clinical elements, especially acuity and nursing assessments differed across the institutions because of different EHR systems and/or different customizations made by each institution. No attempt was made to standardize such elements. Although laboratory results would have been possible to harmonize across the institutions, acuity scores did not map from one institution to another. Nursing assessments provided even more variability; besides a lack of a one-to-one mapping between institutions, nursing assessments sometimes used values chosen from a fixed set (in a drop-down menu) and in other cases allowed for free text.

Table 2. Summary of clinical elements.

Clinical category	Clinical elements
Vitals	Temperature
	Heart rate
	Respiratory rate
	Systolic blood pressure
	Oxygen saturation
Laboratory results	Sodium
	Potassium
	Glucose
	Creatinine
	Bicarbonate
	White blood cell count
	Hermatocrit
	Hemoglobin
Acuity scores	PEWS[a] total score
	Total acuity score
	Acuity level
Nursing assessments	Braden risk
	Activity
	Adult Glasgow coma score
	Audible sounds w/o stethoscope
	Best verbal response
	Brachial bilateral pulse
	Brachial left pulse
	Brachial right pulse
	Cardiac
	Cardiovascular
	Central perfusion cap refill
	Cough
	Eye opening
	Faces pain classification
	Faces pain score
	Femoral bilateral pulse
	Femoral left pulse
	Femoral right pulse
	FLAAC[b] activity
	FLAAC consolability
	FLAAC cry/face/legs
	FLAAC pain classification
	FLAAC total pain score
	Fluid balance
	Friction sheer

Clinical category	Clinical elements
	Heart rate/rhythm
	Patient experiencing pain?
	Level of consciousness
	Left lower extremity perfusion cap
	Left upper extremity perfusion cap refill
	Minimum stimulus to invoke response
	Mobility
	Moisture
	Neurological
	Neurovascular check
	NRS[c] pain classification
	Nutrition
	Orientation level
	Orientation
	Ped Glasgow coma score
	Perfusion cap refill
	Perfusion color
	Perfusion skin temperature
	Peripheral pulses
	PERRLA[d]
	Pupil reaction
	Respirations/respiratory
	Respiratory status
	Response to stimuli
	Retractions
	Rhythm
	Right lower extremity perfusion cap
	Right upper extremity perfusion cap refill
	Secretion/sputum color
	Skin within normal limits
	Temperature condition
	Total pain score for site
	Upper perfusion cap refill
	Work of breathing

[a]Pediatric Early Warning Score.

[b]Faces, Legs, Activity, Cry, Consolability.

[c]Numeric Rating Scale.

[d]Pupils, Equal, Round, React to Light, Accomodation.

Table 3. Feature types used to construct features from clinical elements.

Feature type description	Feature examples
Vitals	
Linear regression slope over scalar vitals of given type	HR[a] slope=−2.1
Magnitude of linear regression slope	HR magnitude slope=2.1
Sign of slope of linear regression	HR slope is negative
Binned category C1-C4 of MAXIMUM value	Maximum HR is C4
Binned category C1-C4 of MINIMUM value	Minimum HR is C1
Binned category C1-C4 of AVERAGE value	Average HR is C3
Binned category C1-C4 of NEWEST (most recent) value	Newest HR is C4
Binned category C1-C4 of OLDEST (least recent) value	Oldest HR is C1
Normalized histogram values over categories computed by counting category assignments for each measurement and normalizing to 1	HR C1 Histogram=0.3; HR C2 Histogram=0.4; HR C4 Histogram=0.3
Labs	
Category (Low, Normal, High) pairs: 2nd Newest and Newest	Glucose low≥normal
Change or lack of change in category (Low, Normal, High) from 2nd Newest to Newest	Creatinine high≥high
Binned percentage change in value from Oldest to Newest value	WBC[b] new/old>1.5
Newest/Oldest>{1.25, 1.5, 2.0, 3,0}	
Newest/Oldest<{0.8, 0.67, 0.5, 0.33}	
Binned percentage change in value from 2nd Newest to Newest value	Glucose new/2nd<0.5
Newest/Oldest>{1.25, 1.5, 2.0, 3,0}	
Newest/Oldest<{0.8, 0.67, 0.5, 0.33}	
Attribute type is present in prediction window 1 or more times	Glucose is present; WBC is present
Last category (Low, Normal, High)	Last WBC is high
Acuity	
Score MINIMUM is 0 or value>{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}	Minimum PEWS[c] score is 0
Score MAXIMUM is 0 or value>{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}	Maximum PEWS score>0
NEWEST score is 0 or value>{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}	Newest PEWS score>0
Linear regression slope over scores	Slope PEWS score=-1.5
Linear regression slope over scores over last 6 hours	Slope PEWS score last 6 hours=3.1
Magnitude of slope over scores	Magnitude slope PEWS score=1.5
Magnitude of slope over scores over last 6 hours	Magnitude slope PEWS score=1.5
Number of measurements of type over last 6 hours>{0, 1,2,4,6,10} (multiple overlapping features included)	Number of PEWS Measurements >0; Number of PEWS Measurements >1; Number of PEWS Measurements >2
Assessments	
Nursing assessment attribute value pair (whether value is scalar or a string)	Cough is productive; Mobility is 1; Mobility is 2
NEWEST assessment attribute-value pair for given attribute	Newest mobility is 2

[a]Heart rate.

[b]White blood cell count.

[c]Pediatric early warning score.

Clinical elements present within the prediction window for each clinical element type were used for feature extraction. The prediction window for vitals included the time frame of 24 hours leading up to the prediction time and for all other elements the length of the prediction window was 72 hours. From raw clinical elements, extracted features aim to capture the state of the patient and patient trajectory. Many of our features using vitals follow the approach taken by Zhai et al [12]. For example, vitals

were binned into risk categories C1 to C4. Features are then derived from these categorized/binned vitals. The types of features derived from the various clinical elements are summarized in Table 3**.** Noteworthy is that we made no attempt to impute missing values.

Machine Learning Methods

Our experiments used logistic regression models and a nonlinear extension to logistic regression in the form of multilayer perceptrons (MLPs), also known as feed-forward neural networks. Neural networks have seen a resurgence in recent years with improved techniques to train them efficiently and effectively.

For binary classification, logistic regression can be written as:

$$p(y=1|x) = logistic(x) = 1/ (1 + exp(-wx)) (1)$$

where *w* is the set of weights (or coefficients) in the model and x represents a vector of input variables, that is, features. Hidden layers consist of sets of neurons; each layer can be viewed as successive (nonlinear) transformations of the input, each having the form:

$$H_i(z) = g(W_i(z)) (2)$$

Where *z* is the input vector to layer *i*, *g* is an activation function and W_i is a matrix of weights. In our models here, we use a rectified linear activation function of the form $g(x)=\max(0, x)$. Given this form, a MLP with *n* hidden layers can be written as:

$$p (y=1| x) = logistic (H_n(H_{n-1}(\ldots(H_1(x))))) (3)$$

As with logistic regression, the model is fit by maximizing the likelihood of the training data . However, given the large number of parameters in our models caused by so many features, there is a strong tendency to *overfit* the training data leading to poor generalization on unseen data. Accordingly, we heavily regularize our models using L1 and L2 regularization terms [18], their joint use sometimes referred to as elastic net regularization. L1 regularization is especially useful as it implicitly performs feature selection. This is beneficial in our case with potentially many irrelevant features [19]. MLPs are even more prone to overfitting as they include more parameters and capture complex nonlinear interactions between the inputs. Our experiments using MLPs make use of dropout [20], a technique in which a certain percentage of the neurons are randomly elided upon processing each data point during training.

Regularization can be achieved by adding penalty terms to the likelihood based on the L1 and L2 norms of the model weights. The penalized log-likelihood has the form:

$$L(D, W) = \sum_{i=1}[log\ p\ (y= y^{(i)}| x^{(i)})] - a_1|| W ||_1 - a_2|| W ||_2 (4)$$

where *W* refers to *all* the weights in the model (including any hidden layer weights) and where a_1 and a_2 are "hyper"-parameters that determine the "strength" of the two regularizer components: $|| W ||_1$ denoting the L1 norm of the parameters and $|| W ||_2$ the L2 norm. These regularizers penalize large-magnitude weights and prevent the model from fitting the training data too closely at the expense of its ability to generalize. Modern machine learning techniques rely heavily on regularization to develop accurate prediction models with large numbers of features and modest amounts of training data.

Estimating the parameters for all models (logistic regression and MLPs) was done by maximizing the penalized likelihood with stochastic gradient descent [21,22]. All machine learning models were trained and used for prediction with the Mandolin machine learning toolkit available as open source on Github.

Model Preparation

The training data used to construct our models leveraged patient encounters from January 1, 2011 through December 31, 2013. Separate models were trained for each institution because clinical elements are not standardized across EHR systems. Model settings such as the regularization coefficients, the number of hidden layers for MLP models, and the number of training iterations were tuned using 5-fold cross validation on the training set. Given the low prevalence of unplanned transfers, we subsampled the controls so that our training data had roughly a 1:5 ratio of cases to controls. We measured the area under the receiver operating characteristic (ROC) curve, the specificity at the threshold corresponding to 80% sensitivity, and also computed the estimated positive predictive value (PPV) given the overall 1.3% prevalence in our dataset. The estimated PPV was derived from the sensitivity, specificity, and prevalence [23].

Experimental Design

We carried out three sets of experiments across all three institutions to measure the contributions of four different clinical element types. The first set of experiments looked at the performance of predictive models using only clinical elements of a single type. A second set of experiments looked at performance when features from each the clinical element types were added successively, in the order: *vitals*, *lab results*, *acuity scores*, and *nursing assessments*. Finally, we carried out a set of ablation experiments comparing the full model, making use of all features with feature sets constructed by removing features for each clinical element type separately. These experiments were carried out with regularized logistic regression.

A key concern in the practical use of a predictive model for detecting patient deterioration is how sensitive the model might be to varying lengths of time between when a prediction is made and when a patient is transferred to the ICU or prediction horizons. For controls, the prediction horizon is the time between when the prediction is made and the patient leaves the floor.

We provided results on experiments training the model to predict deterioration at prediction horizons varying from 1 hour to 16 hours, at 1-hour intervals. Evaluation on the test set was done using the same prediction horizon as was used to train the model.

We examined how well models with different feature sets performed across different prediction horizons.

This set of experiments examined how well models trained to identify deterioration with a given prediction horizon performed when evaluated across different prediction horizons. For example, how a model fared when asked to predict deterioration 16 hours in advance if it was trained to identify deterioration with just a 2-hour prediction horizon. Conversely, how might

a model predict risk of deterioration for a patient just 2 hours away from an unplanned transfer if trained to identify deterioration 16 hours in advance.

Finally, experiments were carried out to measure the effect of regularization on logistic regression models, reducing the number of input features by feature selection and also provide more detailed comparison of MLP models versus logistic regression.

All models were binary classifiers designed to predict whether a patient will have an unplanned transfer to the ICU or not. Evaluation is carried out on the test data from 2014. Our primary evaluation metric is the area under the ROC curve. We also considered the models' specificity at 80% sensitivity and examined the PPV at this cut-point, assuming a prevalence of 1.3% which matched the prevalence of deterioration across the three institutions.

Ethics Approval

The study was reviewed and approved by the institutional review boards at Boston Children's Hospital, Children's National Medical Center, and Cincinnati Children's Hospital Medical Center.

Results

Clinical Element Analysis

Our experiments followed the case and control selection methodologies described, and subsampled the controls. The total case and control counts are shown in Table 1. Primary results are detailed in Table 4. The feature configurations prefixed with All- *X* involved using all features except for those of type *X*. Interestingly, removing any single feature type from all available features generally resulted in minor, nonstatistically significant, reductions in the area under the ROC curve. This held except for the case of removing nursing assessments, which resulted in statistically significant degradations for CCHMC and CNHC but not BCH.

The prediction horizon used for training models and evaluating them was 6 hours. The models used all available features for all experiments and all models were regularized logistic regression except for the rows with MLP denoting a multilayer perceptron model. All logistic regression models across all features sets and institutions used a_1=0.001 and a_2=0.01 (see the likelihood equation above); these values were determined empirically using 5-fold cross validation on the training data. The MLP experiments here used three hidden layers with the rectified linear activation function. The first, second, and third layers had 60, 40 and 40 nodes, respectively. Each layer used a 50% dropout rate [20], with L1 regularization (a_1=0.0003). Again, these model settings were determined through 5-fold cross validation experiments on the training sets. As with regularized logistic regression, the same MLP model settings were used across all three institutions' datasets.

Varying the Prediction Horizon

Although our focus involved predicting deterioration 6 hours before the event, we also considered how well the models performed across different prediction horizons. Additionally, we wanted to further examine the contributions of different groups of features from the various clinical elements to determine how particular feature groups performed at each horizon interval. Figure 2 shows the results for models trained and evaluated at prediction horizons ranging from 1 to 16 hours. We examined four different models where feature groups were successively added, starting with *vitals*, then adding *labs*, then *acuity* and finally *assessments* to arrive at the full model. As we were also interested in understanding how each feature group performed independently of the others; Figure 3 presents results over different prediction horizons considering at each group of features separately. These results show robustness in the models' ability to predict deterioration even 10 to 16 hours before the event.

In addition, we examined how well a model trained for a particular prediction horizon performed when evaluated against varying prediction horizons. We carried this out by looking at a set of cross horizon experiments taking the 16 models trained across prediction horizons from 1 to 16 hours (using all available features) and evaluating each of those models against horizons ranging from 1 to 16. These results are presented as surface plots shown in Figure 4.

Model Comparison

A final set of experiments compared the performance of MLP and regularized logistic regression models, shown in Figure 5. The MLP models perform slightly better for shorter prediction horizons at BCH and CNHS.

Table 4. Evaluation results across all three institutions with various feature sets using a fixed prediction horizon of 6 hours for both training and testing.

Feature set	auROC[a] (95% CI)	Specificity at 0.8 sensitivity (95% CI)	PPV[b] % at 0.8 sensitivity (95% CI)
CCHMC			
All features (MLP[c])	0.890 (0.875-0.904)	0.805 (0.774-0.837)	5.19 (4.48-6.19)
All features	0.886 (0.871-0.901)	0.811 (0.773-0.839)	5.45 (4.47-6.30)
All-Vitals	0.881 (0.866-0.896)	0.784 (0.749-0.830)	4.67 (4.04-5.95)
All-Labs	0.878 (0.862-0.893)	0.802 (0.762-0.830)	5.17 (4.26-5.64)
All-Acuity	0.880 (0.865-0.895)	0.791 (0.761-0.820)	4.91 (4.23-5.63)
All-Assessments[d]	0.865 (0.849-0.880)	0.763 (0.718-0.791)	4.29 (3.59-4.85)
Vitals[d]	0.751 (0.728-0.775)	0.539 (0.470-0.599)	2.21 (1.91-2.53)
Labs[d]	0.651 (0.618-0.685)	0.315 (0.263-0.403)	1.47 (1.37-1.70)
Acuity[d]	0.746 (0.719-0.774)	0.551 (0.474-0.618)	2.24 (1.93-2.65)
Assessments[d]	0.846 (0.828-0.865)	0.738 (0.691-0.775)	3.88 (3.28-4.50)
BCH			
All features (MLP)	0.911 (0.891-0.930)	0.875 (0.834-0.908)	7.73 (5.97-10.3)
All features	0.902 (0.880-0.923)	0.873 (0.832-0.898)	7.66 (5.90-9.36)
All-Vitals	0.902 (0.882-0.922)	0.863 (0.807-0.901)	7.14 (5.18-9.62)
All-Labs	0.880 (0.857-0.903)	0.813 (0.722-0.874)	5.33 (3.65-7.71)
All-Acuity	0.884 (0.862-0.907)	0.831 (0.773-0.878)	5.87 (4.44-7.95)
All-Assessments	0.885 (0.862-0.909)	0.855 (0.798-0.889)	6.77 (4.96-8.67)
Vitals[d]	0.732 (0.699-0.765)	0.479 (0.409-0.579)	1.98 (1.75-2.44)
Labs[d]	0.803 (0.771-0.835)	0.601 (0.518-0.665)	2.57 (2.14-3.05)
Acuity[d]	0.812 (0.782-0.842)	0.590 (0.515-0.722)	2.51 (2.13-3.65)
Assessments[d]	0.814 (0.788-0.842)	0.668 (0.543-0.738)	3.08 (2.25-3.87)
CNHS			
All features (MLP)	0.890 (0.872-0.910)	0.771 (0.718-0.826)	4.40 (3.62-5.71)
All features	0.884 (0.862-0.905)	0.803 (0.740-0.863)	5.08 (3.89-7.14)
All-Vitals	0.899 (0.879-0.919)	0.856 (0.805-0.887)	6.82 (5.13-8.53)
All-Labs	0.869 (0.845-0.893)	0.761 (0.676-0.840)	4.22 (3.15-6.18)
All-Acuity	0.866 (0.842-0.890)	0.761 (0.678-0.823)	4.22 (3.17-5.62)
All-Assessments[d]	0.853 (0.828-0.879)	0.700 (0.635-0.788)	3.39 (2.81-4.73)
Vitals[d]	0.722 (0.689-0.755)	0.471 (0.412-0.569)	1.95 (1.76-2.39)
Labs[d]	0.700 (0.661-0.740)	0.458 (0.359-0.533)	1.91 (1.62-2.21)
Acuity[d]	0.735 (0.695-0.775)	0.345 (0.276-0.451)	1.58 (1.43-1.88)
Assessments[d]	0.844 (0.818-0.871)	0.683 (0.629-0.745)	3.22 (2.76-3.97)

[a]Area under the receiver operator characteristic curve.

[b]Positive predictive value.

[c]MLP: multilayer perceptrons.

[d]Indicates results that are statistically significant compared to the best result for each institution (DeLong test, $P<.05$).

Figure 2. Model performance with increasingly complex (additive) feature sets across prediction horizons, including 95% CIs. ROC: receiver operating characteristic; BCH: Boston Children's Hospital; CCHMC: Cincinnati Children's Hospital and Medical Center; CNHS: Children's National Health System.

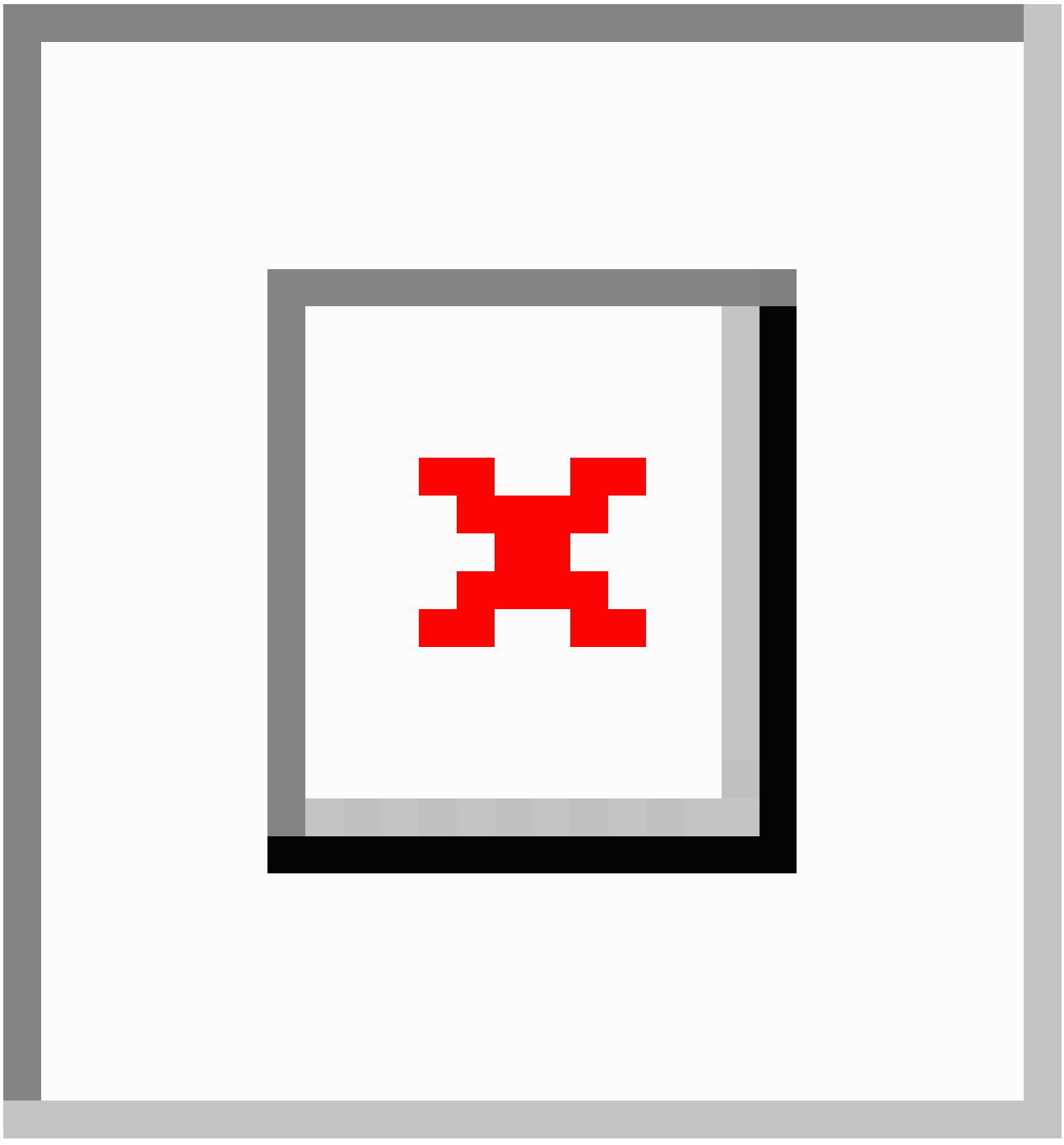


Figure 3. Performance of models with individual feature sets across prediction horizons, including 95% CIs. ROC: receiver operating characteristic; BCH: Boston Children's Hospital; CCHMC: Cincinnati Children's Hospital and Medical Center; CNHS: Children's National Health System.

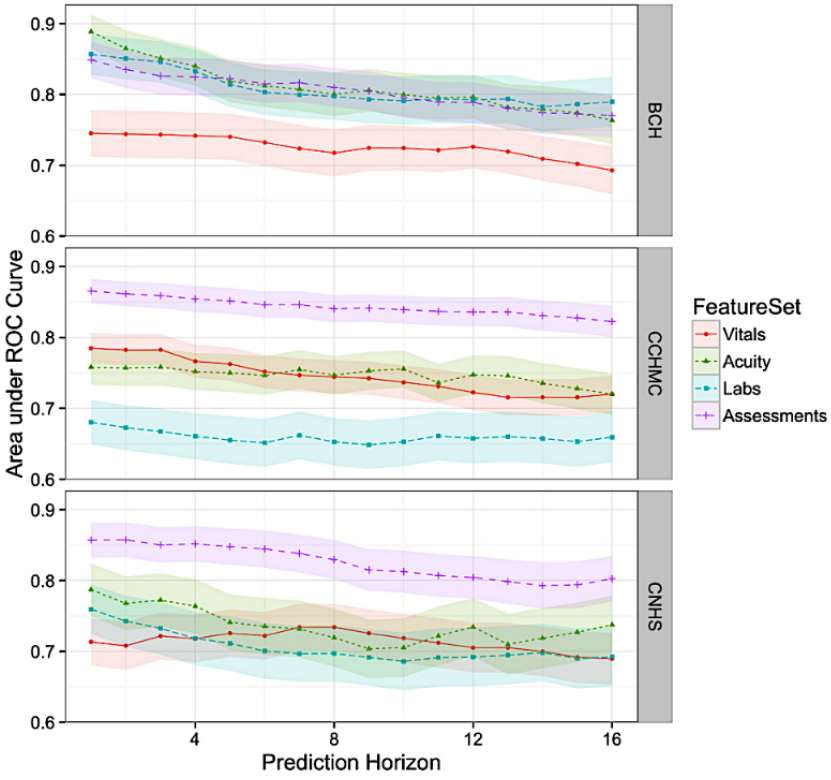


Figure 4. Area under receiver operating characteristic (ROC) curve when training and evaluating models across prediction horizons ranging from 1 hour to 16 hours. BCH: Boston Children's Hospital; CCHMC: Cincinnati Children's Hospital and Medical Center; CNHS: Children's National Health System.

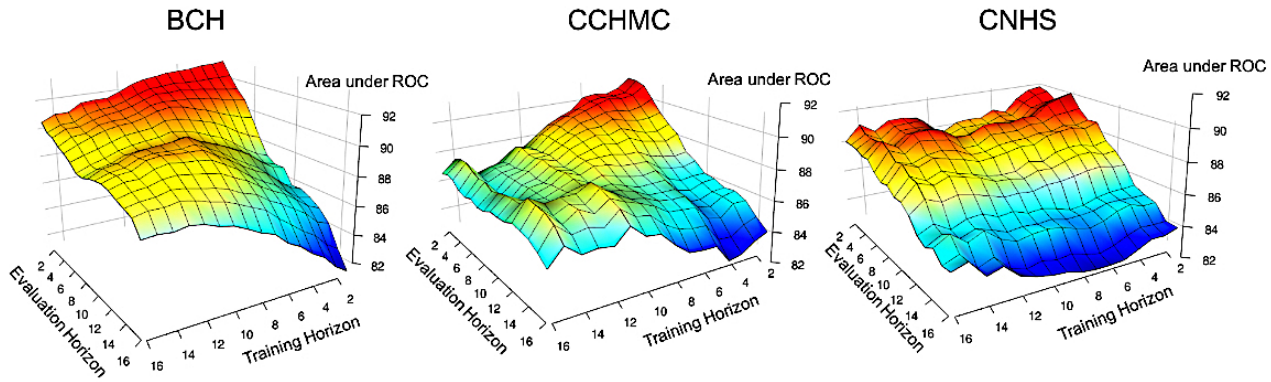
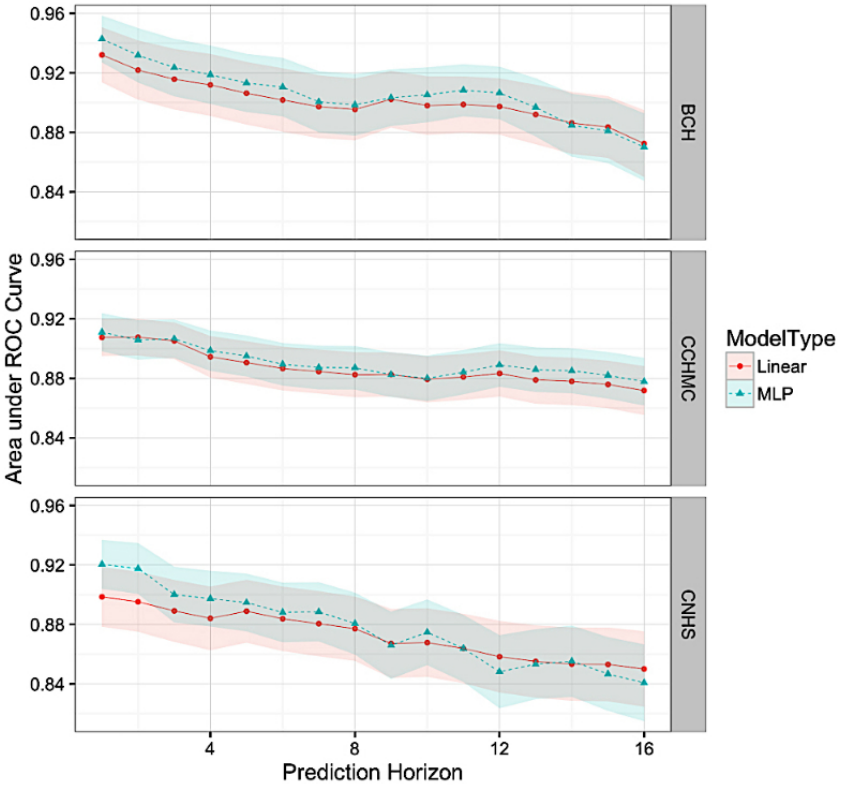


Figure 5. Best regularized logistic regression (linear) model in comparison with a multilayer perceptron (MLP) across different prediction horizons. ROC: receiver operating characteristic; BCH: Boston Children's Hospital; CCHMC: Cincinnati Children's Hospital and Medical Center; CNHS: Children's National Health System.



Discussion

Analysis of Results

Across all three institutions, our best models generally use all available features. The results show a somewhat consistent pattern across institutions, with the CNHS results generally lower, possibly beacause of less available data. The MLP provides a nonstatistically significant, but consistent, benefit over the linear model in terms of area under the ROC curve. Noteworthy is how the combination of the four different types of features generally provides the best performance, though we do note that removing vitals from the feature sets does not affect the BCH model and, in fact, slightly improves the CNHS model. Nursing assessments provide a strong indication of future deterioration, a finding that holds across all three institutions. This finding is consistent with recent work predicting sepsis that demonstrated significant benefits to utilizing text comment fields [24]. We anticipate that with additional labeled data, the nonlinear MLP model may outperform the logistic regression model. Recent work at predicting deterioration has demonstrated the utility of nonlinear models for predicting deterioration [14], when sufficient data are available. These results are encouraging, showing that a complex MLP with three hidden layers can be regularized sufficiently to avoid overfitting.

Features based on laboratory results, acuity scores, and nursing assessments differed across the three hospitals. These differences were because of the fact that some types of clinical data, nursing assessments in particular, lack a one-to-one mapping across institutions. In addition, vocabularies differ across EHR systems and institutions. For example, one institution might have a nursing assessment "Level of Consciousness" while another abbreviates it to "LOC." In a similar vein, the values (eg, "drowsy," "sleepy," and "alert") are institution-specific terms, some of which may not map to values at another institution. Rather than attempting to normalize all these clinical elements to the same vocabulary, features were constructed by simply taking the attribute-value pairs as they were realized in the EHR, directly from the corresponding database fields. This has an advantage of reducing the time and labor involved for building a model for new institutions' EHR systems as it obviates the need to map to a standard feature vocabulary. On the downside, however, each model is specific to a single institution.

When considering deploying deterioration prediction models in the hospital setting, a natural question arises as to the robustness of models across different prediction horizons. For example, if the model is trained to forecast deterioration 10 hours in advance but a patient is, in fact, just 2 hours away from a deterioration event, how well might the model perform? Not surprisingly, our results here demonstrated that ideally models should be trained and used to predict deterioration at a fixed horizon. For example, models trained at predicting deterioration only a few hours away perform very poorly at predicting deterioration 10 to 16 hours prior.

Most previous methods to detect deterioration are more limited than ours. The use of early warning scores, such as the Pediatric Early Warning Scores (PEWS) [25] and Children's Hospitals Early Warning Score (CHEWS) [26] to assess the severity of a patient's illness can provide warnings up to 11 hours before code and rapid response team (RRT) events [27]. Yet, these scores require manual entry by nurses and only consider small

sets of clinical elements. Other work predicting deterioration uses markedly smaller feature sets than ours [12,14] and make use of 29 predictor variables. The Rothman Index (RI) uses 26 variables [28]. In contrast, we have upwards of 4000 predictor variables across four different types of clinical elements. We believe our rich set of predictor variables not only improves the accuracy of our models but increases their robustness to missing data. Indeed, removing any single feature group only mildly degrades the models' accuracies, except for the case of nursing assessments. The RI [29-31] and the pediatric RI [32] use stepwise logistic regression for the purpose of predicting 1-year postdischarge risk of mortality and other adverse outcomes. It demonstrated the usefulness of including nursing assessments in predicting patient outcomes; however, it is not used to predict unplanned ICU transfers.

Other previous research also focused on physiologic patient characteristics to predict deterioration. Zhai et al [12] developed an EHR-based logistic regression algorithm to predict escalations to the pediatric ICU (PICU) in the first 24 hours after admission from the emergency department (ED). This work highlights several clinical elements that can be leveraged and while the study focuses on pediatric patients, it limits the patient population to only those who had an unplanned transfer to the ICU within 24 hours. Although direct comparisons are not possible because of different experimental conditions, we note Zhai et al [12] achieved 0.912 area under ROC, predicting deterioration 1 hour in advance. Churpek et al [14] obtained lower results (0.79 area under ROC); their prediction horizon ranges from 8 to 16 hours, and they make use of fewer clinical elements than our models. Recent work by Horng et al [24] predicted the occurrence of infection for purposes of sepsis clinical decision support, showing the importance of text analysis in conjunction with vitals for the task.

In contrast to previous studies, we looked carefully at a range of prediction horizons. Zhai et al [12], included predictions with a fixed horizon of 1 hour, whereas in the study by Churpek et al [14], the horizon effectively varied from 8 to 16 hours. Understanding the model's predictive power at specific horizons is necessary to determine how frequently the model should be invoked to provide a new risk assessment for deterioration. Here, our model shows robustness to longer horizons, meaning that it may prove beneficial even in settings in which the model can only be run infrequently because of strains it may place on EHR system infrastructure.

Limitations

There are many directions for future work. Improved methods for handling the nonstationary properties and sampling bias underlying health care data may provide better features through alternative parameterizations of time such as sequence time [33]. Across time scales of months or years, there is potential for data drift as patient populations and practice within the hospital setting change. Methods to detect data drift [34] and ameliorate them [35,36] would increase robustness and provide indications as to when models need to be retrained. Relatedly, models that capture the nonindependent sequence of predictions over time for the same patient, in a state-space or Markov model, may perform better and indicate trends.

In some cases models performed slightly better at longer prediction horizons; we hypothesize some of these trends are caused by noisy or missing inputs. Better features such as those derived from procedures, medication ordering, and administration may provide measures of the patient's complexity and acuity. Finally, rich information is present in various free-text fields [37,38,24] that may provide indicators of clinician concern.

Methods for providing explanations of model predictions in terms of the predictor variables present may have benefits in terms of validation and clinician acceptance [15]. On the other hand, minimizing labor-intensive feature extraction altogether is an interesting avenue to explore. Specifically, deep learning techniques [39,40] that help to learn representations automatically appear promising.

Adjustment of the outcome variable itself is another area for refinement. Many patients not identified as cases, as they were never transffered to the ICU, could be considered cases by virtue of their potential to have resulted in a deterioration event, had interventions not occurred. Expanding the cases to include patients based on certain interventions may be worth exploring. Another formulation would be to train the model to predict deterioration for some *interval of time* in the future, for example, 4 to 6 hours. This may improve the robustness of the model. Survival analysis based on hazard models is another approach where the goal is to measure the time until deterioration, yet challenges arise from censoring [41] and competing events [42] based on the fact that many patients never go on to have a deterioration event. Finally, in cases where the outcome variable of interest can be observed (eg, acuity scores) or computed (eg, sequential organ failure assessment [SOFA], scores for sepsis [43,44]) as a scalar value at various points in time from EHR retrospectively, deterioration could be formulated as a *forecasting* problem. Although forecasting models are inherently more complex (as they provide a series of nonindependent predictions), they may provide better interpretability, especially in conjunction with CIs associated with the forecast.

Practical Implications

Deployed in the hospital setting, this model may supplement existing detection tools in use such as safety huddles or rapid response teams to improve the recognition of patients at risk of experiencing an unplanned ICU transfer. Ultimately, the results of the model could lead clinicians to detect deterioration and act sooner. This may avoid serious events that lead to higher rates of morbidity and mortality. There is also great potential to reduce cost through fewer inpatient days, shorter ICU stays, and fewer and less extreme medical interventions.

Conclusions

This paper described a machine learning approach to predict deterioration in pediatric patients as indicated by an unplanned ICU transfer by leveraging rich sets of clinical elements in the EHR. Our study, carried out at three separate institutions with different EHR systems, suggests that such approaches to predicting deterioration have a great potential to improve care and reduce costs [5]. By analyzing how prediction quality changes across different prediction horizons, we have provided

insight into how such a model would fare in a real clinical setting. In addition, our research suggests that feature-rich, data-driven models may perform at a superior level to existing models reported in the literature based on small numbers of carefully tuned variables. Ultimately, the model output may be integrated in workflows of rapid response teams and safety leads so that deterioration could be recognized earlier.

Acknowledgments

The authors wish to thank the participating hospitals: Boston Children's Hospital, Children's National Medical Center, and Cincinnati Children's Hospital Medical Center. The participating members from each hospital were instrumental in providing clinical expertise that drove the development and refinement of the models. In particular, we wish to thank Steve Muething, MD.

Conflicts of Interest

None declared.

References

1. Buist MD, Jarmolowski E, Burton PR, Bernard SA, Waxman BP, Anderson J. Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care. A pilot study in a tertiary-care hospital. Med J Aust 1999 Jul 05;171(1):22-25. [Medline: 10451667]
2. Kause J, Smith G, Prytherch D, Parr M, Flabouris A, Hillman K, Intensive Care Society (UK), AustralianNew Zealand Intensive Care Society Clinical Trials Group. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom--the ACADEMIA study. Resuscitation 2004 Sep;62(3):275-282. [doi: 10.1016/j.resuscitation.2004.05.016] [Medline: 15325446]
3. Schein RM, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. Chest 1990 Dec;98(6):1388-1392. [Medline: 2245680]
4. Chan PS, Jain R, Nallmothu BK, Berg RA, Sasson C. Rapid Response Teams: A Systematic Review and Meta-analysis. Arch Intern Med 2010 Jan 11;170(1):18-26. [doi: 10.1001/archinternmed.2009.424] [Medline: 20065195]
5. Bonafide CP, Localio AR, Song L, Roberts KE, Nadkarni VM, Priestley M, et al. Cost-benefit analysis of a medical emergency team in a children's hospital. Pediatrics 2014 Aug;134(2):235-241 [FREE Full text] [doi: 10.1542/peds.2014-0140] [Medline: 25070310]
6. Nadkarni VM, Larkin GL, Peberdy MA, Carey SM, Kaye W, Mancini ME, et al. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. JAMA 2006 Jan 4;295(1):50-57. [doi: 10.1001/jama.295.1.50] [Medline: 16391216]
7. Brilli RJ, Gibson R, Luria JW, Wheeler TA, Shaw J, Linam M, et al. Implementation of a medical emergency team in a large pediatric teaching hospital prevents respiratory and cardiopulmonary arrests outside the intensive care unit. Pediatr Crit Care Med 2007 May;8(3):236-46; quiz 247. [doi: 10.1097/01.PCC.0000262947.72442.EA] [Medline: 17417113]
8. Das S, Eisenberg LD, House JW, Lee KJ, Lusk RP, Nielsen DR, et al. Meaningful use of electronic health records in otolaryngology: recommendations from the American Academy of Otolaryngology--Head and Neck Surgery Medical Informatics Committee. Otolaryngol Head Neck Surg 2011 Feb;144(2):135-141. [doi: 10.1177/0194599810393441] [Medline: 21493407]
9. Badawi O, Brennan T, Celi LA, Feng M, Ghassemi M, Ippolito A, et al. Making big data useful for health care: a summary of the inaugural mit critical data conference. JMIR Med Inform 2014;2(2):e22 [FREE Full text] [doi: 10.2196/medinform.3447] [Medline: 25600172]
10. Kellermann AL, Jones SS. What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Aff (Millwood) 2013 Jan;32(1):63-68. [doi: 10.1377/hlthaff.2012.0693] [Medline: 23297272]
11. Johnson AEW, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine Learning and Decision Support in Critical Care. Proc IEEE Inst Electr Electron Eng 2016 Feb;104(2):444-466 [FREE Full text] [doi: 10.1109/JPROC.2015.2501978] [Medline: 27765959]
12. Zhai H, Brady P, Li Q, Lingren T, Ni Y, Wheeler DS, et al. Developing and evaluating a machine learning based algorithm to predict the need of pediatric intensive care unit transfer for newly hospitalized children. Resuscitation 2014 Aug;85(8):1065-1071 [FREE Full text] [doi: 10.1016/j.resuscitation.2014.04.009] [Medline: 24813568]
13. Tsoukalas A, Albertson T, Tagkopoulos I. From data to optimal decision making: a data-driven, probabilistic machine learning approach to decision support for patients with sepsis. JMIR Med Inform 2015 Feb 24;3(1):e11 [FREE Full text] [doi: 10.2196/medinform.3445] [Medline: 25710907]
14. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. Crit Care Med 2016 Feb;44(2):368-374. [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]

15. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. 2015 Aug Presented at: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015; Sydney, Australia.
16. Luo G. Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. Health Inf Sci Syst 2016;4:2 [FREE Full text] [doi: 10.1186/s13755-016-0015-4] [Medline: 26958341]
17. Tee AC, Ching NF, Keow NS. In-hospital mortality among unplanned admissions to a medical intensive care unit. Critical Care and Shock 2008;11:149-153.
18. Ng A. Feature selection, l1 vs l2 regularization and rotational invariance. 2004 Presented at: Twenty-first International Conference on Machine Learning; 2004; New York, NY, USA.
19. Demir-Kavuk O, Kamada M, Akutsu T, Knapp E. Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features. BMC Bioinformatics 2011 Oct 25;12:412 [FREE Full text] [doi: 10.1186/1471-2105-12-412] [Medline: 22026913]
20. Srivastava N, Hinton G, Krizhevsky A. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research. January 2014;15(1):1929-1958.
21. Zhang T. Solving large scale linear prediction problems using stochastic gradient descent algorithms. 2004 Presented at: Proceedings of the twenty-first international conference on Machine learning; 2004; Banff, Alberta, Canada.
22. Recht B, Re C, Wright S, Liu F. Hogwild: A lock-free approach to parallelizing stochastic gradient descent. Advances in Neural Information Processing Systems 2011:693-701.
23. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. Crit Care 2015 Aug 13;19:285 [FREE Full text] [doi: 10.1186/s13054-015-0999-1] [Medline: 26268570]
24. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One 2017;12(4):e0174708 [FREE Full text] [doi: 10.1371/journal.pone.0174708] [Medline: 28384212]
25. Duncan H, Hutchison J, Parshuram CS. The Pediatric Early Warning System score: a severity of illness score to predict urgent medical need in hospitalized children. J Crit Care 2006 Sep;21(3):271-278. [doi: 10.1016/j.jcrc.2006.06.007] [Medline: 16990097]
26. Kleinman M, Romano J. Children's Hospital Boston Early Warning Score Early detection + Early intervention = Better outcomes. FIRST Do no harm 2010:2-3 [FREE Full text]
27. Akre M, Finkelstein M, Erickson M, Liu M, Vanderbilt L, Billman G. Sensitivity of the pediatric early warning score to identify patient deterioration. Pediatrics 2010 Apr;125(4):e763-e769. [doi: 10.1542/peds.2009-0338] [Medline: 20308222]
28. Finlay GD, Rothman MJ, Smith RA. Measuring the modified early warning score and the Rothman index: advantages of utilizing the electronic medical record in an early warning system. J Hosp Med 2014 Feb;9(2):116-119 [FREE Full text] [doi: 10.1002/jhm.2132] [Medline: 24357519]
29. Rothman MJ, Solinger AB, Rothman SI, Finlay GD. Clinical implications and validity of nursing assessments: a longitudinal measure of patient condition from analysis of the Electronic Medical Record. BMJ Open 2012;2(4) [FREE Full text] [doi: 10.1136/bmjopen-2012-000849] [Medline: 22874626]
30. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. J Biomed Inform 2013 Oct;46(5):837-848 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.011] [Medline: 23831554]
31. Finlay GD, Rothman MJ, Smith RA. Measuring the modified early warning score and the Rothman index: advantages of utilizing the electronic medical record in an early warning system. J Hosp Med 2014 Feb;9(2):116-119 [FREE Full text] [doi: 10.1002/jhm.2132] [Medline: 24357519]
32. da SYS, Hamilton MF, Horvat C, Fink EL, Palmer F, Nowalk AJ, et al. Evaluation of Electronic Medical Record Vital Sign Data Versus a Commercially Available Acuity Score in Predicting Need for Critical Intervention at a Tertiary Children's Hospital. Pediatr Crit Care Med 2015 Sep;16(7):644-651. [doi: 10.1097/PCC.0000000000000444] [Medline: 25901545]
33. Hripcsak G, Albers DJ, Perotte A. Parameterizing time in electronic health record studies. J Am Med Inform Assoc 2015 Jul;22(4):794-804. [doi: 10.1093/jamia/ocu051] [Medline: 25725004]
34. Tsymbal A. The problem of concept drift: definitions and related work. Computer Science Department, Trinity College Dublin 2004;106(2).
35. Elwell R, Polikar R. Incremental learning of concept drift in nonstationary environments. IEEE Trans Neural Netw 2011 Oct;22(10):1517-1531. [doi: 10.1109/TNN.2011.2160459] [Medline: 21824845]
36. Hulten G, Spencer L, Domingos P. Mining time-changing data streams. 2001 Presented at: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining; 2001; San Francisco.
37. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009 Oct;42(5):760-772 [FREE Full text] [doi: 10.1016/j.jbi.2009.08.007] [Medline: 19683066]
38. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]

39. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
40. Zheng Y, Liu Q, Chen E, Ge Y, Zhao JL. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. Front. Comput. Sci 2015 Sep 8;10(1):96-112. [doi: 10.1007/s11704-015-4478-2]
41. Singer JD, Willett JB. It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. Journal of Educational and Behavioral Statistics 1993 Jan 01;18(2):155-195. [doi: 10.3102/10769986018002155]
42. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association 1999 Jun;94(446):496-509. [doi: 10.1080/01621459.1999.10474144]
43. Vincent JL, Moreno R, Takala J, Willatts S, De MA, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996 Jul;22(7):707-710. [Medline: 8844239]
44. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016 Feb 23;315(8):801-810. [doi: 10.1001/jama.2016.0287] [Medline: 26903338]

Abbreviations

ADT: admission-discharge-transfer
auROC: area under the receiver operator characteristic curve
BCH: Boston Children's Hospital
CCHMC: Cincinnati Children's Hospital Medical Center
CDE: critical deterioration event
CHEWS: children's hospitals early warning score
CNHS: Children's National Health System
ED: emergency department
EHR: electronic health record
ICD-9: international classification of diseases - 9
ICU: intensive care unit
MLP: multilayered perceptron
NICU: neonative intensive care unit
OR: operating room
PACU: postanesthesia care unit
PEWS: pediatric early warning score
PPV: positive predictive value
RI: Rothman index
RRT: rapid response team
ROC: receiver operating characteristic
SOFA: sequential organ failure assessment
SQL: structured query language

Edited by G Eysenbach; submitted 11.08.17; peer-reviewed by S Purkayastha, I Mircheva; comments to author 06.09.17; revised version received 22.09.17; accepted 23.09.17; published 22.11.17.

Please cite as:
Wellner B, Grand J, Canzone E, Coarr M, Brady PW, Simmons J, Kirkendall E, Dean N, Kleinman M, Sylvester P
Predicting Unplanned Transfers to the Intensive Care Unit: A Machine Learning Approach Leveraging Diverse Clinical Elements
JMIR Med Inform 2017;5(4):e45
URL: http://medinform.jmir.org/2017/4/e45/
doi:10.2196/medinform.8680
PMID:29167089

Original Paper

Adverse Drug Event Discovery Using Biomedical Literature: A Big Data Neural Network Adventure

Ahmad P Tafti[1], PhD; Jonathan Badger[1], PharmD; Eric LaRose[1], BS; Ehsan Shirzadi[2], MSc; Andrea Mahnke[1], MSc; John Mayer[1], PhD; Zhan Ye[1], PhD; David Page[3], PhD; Peggy Peissig[1], MBA, PhD

[1]Biomedical Informatics Research Center, Marshfield Clinic Research Institute, Marshfield, WI, United States

[2]Institute of Electrical and Electronics Engineers, Dublin, Ireland

[3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Ahmad P Tafti, PhD
Biomedical Informatics Research Center
Marshfield Clinic Research Institute
1000 N Oak Ave
Marshfield, WI, 54449
United States
Phone: 1 7153893514
Email: pahlavantafti.ahmad@marshfieldresearch.org

Abstract

Background: The study of adverse drug events (ADEs) is a tenured topic in medical literature. In recent years, increasing numbers of scientific articles and health-related social media posts have been generated and shared daily, albeit with very limited use for ADE study and with little known about the content with respect to ADEs.

Objective: The aim of this study was to develop a big data analytics strategy that mines the content of scientific articles and health-related Web-based social media to detect and identify ADEs.

Methods: We analyzed the following two data sources: (1) biomedical articles and (2) health-related social media blog posts. We developed an intelligent and scalable text mining solution on big data infrastructures composed of Apache Spark, natural language processing, and machine learning. This was combined with an Elasticsearch No-SQL distributed database to explore and visualize ADEs.

Results: The accuracy, precision, recall, and area under receiver operating characteristic of the system were 92.7%, 93.6%, 93.0%, and 0.905, respectively, and showed better results in comparison with traditional approaches in the literature. This work not only detected and classified ADE sentences from big data biomedical literature but also scientifically visualized ADE interactions.

Conclusions: To the best of our knowledge, this work is the first to investigate a big data machine learning strategy for ADE discovery on massive datasets downloaded from PubMed Central and social media. This contribution illustrates possible capacities in big data biomedical text analysis using advanced computational methods with real-time update from new data published on a daily basis.

(JMIR Med Inform 2017;5(4):e51) doi:10.2196/medinform.9170

KEYWORDS

adverse drug event; adverse drug reaction; drug side effects; machine learning; text mining

Introduction

Background

Adverse drug events (ADEs), defined as the set of detriments or injuries caused by a medication, have led to additional medical costs, prolonged hospitalization, morbidity, and ascribable disability worldwide [1-4]. ADEs encompass all adverse drug reactions but also include preventable causes of errors such as inappropriate dosing, dispensing errors, and drug abuse. Discovery of ADEs has gained great attention in the health care community, and in the last few years, several drug risk-benefit assessment strategies have been developed to analyze drug efficacy and safety using different medical data

sources, ranging from electronic health records (EHRs) to human-health–related social media and drug reviews [5-14]. A variety of combined computational methods using natural language processing (NLP), machine learning strategies, and text retrieval algorithms have been employed to extract ADEs from such data sources [15-23]. Clinical trials, EHRs, and medical case reports are additional biomedical data-rich sources that have been utilized for ADE extraction [24-28].

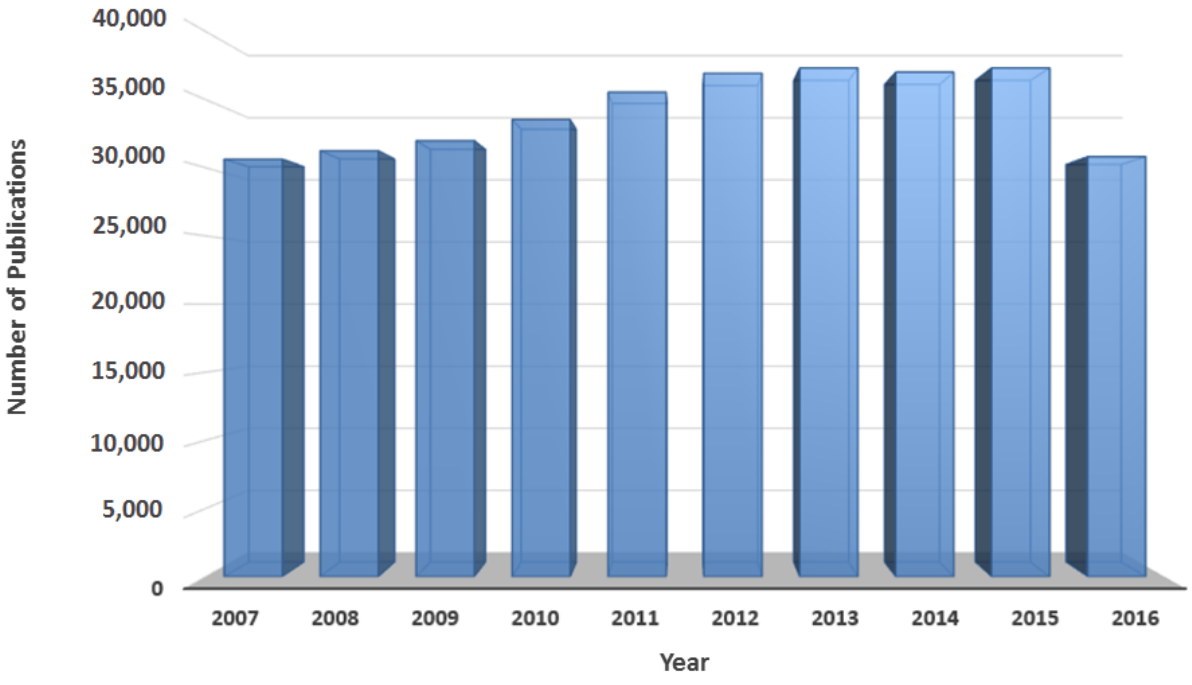
In recent years, biomedical articles produced by scientists all across the world have grown extensively. Figure 1 [29] shows that the number of journal and conference papers published in different medication studies (eg, ADEs and drug analysis, drug evaluation, and drug repositioning) rapidly grew in number from year 2007 to 2016. The total number of publications in those years is approximately 342,301 articles. To roughly estimate the size of such scientific papers, we assumed a PDF file format for each article. The size of a PDF file depends on the number of pages and pictures or metadata inside the file. Considering a 9 to 11 page PDF including plain text along with a few pictures, it may equal almost 3 MB size in average, and it appears, approximately 1.02 TB articles were generated in drug associated studies from 2007 to 2016. The other file formats such as extensible markup language (XML), may be much larger in size. Scientific articles published in biomedical research are usually generated using standardized and principled methods and therefore, are especially valuable for high-quality knowledge discovery. This great deluge of information includes an enormous number of scientific publications on ADEs' study, an area of focus into which many biomedical researchers have entered, developing a variety of research activities for discovering, analyzing, and monitoring ADEs [30-38].

It is impossible for researchers, scientists, and physicians to read and process the large body of scientific articles and remain abreast of the foremost information regarding ADEs. Therefore, there is a pressing need to develop intelligent computational methods, particularly big data analytics solutions, to efficiently process this wealth of data. Big data biomedical text analysis utilizes advanced computational technologies including big data infrastructure, NLP, statistical analytics, and machine learning algorithms to extract facts from text data. This in turn generates new hypotheses by systematically analyzing large numbers of scientific publications.

Objectives and the Main Contributions

Whereas ADE discovery from diverse biomedical data sources in general has been studied historically in health care informatics, the use of big data scientific articles and health-related social media for ADE discovery has been very limited so far. The motivation of this work is to study big data machine learning solutions, particularly big data neural networks (bigNN), to analyze ADEs from large-scale biomedical text data, developing a scalable framework to fulfill the following objectives: (1) to extract current knowledge and high-quality information about ADEs using full text scientific articles and social media, (2) to utilize and adapt advanced NLP and machine learning algorithms in a large-scale fashion by the use of big data infrastructures, and (3) to provide better insights and tendencies in large-scale biomedical text analytics and identify the challenges and potential enhancements toward efficient and accurate ADE discovery. We briefly summarize our *main contributions* as follows:

Figure 1. The number of publications in several medication studies available at PubMed over the last 10 years. The results obtained by submitting a query: ((((((((((drug analysis[MeSH Terms]) OR drug analysis[MeSH Subheading]) OR adverse drug event[MeSH Terms]) OR adverse drug event[MeSH Subheading]) OR adverse drug reaction[MeSH Terms]) OR adverse drug reaction[MeSH Subheading]) OR drug evaluation[MeSH Terms]) OR drug evaluation[MeSH Subheading])) OR drug repositioning[MeSH Terms]) OR drug repositioning[MeSH Subheading].



We initiated a study of big data literature mining for ADE discovery with the use of two different data sources: (1) published full text scientific articles available on PubMed Central [39], and (2) posts available in health-related social media, including MedHelp [40], patient [41], and WebMD [42]. Although several promising approaches have been designed for biomedical text mining, the development of scalable machine learning frameworks capable of ADE extraction from big data is very limited. To the best of our knowledge, our work is the first to investigate a bigNN strategy for ADE discovery on massive datasets downloaded from PubMed Central and social media.

- With the current work and using big data analytics platforms such as Elasticsearch and Apache Spark, we developed a scalable framework to analyze and visualize ADEs from hundreds of thousands of published scientific articles and social media blog posts.
- Combining a variety of the internal neural network parameters, we presented a predictive model that obtained accuracy, precision, recall, and area under 92.7%, 93.6%, 93.0%, and 0.905, respectively, on a massive dataset downloaded from PubMed Central plus health-related social media.
- This paper opens the door to pursue large-scale biomedical literature mining and its application in health care informatics in general and introduces several possible enhancements to advance the level of the impact of this research area.

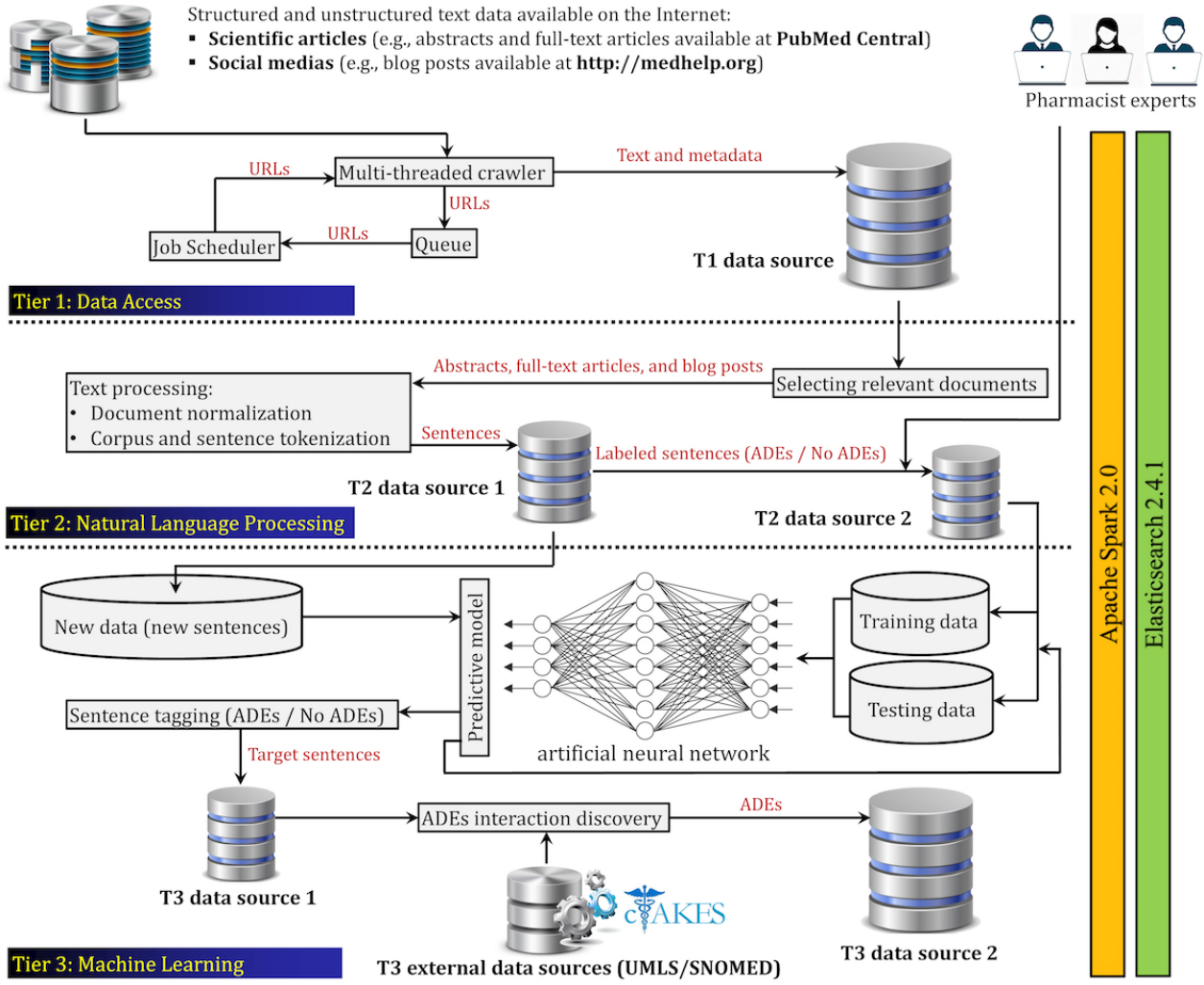
Methods

The general pipeline of the proposed ADE extraction framework is illustrated in Figure 2. In this section, we shall explain the underlying tiers of the proposed bigNN framework.

Tier 1: Data Access

Tier 1 systematically collects the expanding body of scientific articles and social media blog posts through different data sources available on the Internet. A multi-threaded crawler or downloader was developed to provide timely and efficient processing of the diverse big data content found on the Internet. *Scrapy* [43], a free and open source Web crawling system has been used to allow multiple threads to automatically fetch URLs from different sources. A queuing system and a scheduler have also been established as a part of the data access tier.

Figure 2. The proposed system for adverse drug event (ADE) discovery. All tiers developed on top of the Apache Spark 2.0 that utilizes an Elasticsearch database 2.4.1 to data storage and retrieval.



All collected data (eg, XML files) are first turned into plain text data, and together with associated metadata (eg, journal name, author list, and publication date) are stored in a single type (table type) inside a No-SQL database, namely Elasticsearch [44], which provides a distributed, open source, RESTful and full text search and analytics engine.

Tier 2: Natural Language Processing

Tier 2 includes several computational procedures to process raw text data, preparing potential ADE sentences to feed the next tier.

Selecting Relevant Documents

One of the major goals of the proposed system was to collect groups of sentences that provide evidence about drug-event pairs. Toward that outcome, there is an emergent need to identify trustworthy and reputable data sources (eg, well-founded and prestigious journals such as *Nature*, *PNAS*, and *PLOS*). As the proposed framework accumulates data from two separate data sources, including scientific journals and messages posted on social media, we established two different criteria to yield more credible data. Section A.1 of Multimedia Appendix 1 further discusses the proposed method and criteria.

Text Processing

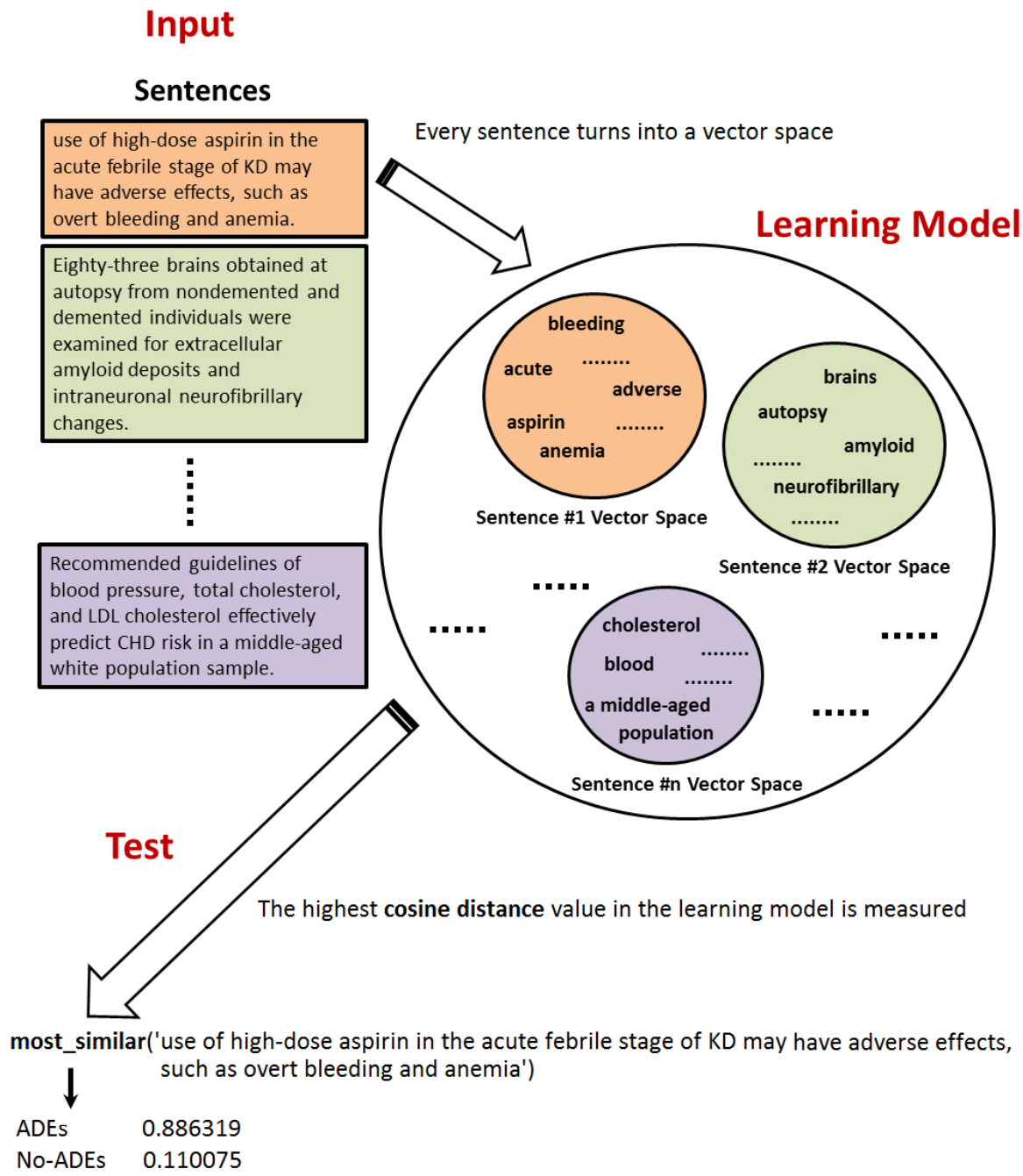
We first normalize all documents by converting corpora into a standard consistent form. This process (1) converts all characters to lower case, (2) transliterates to American Standard Code for Information Interchange if needed, and (3) deletes a set of existing substrings and patterns (eg, [], , ?, !, and ()). Once we complete the proposed text normalization process, we convert every document into a set of sentences. Although several ADEs could be captured among different sentences, extracting ADEs interactions across sentences is significantly more challenging than within sentences [6,45]. To feed the bigNN system, a random subset of sentences was selected for manual annotation by three domain experts; see Section A.2 of Multimedia Appendix 1 for details. The random subset of sentences includes health and medical-related text data either with or without ADE interactions. Sentences that are missing either a drug name or an adverse drug effect term were excluded. Using a Web application (Multimedia Appendix 1), the domain experts labeled individual sentences as ADEs, No-ADEs, or Not Decided. To focus on binary classification, we omitted sentences labeled “Not Decided” leaving two different classes: (1) ADEs and (2) No-ADEs. Section A.2 of Multimedia Appendix 1 explains how we made a training set for the machine learning tier.

Tier 3: Machine Learning

Tier 3 implements the core functionality needed for making a predictive model to distinguish ADEs sentences versus those that are No-ADEs. From the machine learning perspective, this means making a binary classification predictive model that assigns one of two trained labels to new unlabeled sentences. Therefore, when an arbitrary sentence arrives, the predictive model chooses one category exclusively from two predefined classes as ADEs or No-ADEs. There are two basic steps. The first step is to extract effective content as a set of features. The next step is the text classification assignment. The bag-of-words (BoW) representation, a widely used content extractor algorithm, has been around for several years in the text analytics domain, and it provides an easy way to turn text-based data records into a set of feature vectors such that the frequency of occurrence of words (eg, uni-grams and/or bi-grams, along with part-of-speech [POS] tagging) in the corpus is used as a feature vector to train a classifier (eg, support vector machine [SVM], decision tree, and/or logistic regression) [46-51]. The BoW representation maintains word intensities across the corpus, but it dissembles grammar, syntactic, semantic, and word order. In contrast to the BoW representation, the word2vector (word2vec) algorithm, originally developed at Google [52], includes a set of computational methods that turns a corpus of text data into a meaningful vector space that encompasses grammar, semantic, and word order. Word2vec comes with a two-layer neural network and is able to tackle several text analytics functions including dependency parsing [53,54], named entity recognition [55,56], text classification [57,58], and word clustering [59]. The model takes a text corpus as an input and turns each word in the corpus into a vector as illustrated in Section A.3 of Multimedia Appendix 1. It then groups vectors of similar words together in a vector space, training words against other words that neighbor them in the input text corpus [60,61]. Our bigNN system implemented the word2vec neural network, which is fully explained in Section A.3 of Multimedia Appendix 1.

An abstract view of the proposed learning algorithm is shown in Figure 3. As mentioned in Section A.3 of Multimedia Appendix 1, the algorithm builds a vector representation for words and/or sentences. If we train a learning model with 98-dimension, then we will obtain a 98-digit in front of each word (eg, “aspirin”). The cosine distance similarity [62,63], which is the normalized dot product between vectors, is then measured to find the best fitness class for a sentence. Once we have the labeled sentences as ADEs or No-ADEs, we focus on ADE sentences and find the positive adverse-drug interactions using cTAKES [64].

Figure 3. This figure depicts the abstract view of the proposed learning model where the cosine distance similarity is measured to select an appropriate category for a given sentence. Every single word is represented as a vector, and eventually every sentence (eg, adverse drug events [ADEs] or No-ADEs) turns in to a vector space. Once we have done with training the model, for every new coming sentence, the highest cosine distance value is then measured to find the best fitness class.



Results

Implementation and Test Bed

We investigated two different data sources: (1) biomedical articles and (2) health-related social media blog posts. The first data source included 97,246,719 sentences obtained from almost 1,451,413 abstracts and full text articles available on PubMed Central, and the second one consisted of 2,524,622 sentences obtained from 419,915 blog posts at MedHelp, Patient, and WebMD. To train our proposed predictive model illustrated in Figure 2, we first randomly selected different subsets for the purpose of manual annotation. Those subsets were annotated by three domain experts who have been working in medical and pharmacy domains. We defined two different classes as ADEs and No-ADEs to indicate drug-event interactions in a sentence. The sentences that included positive drug-event interaction were tagged as ADEs, and the others were tagged as No-ADEs. Before the large-scale manual extraction, we evaluated the interrater reliability among our three domain experts, the results of kappa statistics, .84, which indicates very good concordance and agreement; more details are reported in Section A.2 of Multimedia Appendix 1.

Table 1. Nine datasets were employed to make and evaluate the big data neutral network (bigNN) system. Each row identifies the dataset along with the number of ADEs and No-ADEs sentences within the dataset. Every row also shows how many of these sentences are human-labeled and/or machine-labeled. The datasets are separated into three main categories: biomedical articles (*_BA), social media posts (*_SM), and the combination of the two (*_Combined).

Dataset ID	Total number of sentences	Number of ADEs[a] sentences	Number of No-ADEs sentences	Number of human-labeled sentences	Number of machine-labeled sentences
ADEs#1_BA	6960	3311	3649	6960	0
ADEs#1_SM	400	160	240	400	0
ADEs#1_Combined	7360	3471	3889	7360	0
ADEs#2_BA	13,545	6359	7186	7015	6530
ADEs#2_SM	472	195	277	405	67
ADEs#2_Combined	14,017	6554	7463	7420	6597
ADEs#3_BA	21,278	10,307	10,971	7015	14,263
ADEs#3_SM	565	241	324	405	160
ADEs#3_Combined	21,843	10,548	11,295	7420	14,423

[a]ADEs: adverse drug events.

Next, we developed nine different datasets from those sentences to train the proposed bigNN system. The first set of datasets (the first three rows in Table 1) used only human-labeled datasets from our domain experts. Within these three datasets, three types of sentences were generated: the first type was biomedical articles, the second was social media, and the third was the combination of the two. For the second set of datasets (the three datasets in the middle of Table 1), we added some machine-labeled datasets (the number of datasets were reported in the last column, “number of machine-labeled sentences” of Table 1) with human-labeled datasets to increase the sample sizes. We also generated three types of sentences as illustrated above. For the third set of datasets (the last three rows of Table 1), we added in more machine-labeled datasets with human-labeled datasets; similar three types of sentences were generated.

We also utilized three smaller datasets, including 2600, 3500, and 4000 human-labeled data records and plotted learning curves to see how the accuracy of the predictive model varies with increasing amount of training data. The results were not promising enough, and we started with a larger human-labeled dataset as illustrated in the first three rows in Table 1. For all the experiment, we utilized 75% of every dataset to train the model and 25% to test it using four-fold cross validation, with no sentence to appear in both the training and testing sets at the same time.

Experimental Setup

Every programming module in Tier 1 and Tier 2 was developed by Python 2.7.13. Tier 3, the machine learning tier, was implemented by Java j2SE 8. All of these tiers were developed on top of the Apache Spark 2.0 [65] and Elasticsearch DB 2.4.1 [44] just to tackle the problem of big data analytics in an efficient and timely fashion. From the computational side, a dedicated computational resource, including two virtual machines in a VMWARE cluster environment, each running a 64-bit CentOS 6.8 operating system with 8 vCPUs, 16 GB RAM, and 1 TB HDD in total, hosted on a Xeon E5-2690V3 2.6 GHz CPU, were used to obtain the experimental results.

Experimental Validations

We analyzed the performance of the predictive model across all the datasets. Accuracy, precision, and recall obtained by the experiments are shown in Multimedia Appendix 2. The first column shows the dataset used to in the experiments. The second column describes a configuration setup for a set of internal parameters of the proposed neural network model. Minimum word frequency (MWF) allows for ignoring all words in the vocabulary with total occurrences lower than MWF value. Epoch (EP) is the number of forward and backward passes of all training examples. Window size (WS) defines context windows size to generate a vector representation for words across the documents. Iteration (ITR) defines the number of iterations done for each mini-batch during a training process. The last column shows elapsed time for the training stage. This does not reflect the time of text-preprocessing tasks such as normalization and tokenization. The current table shows that greater EP and ITR with the use of WS of two will provide better performance across all three datasets. To further analyze our proposed ADEs sentence discovery system, we also compared the proposed predictive model with the combination of BoW feature selection method and SVM, decision tree, and naïve Bayes classifiers. Uni-grams, bi-grams, along with POS tagging were used as BoW features to make a predictive model. One can see in Multimedia Appendix 2, the most promising accuracy results across all datasets obtained by the (MWF=2%, EP=25, WS=2, and ITR=10) configuration. Furthermore, we demonstrated that the performance of bigNN system, both accuracy and time of completing the task, is comparable with traditional SVM, naïve Bayesian, and decision tree with BoW strategy. The results of this experiment are shown in Table 2. All the measures in Table 2 are selected from the best performed model by tuning the models using different parameters for all the bigNN system and SVM, naïve Bayesian, and decision tree with BoW strategy. Regarding the BoW feature set, we obtained the best results by utilizing a combination of uni-grams and bi-grams, together

with POS tagging. With respect to the traditional machine learning classifiers, and for example SVM, the best performance was achieved with the use of radial basis function kernel, loss of 0.12, seed of 1, and without normalization of input data.

For each of the datasets shown in Table 2, we split the data randomly as 75% to train and 25% to test the proposed sentence classifier system. The best accuracy results using our proposed predictive model were obtained by MWF=2%, EP=25, WS=2, ITR=10 configuration. Regarding the BoW features set, we utilized a combination of uni-grams, bi-grams, and POS tagging. Using the proposed predictive model, the vocabulary size for ADEs#1_Combined, ADEs#2_Combined, and ADEs#3_Combined datasets were 15125, 26,448, and 37,524, respectively.

Whereas with the use of BoW, when it utilized uni-grams, bi-grams, and POS, the vocabulary size was 33,567, 59,941, and 76,758 for datasets ADEs#1_Combined, ADEs#2_Combined, and ADEs#3_Combined. For SVM, decision tree, and naïve Bayes classifiers, we utilized Weka library (version 3.7.12) [66] running on hadoop-2.7 [67] by the use of Hadoop distributed file system.

The study shows that bigNN system generates better results in comparison with traditional BoW along with SVM, decision tree, and naïve Bayes classification algorithms. Area under the curve (AUC) of our proposed predictive model across all three datasets was also analyzed, and it is shown in Figure 4.

The current test results are only as good as the predictive model developed in the training phase. We accomplished further experiments just to make sure that the predictive model is sufficiently accurate in assigning appropriate classes for new unlabeled data records. Our approach was to rely on human reviews. We fed the proposed word2vec predictive model with new unlabeled data records and gave a random subset of the system output five hundred system-labeled instances to two domain experts to review. We got 87.6%, 86.1%, and 88.7% in average for accuracy, precision, and recall, respectively.

We briefly summarized the experimental results as follows:

The results we obtained showed that the use of combined dataset was better than the use of either source individually (Table 1) for all the models and is statistically significant (at P=.04).

The results illustrated in the Multimedia Appendix 2 show that a greater *epoch* along with a greater *iteration* with the use of *window size* of two tend to be useful over all datasets using bigNN system, and the result is statistically significant (at P=.02). However, it requires a longer training time.

The comparative study shown in Table 2 demonstrates that the bigNN system was able to generate better results in comparison with traditional BoW along with SVM, decision tree, and naïve Bayes classification algorithms. Performing a t test on AUC matched by those models shows statistically significant differences (at P=.03) between our bigNN system and those two models utilizing BoW along with decision tree and naïve Bayes. It also shows no statistically differences employing BoW and SVM.

Table 2. The comparisons of our big data neutral network (bigNN) system with traditional bag-of-words (BoW) method using support vector machine (SVM), decision tree, and naïve Bayes classifiers.

Dataset ID	Learning method	Number of sentences	Accuracy (%)	Precision (%)	Recall (%)	Area under the receiver operating characteristic	Training time (min)
ADEs#1_Combined[a]	bigNN[b] system	7360	88.7	88.5	89.4	0.842	45.7
ADEs#1_Combined	BoW[c] + SVM[d]	7360	89.4	88.3	88.0	0.841	66.3
ADEs#1_Combined	BoW + decision tree	7360	84.0	83.7	82.1	0.775	49.5
ADEs#1_Combined	BoW + naïve Bayes	7360	83.7	82.1	83.5	0.763	48.9
ADEs#2_Combined	bigNN system	14,017	89.1	88.9	89.3	0.874	69.5
ADEs#2_Combined	BoW + SVM	14,017	89.5	88.0	89.7	0.875	88.9
ADEs#2_Combined	BoW + decision tree	14,017	85.5	84.9	84.5	0.861	75.2
ADEs#2_Combined	BoW + naïve Bayes	14,017	84.3	84.0	85.7	0.855	73.8
ADEs#3_Combined	bigNN system	21,843	92.7	93.6	93.0	0.905	121.7
ADEs#3_Combined	BoW + SVM	21,843	92.5	94.0	93.2	0.911	159.5
ADEs#3_Combined	BoW + decision tree	21,843	88.3	87.5	87.2	0.868	131.5
ADEs#3_Combined	BoW + naïve Bayes	21,843	87.5	86.2	85.8	0.851	135.3

[a]ADEs: adverse drug events.

[b]bigNN: big data neutral network.

[c]BoW: bag-of-words.

[d]SVM: support vector machine.

Figure 4. This figure shows the area under the curve (AUC) of our proposed predictive model. ADEs: adverse drug events.

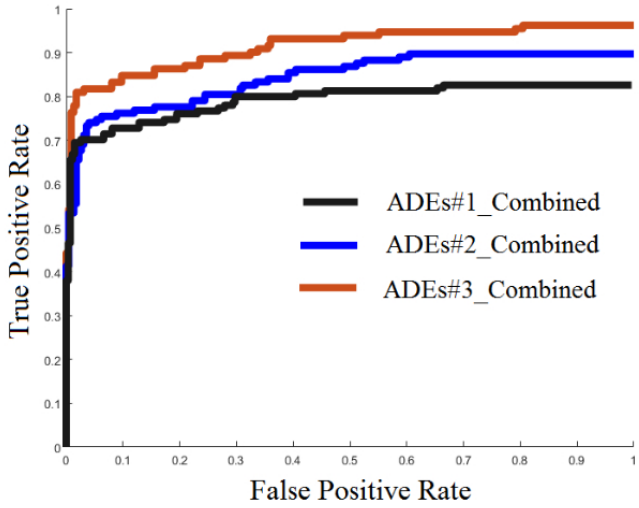


Table 2 also shows that our bigNN system was faster than the BoW method along with SVM, decision tree, and naïve Bayes classification algorithms. Performing a *t* test on training time matched by those methods presents statistically significant differences (at *P*=.03) between our proposed model and all those three models developed by BoW along with SVM, decision tree, and naïve Bayes.

The AUCs presented in Figure 4 show that larger training samples tend to add benefit when making an accurate and reliable predictive model.

To further analyze the results, the next section will present a set of ADEs' scientific visualizations obtained using the proposed framework.

Scientific Visualization of ADEs

Scientific visualization is concerned with representing large and highly dimensional information by means of charts, graphs, and images. The general objective of any scientific visualization is to improve understanding of the data being investigated. In this section, we scientifically visualize the ADE information extracted by our proposed system. Once the system was trained across the different datasets, we fed the system new unlabeled data. This included 92,681,359 sentences from biomedical articles downloaded from PubMed Central and 1,624,117 sentences from social media, including MedHelp, Patient, and WebMD. There are considerable amounts of drugs and adverse events, and it is beyond the scope of this paper to visualize all of them. Here, we limited ourselves to the list of 28 drugs as illustrated in Table 3. The proposed system could find 12,265 ADE sentences from biomedical articles and 181 ADE sentences from three social media sources using the drug list defined.

Figures 5 and 6 show ADE discovery visualization results obtained from the biomedical articles and social media, respectively. Figure 7 represents a set of word cloud examples generated for ADEs associated with "aspirin," "atenolol," "gabapentin," and "statins." A word cloud is a graphical representation composed of different words contained in a corpus, in which the size of every word indicates its frequency or importance to the text.

Table 3. The list of the drugs used to make the scientific visualization results.

Row ID	Drug name
1	Anesthesia
2	Antihistamine
3	Antipsychotic
4	Aspirin
5	Atenolol
6	Atorvastatin
7	Azithromycin
8	Dexamethasone
9	Diazepam
10	Dopamine
11	Ephedrine
12	Gabapentin
13	Galantamine
14	Heparin
15	Ibuprofen
16	Lamotrigine
17	Lorazepam
18	Melatonin
19	Meloxicam
20	Metformin
21	Methylphenidate
22	Ondansetron
23	Orlistat
24	Sildenafil
25	Statins
26	Vioxx
27	Warfarin
28	Wellbutrin

Figure 5. The adverse drug events' (ADEs') visualization results obtained from biomedical articles. One can see the number of "anesthesia" observations is 2186, where the most frequent adverse drug events are "hypotension," "nausea," "aspiration," and "depression," respectively. Using the CI of 95%, Pr(hypotension|anesthesia) is between 20.5% and 23.5%, and Pr(nausea|anesthesia) is some point between 12.0% and 14.7%. Another example is "gabapentin" where the most frequent adverse drug events based on ADE sentences extracted from biomedical articles are "dizziness," "nausea," "fatigue," and "edema" whereas the number of "gabapentin" is 261. Using the CI of 95%, Pr(dizziness|gabapentin) is between 33.0% and 44.2%.

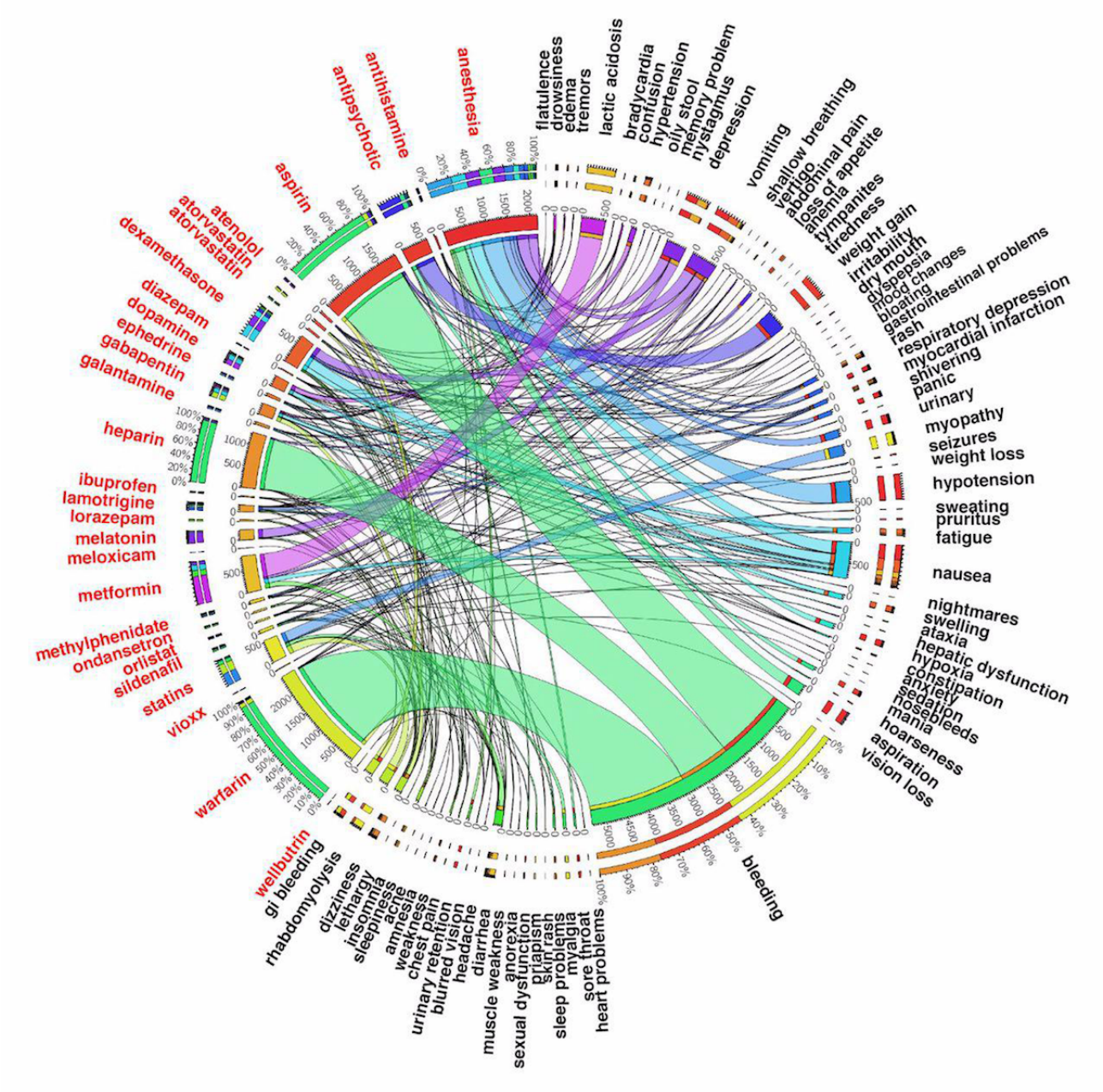


Figure 6. The adverse drug events' (ADEs') visualization results obtained from health-related social media. One can see the number of "metformin" observations is 28, where its most frequent ADEs are "nausea," "diarrhea," "vomiting," "dizziness," and "stomach pain," respectively. Another example is "atenolol" and its most frequent adverse events are "depression," "bradycardia," "hypotension," "tiredness," and "dizziness." From the adverse events aspect, this figure shows, for example, "nausea" as an adverse event is mostly associated with "metformin," "dexamethasone," "antihistamine," "wellbutrin," and "sildenafil".

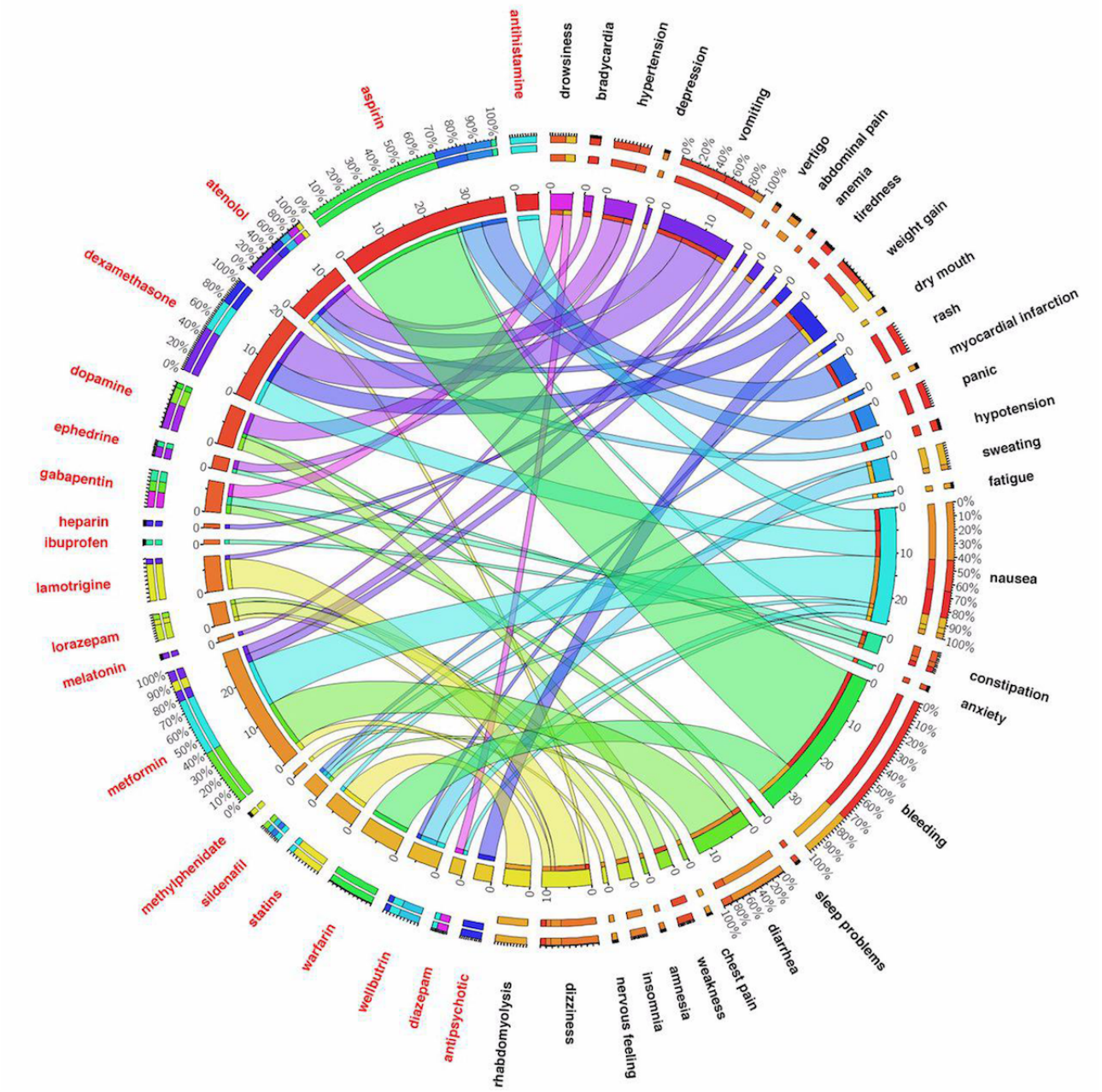
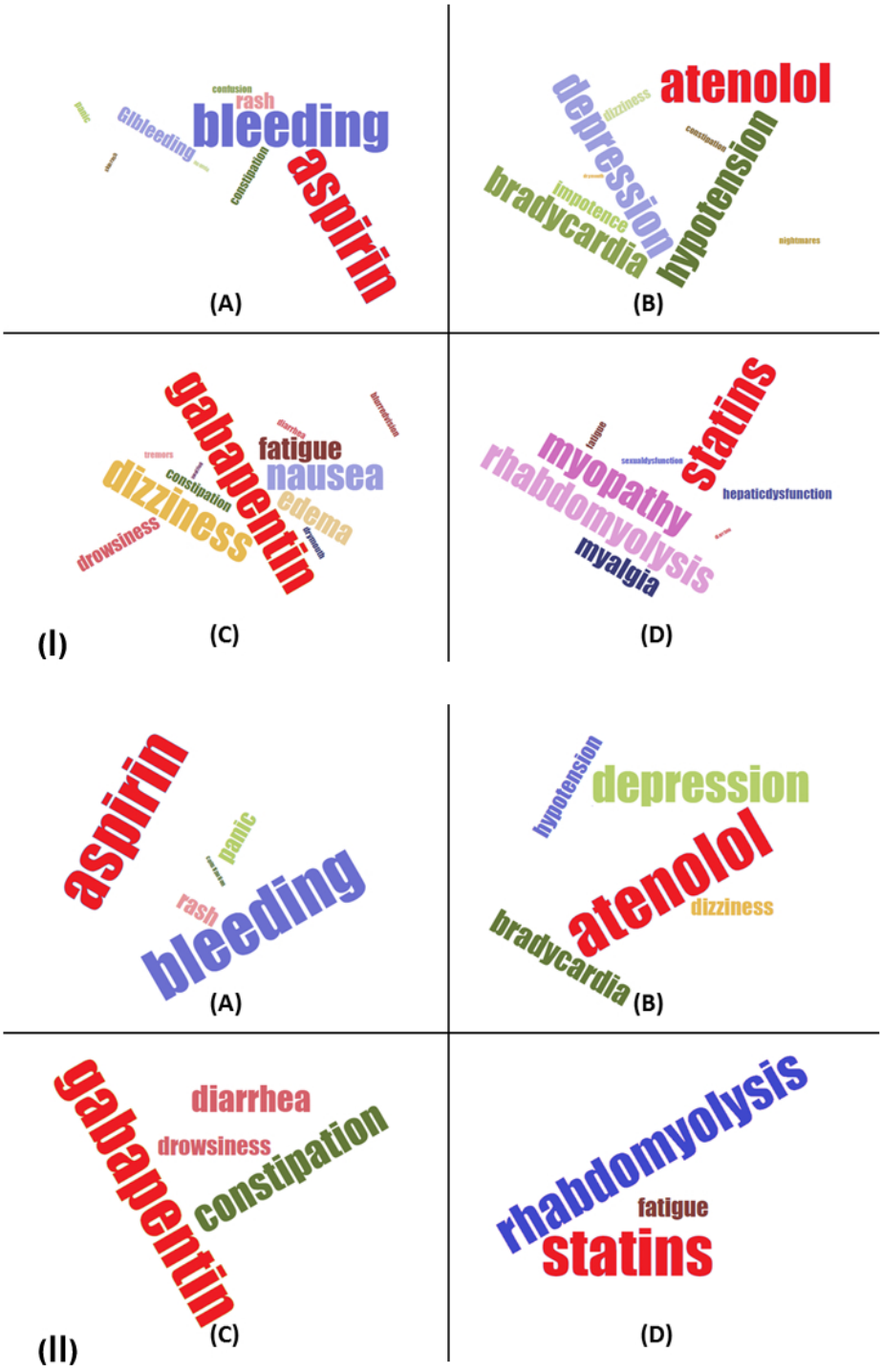


Figure 7. The word cloud representations for “aspirin,” “atenolol,” “gabapentin,” and “statins.” (1) These results are obtained from biomedical articles. Whereas the most frequent adverse dug events for “statins” are “myopathy,” “rhabdomyolysis,” “myalgia,” “fatigue,” and “hepatic dysfunction” as shown in (D), the most frequent ADEs for “atenolol” are “depression,” “hypotension,” “bradycardia,” and “impotence,” respectively (B). (2) These results are extracted from social media. Whereas the most frequent ADEs for “statins” are “rhabdomyolysis” and “fatigue” (D), the most frequent ADEs for “atenolol” are “depression,” “bradycardia,” “hypotension,” and “dizziness,” respectively (B).



Discussion

Principal Findings

Modern medical data sources ranging from clinical trials, EHRs, and medical case reports to scientific articles and patients’ blog posts are rapidly growing in size and complexity, and scientific biomedical articles, as well as health-related social media are under-researched data sources for biomedical studies. Thus, there is a pressing need to develop efficient solutions to harness this wealth of data using advanced computational methods such as artificial intelligence and big data machine learning.

With this contribution, an attempt was made to design and develop a computational bigNN system to detect, analyze, and visualize ADEs from massive data sources obtained from PubMed Central, a widely referenced repository of scientific articles, and the social media blog posts existing in MedHelp,

Patient, and WebMD. The model was developed using the word2vec neural network architectural on top of the Apache Hadoop cluster, Apache Spark, and Elasticsearch No-SQL distributed database to tackle efficient big data ADE identification. We accomplished extensive experimental validations to ensure that the proposed predictive model can accurately assign appropriate classes (eg, ADEs or No-ADEs) for both current and also new unlabeled data records. Our trained system was able to detect a number of well-known ADEs from unlabeled data taken directly from the literature. A list of well-known ADEs can be found in Multimedia Appendix 1. Warfarin is an extremely effective anticlotting agent primarily used for patients at high risk for stroke or heart attack because of atrial fibrillation that, up until recently, has been a cornerstone of treatment. One common and potentially serious adverse effect of warfarin therapy is bleeding. This can occur because of changes in diet, drug interactions, or spurious physiological changes. The effectiveness of warfarin coupled with a high risk for serious bleeding events led to extensive research and publication in the medical literature, which is very apparent in our data visualization. Similarly, the widespread use of aspirin, which also carries a risk of bleeding, has been extensively studied for primary prevention of heart attack, colorectal cancer, and for secondary preventions of cardiovascular events, which is also captured in our results. Our system also identifies serious but rare side effects such as lactic acidosis caused by metformin use and rhabdomyolysis attributed to statin therapy. Common drug side effects are also captured, although with a smaller number of hits. Examples in Figure 5 include ADE pairs of metformin and diarrhea, antipsychotics and weight gain, gabapentin and dizziness, and drowsiness with antihistamines.

The findings of our system shed light on areas for future work and on inherent challenges with semantics and context in NLP. The following examples will illustrate some of these challenges. Our system identifies nausea and vomiting as ADEs associated with dexamethasone. Although it is true that dexamethasone can cause these reactions, it is commonly prescribed to prevent nausea and vomiting associated with chemotherapy. Without more contextual clues from the free text in the articles, it is impossible for us to decide whether dexamethasone is being identified as a treatment or a causal agent in this case. Another example where our system identified an indication as an ADE was with lamotrigine and seizures. Seizure is a primary indication for the use of lamotrigine and not a causal agent. Successfully classifying these edge cases may require additional labeled data, a larger window size, or other unexplored techniques and is an area for further study.

The ADEs identified from our analysis of social media provide a number of interesting similarities and differences with those in the literature. In our social media results, as might be expected, we see a larger proportion of ADEs related to the more common side effects of drugs as compared with the literature. For example, we see a large proportion of sentences identified for nausea and diarrhea with the use of metformin and fewer mentions of abdominal pain and vomiting. This parallels nicely with the incidence expected in real-world use. In contrast, we see a high proportion of sentences labeled for lactic acidosis, an extremely rare ADE associated with metformin use from the literature. The number of sentences describing adverse drug events in biomedical text articles is highly variable and includes influencing factors such as the severity of the drug reaction, safety concerns eliciting directed study, and the goals and intent of the research paper. Non-life-threatening ADEs are less important to clinical researchers, assuming they do not result in discontinuation, compared with serious reactions. In a similar way, side effects reported in social media will naturally include more common side effects, particularly because they are impactful to the patient taking the medication. In future work, text mining should take advantage of these naturally occurring differences. Publications in the biomedical literature or postings in social media, especially early after the release of a novel drug, may include case studies or reports of side effects not seen in clinical trials that could be detected by our system before the signal reaches the critical detection threshold of reporting systems such as the FDA Adverse Event Reporting System.

Limitations

We acknowledge some limitations to this research study. Assessing the quality of scientific journals is a difficult task but important for narrowing the search space of candidate articles. In this work, we attempt to combine three ranking indices in the hope of identifying journals with the most credible information without generating a hand-curated list. It may be the case that our approach excludes journals that would be extremely useful in identifying ADEs but are excluded based on a low combined score. In addition, some journals will provide a richer source of information on ADEs than others based on their intended audience and subject matter irrespective of any ranking criterion. This leaves the question of which journals to focus on for ADE text mining open for further exploration. In the scientific articles, as well as social media blog posts, we noticed that different people may use different terms to discuss a similar single adverse event. For example, the terms “mood changes” and “mood swings” are often used interchangeably, equally meaning “mood changes.” A robust dictionary-based methodology may help address this issue. Additionally, text mining of the social media comments posted by patients is a really challenging task, as the comments are often written in an informal way. As we have a smaller number of labeled sentences from this source, we didn’t address overcoming the differences in phrasing, spelling errors, or other problems introduces from these posts. No fuzzy matching or specialized dictionaries were used on either source, so if there are spelling errors in the drug name, adverse event, or indication, the sentence would have been excluded from evaluation. Furthermore, our proposed big data neural network model is more appropriate for the short-length text data (eg, a sentence) classification rather than the long-length text data (eg, full text articles) categorization. Advanced tokenization systems, and in particular, a medical literature–based tokenization system will be useful with short-length text data.

Conclusions

The present contribution utilized a bigNN system to discover only ADEs; however, there are several interesting applications to leverage the proposed system. For example, the proposed big

data analytics pipeline could help study drug repurposing and impact drug development, particularly in analyzing the success rates of new medications. The social media demographic information (eg, age, gender, ethnicity, and location) supports use of the current contribution to explore ADEs and drug indications discussed by different demographic groups. For future work, we intend to further explore the application of the proposed framework to medical informatics, and particularly drug analyses, extending the work for social media–based ADE discovery discussed by different demographics groups. We would enhance the proposed framework to tackle the problem of semisupervised learning with multiple labels, rather than only a single label. We also plan to make a sentence-based ADE discovery dataset and present it publicly and make it freely available to the research community.

Acknowledgments

The authors gratefully acknowledge the support of the Center for Predictive Computational Phenotyping, supported by the National Institute of Health Big Data to Knowledge Initiative, US4AI117924; the University of Wisconsin Institute for Clinical and Translational Research (UL1TR000427) and the support of NIGMS grant 2RO1 GM097618. The authors of the paper wish to thank Richard Berg, Ryan Frahm, and Joseph Ellefson at Marshfield Clinic Research Institute and Mehdi Assefi and Sanaz Haghani at University of Georgia for their valuable contributions in big data analytics and scientific visualizations presented in this work. Special thanks goes to Anne Nikolai at Marshfield Clinic Research Institute for her valuable contributions in manuscript preparation.

Authors' Contributions

PP, DP, and APT conceived and designed the study. PP and DP secured the funding for this contribution. APT, JB, EL, and ES designed and developed the software framework. APT, JB, EL, ES, AM, JM, and ZY conducted the data collection. All authors contributed to the analysis and interpretation of the results and scientific visualization. PP and APT led the writing of this manuscript with all coauthors comments. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Selecting the relevant documents, Building the training set, Word2vec neural network model, Well-known ADEs.

[PDF File (Adobe PDF File), 640KB - medinform_v5i4e51_app1.pdf]

Multimedia Appendix 2

bigNN experimental validations.

[PDF File (Adobe PDF File), 195KB - medinform_v5i4e51_app2.pdf]

References

1. Gottlieb A, Hoehndorf R, Dumontier M, Altman RB. Ranking adverse drug reactions with crowdsourcing. J Med Internet Res 2015;17(3):e80 [FREE Full text] [doi: 10.2196/jmir.3962] [Medline: 25800813]
2. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. J Am Med Assoc 1995 Jul 05;274(1):29-34. [Medline: 7791255]
3. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. J Am Med Assoc 1997;277(4):301-306. [Medline: 9002492]
4. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. J Am Med Assoc 1998 Apr 15;279(15):1200-1205. [Medline: 9555760]
5. Trame MN, Biliouris K, Lesko LJ, Mettetal JT. Systems pharmacology to predict drug safety in drug development. Eur J Pharm Sci 2016 Oct 30;94:93-95. [doi: 10.1016/j.ejps.2016.05.027] [Medline: 27251780]
6. Liu J, Zhao S, Zhang X. An ensemble method for extracting adverse drug events from social media. Artif Intell Med 2016 Jun;70:62-76. [doi: 10.1016/j.artmed.2016.05.004] [Medline: 27431037]
7. Price J. What can big data offer the pharmacovigilance of orphan drugs? Clin Ther 2016 Dec;38(12):2533-2545. [doi: 10.1016/j.clinthera.2016.11.009] [Medline: 27914633]
8. van Walsem A, Pandhi S, Nixon RM, Guyot P, Karabis A, Moore RA. Relative benefit-risk comparing diclofenac to other traditional non-steroidal anti-inflammatory drugs and cyclooxygenase-2 inhibitors in patients with osteoarthritis or rheumatoid

arthritis: a network meta-analysis. Arthritis Res Ther 2015 Mar 19;17:66 [FREE Full text] [doi: 10.1186/s13075-015-0554-0] [Medline: 25879879]
9. Pignatti F, Ashby D, Brass EP, Eichler H, Frey P, Hillege HL, et al. Structured frameworks to increase the transparency of the assessment of benefits and risks of medicines: current status and possible future directions. Clin Pharmacol Ther 2015 Nov;98(5):522-533. [doi: 10.1002/cpt.203] [Medline: 26261064]
10. Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. Drug Saf 2014 Oct;37(10):777-790 [FREE Full text] [doi: 10.1007/s40264-014-0218-z] [Medline: 25151493]
11. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. Trials 2012 Aug 20;13:138 [FREE Full text] [doi: 10.1186/1745-6215-13-138] [Medline: 22906139]
12. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. Value Health 2010 Aug;13(5):657-666 [FREE Full text] [doi: 10.1111/j.1524-4733.2010.00725.x] [Medline: 20412543]
13. Eichler H, Abadie E, Raine JM, Salmonson T. Safe drugs and the cost of good intentions. N Engl J Med 2009 Apr 02;360(14):1378-1380. [doi: 10.1056/NEJMp0900092] [Medline: 19339718]
14. Garrison LP, Towse A, Bresnahan BW. Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. Health Aff (Millwood) 2007;26(3):684-695 [FREE Full text] [doi: 10.1377/hlthaff.26.3.684] [Medline: 17485745]
15. Jonnagaddala J, Jue T, Dai H. Binary classification of Twitter posts for adverse drug reactions. 2016 Jan Presented at: In Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing; 2016; Big Island, HI. USA p. 4-8.
16. Rastegar-Mojarad M, Elayavilli R, Wang L, Prasad R, Liu H. Prioritizing Adverse Drug Reaction and Drug Repositioning Candidates Generated by Literature-Based Discovery. In: Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2016 Presented at: BCB '16; October 02 - 05, 2016; Seattle, WA, USA p. 289-296. [doi: 10.1145/2975167.2975197]
17. Yang M, Kiang M, Shang W. Filtering big data from social media--building an early warning system for adverse drug reactions. J Biomed Inform 2015 Apr;54:230-240 [FREE Full text] [doi: 10.1016/j.jbi.2015.01.011] [Medline: 25688695]
18. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. J Biomed Inform 2015 Feb;53:196-207 [FREE Full text] [doi: 10.1016/j.jbi.2014.11.002] [Medline: 25451103]
19. Liu X, Chen H. AZDrugMiner: an information extraction system for mining patient-reported adverse drug events in online patient forums. 2013 Presented at: International Conference on Smart Health; August 3, 2013; Beijing, China p. 134-150.
20. Bian J, Topaloglu U, Yu F. Towards large-scale twitter mining for drug-related adverse events. In: Proceedings of the 2012 international workshop on Smart health and wellbeing. 2012 Oct 29 Presented at: SHB '12; October 29 - 29, 2012; Maui, Hawaii, USA p. 25-32 URL: http://europepmc.org/abstract/MED/28967001 [doi: 10.1145/2389707.2389713]
21. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. AMIA Annu Symp Proc 2011;2011:217-226 [FREE Full text] [Medline: 22195073]
22. Paul M, Dredze M. Drug Extraction from the Web: Summarizing Drug Experiences with Multi-Dimensional Topic Models. 2013 Presented at: Proceedings of NAACL-HLT 2013; 9-14 June, 2013; Atlanta, Georgia p. 168-178.
23. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med 2012 Mar 14;4(125):125ra31 [FREE Full text] [doi: 10.1126/scitranslmed.3003377] [Medline: 22422992]
24. Liang Z, Huang JX, Zeng X, Zhang G. DL-ADR: a novel deep learning model for classifying genomic variants into adverse drug reactions. BMC Med Genomics 2016 Aug 10;9(Suppl 2):48 [FREE Full text] [doi: 10.1186/s12920-016-0207-4] [Medline: 27510822]
25. Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. J Am Med Inform Assoc 2014;21(2):353-362 [FREE Full text] [doi: 10.1136/amiajnl-2013-001612] [Medline: 24158091]
26. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther 2012 Jun;91(6):1010-1021 [FREE Full text] [doi: 10.1038/clpt.2012.50] [Medline: 22549283]
27. Gurulingappa H, Mateen-Rajput A, Toldo L. Extraction of potential adverse drug events from medical case reports. J Biomed Semantics 2012;3(1):15 [FREE Full text] [doi: 10.1186/2041-1480-3-15] [Medline: 23256479]
28. Aramaki E, Miura Y, Tonoike M, Ohkuma T, Masuichi H, Waki K, et al. Extraction of adverse drug effects from clinical records. Stud Health Technol Inform 2010;160(Pt 1):739-743. [Medline: 20841784]
29. NCBI. PubMed URL: https://www.ncbi.nlm.nih.gov/pubmed [accessed 2017-11-29] [WebCite Cache ID 6vLTyklNq]
30. Naples JG, Hanlon JT, Schmader KE, Semla TP. Recent literature on medication errors and adverse drug events in older adults. J Am Geriatr Soc 2016 Feb;64(2):401-408 [FREE Full text] [doi: 10.1111/jgs.13922] [Medline: 26804210]
31. Noguchi C, Sakuma M, Ohta Y, Bates DW, Morimoto T. Prevention of medication errors in hospitalized patients: the Japan adverse drug events study. Drug Saf 2016 Nov;39(11):1129-1137. [doi: 10.1007/s40264-016-0458-1] [Medline: 27638660]
32. Meier F, Maas R, Sonst A, Patapovas A, Müller F, Plank-Kiegele B, et al. Adverse drug events in patients admitted to an emergency department: an analysis of direct costs. Pharmacoepidemiol Drug Saf 2015 Feb;24(2):176-186. [doi: 10.1002/pds.3663] [Medline: 24934134]

33. Nuckols TK, Smith-Spangler C, Morton SC, Asch SM, Patel VM, Anderson LJ, et al. The effectiveness of computerized order entry at reducing preventable adverse drug events and medication errors in hospital settings: a systematic review and meta-analysis. Syst Rev 2014 Jun 04;3:56 [FREE Full text] [doi: 10.1186/2046-4053-3-56] [Medline: 24894078]
34. Kanaan AO, Donovan JL, Duchin NP, Field TS, Tjia J, Cutrona SL, et al. Adverse drug events after hospital discharge in older adults: types, severity, and involvement of Beers Criteria Medications. J Am Geriatr Soc 2013 Nov;61(11):1894-1899 [FREE Full text] [doi: 10.1111/jgs.12504] [Medline: 24116689]
35. Dormann H, Sonst A, Müller F, Vogler R, Patapovas A, Pfistermeister B, et al. Adverse drug events in older patients admitted as an emergency: the role of potentially inappropriate medication in elderly people (PRISCUS). Dtsch Arztebl Int 2013 Mar;110(13):213-219 [FREE Full text] [doi: 10.3238/arztebl.2013.0213] [Medline: 23596501]
36. Finkelstein J, Chen Q, Adams H, Friedman C. Automated summarization of publications associated with adverse drug reactions from PubMed. AMIA Jt Summits Transl Sci Proc 2016;2016:68-77 [FREE Full text] [Medline: 27570654]
37. Natarajan S, Bangera V, Khot T, Picado J, Wazalwar A, Costa VS, et al. Markov logic networks for adverse drug event extraction from text. Knowl Inf Syst 2017 May;51(2):435-457. [doi: 10.1007/s10115-016-0980-6] [Medline: 29123330]
38. Kang N, Singh B, Bui C, Afzal Z, van Mulligen EM, Kors JA. Knowledge-based extraction of adverse drug events from biomedical text. BMC Bioinformatics 2014 Mar 04;15:64 [FREE Full text] [doi: 10.1186/1471-2105-15-64] [Medline: 24593054]
39. NCBI. PubMed Central URL: https://www.ncbi.nlm.nih.gov/pmc/ [accessed 2017-11-29] [WebCite Cache ID 6vLUJCock]
40. MedHelp. URL: http://medhelp.org/ [accessed 2017-11-29] [WebCite Cache ID 6vLUQL3F0]
41. Patient. URL: https://patient.info/ [accessed 2017-11-29] [WebCite Cache ID 6vLUVsMfF]
42. WebMD. URL: https://www.webmd.com/default.htm [accessed 2017-11-29] [WebCite Cache ID 6vLUcTEob]
43. Scrapy. URL: https://scrapy.org/ [accessed 2017-11-29] [WebCite Cache ID 6vLUjwo7C]
44. Elastic. Elasticsearch URL: https://www.elastic.co/ [accessed 2017-11-29] [WebCite Cache ID 6vLUpqeIv]
45. Henriksson A, Kvist M, Dalianis H, Duneld M. Identifying adverse drug event information in clinical notes with distributional semantic representations of context. J Biomed Inform 2015 Aug 17;57:333-349 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.013] [Medline: 26291578]
46. Sriram B, Fuhry D, Demir E, Ferhatosmanoglu H, Demirbas M. Short text classification in twitter to improve information filtering. 2010 Presented at: The 33rd international ACM SIGIR conference on Research and development in information retrieval; Jul 19, 2010; Geneva, Switzerland p. 841-842.
47. Sivic J, Zisserman A. Efficient visual search of videos cast as text retrieval. IEEE Trans Pattern Anal Mach Intell 2009 Apr;31(4):591-606. [doi: 10.1109/TPAMI.2008.111] [Medline: 19229077]
48. Sahlgren M, Cöster R. Using bag-of-concepts to improve the performance of support vector machines in text categorization. 2004 Presented at: the 20th International Conference on Computational Linguistics; August 23, 2004; Geneva, Switzerland.
49. Lodhi H, Saunders C, Shawe-Taylor J, Cristianini N, Watkins C. Text classification using string kernels. J Mach Learn Res 2002;2:419-444.
50. Pedersen T. A decision tree of bigrams is an accurate predictor of word sense. 2001 Presented at: NAACL '01 Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies; June 01 - 07, 2001; Pittsburgh, Pennsylvania p. 1-8.
51. Scott S, Matwin S. Feature engineering for text classification. 1999 Presented at: ICML '99 Proceedings of the Sixteenth International Conference on Machine Learning; June 27 - 30, 1999; San Francisco, CA, USA p. 379-388.
52. Mikolov T, Chen K, Corrado G, Dean J. arxiv.org. 2013. Efficient estimation of word representations in vector space URL: https://arxiv.org/pdf/1301.3781.pdf [accessed 2017-11-30] [WebCite Cache ID 6vNQxsjjp]
53. Bansal M, Gimpel K, Livescu K. Tailoring Continuous Word Representations for Dependency Parsing. In ACL (2) 2014 Jun (pp. 809-815) 2014.
54. Dyer C, Ballesteros M, Ling W, Matthews A, Smith N. Transition-based dependency parsing with stack long short-term memory. arXiv preprint arXiv.08075 May 2015:29.
55. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Statnlp.org. 2016. Neural architectures for named entity recognition URL: http://www.statnlp.org/wp-content/uploads/2017/06/nner.pdf [accessed 2017-11-30] [WebCite Cache ID 6vNRB665R]
56. Santos CN, Guimaraes V. Aclweb.org. 2015. Boosting named entity recognition with neural character embeddings URL: http://www.anthology.aclweb.org/W/W15/W15-3904.pdf [accessed 2017-11-30] [WebCite Cache ID 6vNRHxjLX]
57. Lilleberg J, Zhu Y, Zhang Y. Support vector machines and word2vec for text classification with semantic features. 2015 Presented at: Cognitive Informatics & Cognitive Computing (ICCI*CC), 2015 IEEE 14th International Conference on; July 6, 2015; Beijing, China p. 136-140.
58. Zhang X, Zhao J, LeCun Y. Courant Institute of Mathematical Sciences, New York University. 2015. Character-level convolutional networks for text classification URL: https://papers.nips.cc/paper/5782-character-level-convolutional-networks-for-text-classification.pdf [accessed 2017-11-30] [WebCite Cache ID 6vNRQDFcC]
59. Xue B, Fu C, Shaobin Z. A new clustering model based on Word2vec mining on Sina Weibo users' tags. IJGDC 2014;7(3):41-48. [doi: 10.14257/ijgdc.2014.7.3.05]

60. Goldberg Y, Levy O. Arxiv.org. 2014. word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method URL: https://arxiv.org/pdf/1402.3722.pdf [accessed 2017-12-01] [WebCite Cache ID 6vNReVCAM]
61. Rong X. arxiv.org. 2014. word2vec parameter learning explained URL: https://arxiv.org/pdf/1411.2738.pdf [accessed 2017-12-01] [WebCite Cache ID 6vNRhUI42]
62. Ye J. Cosine similarity measures for intuitionistic fuzzy sets and their applications. Math Comput Model 2011;53(1-2):91-97. [doi: 10.1016/j.mcm.2010.07.022]
63. Muflikhah L, Baharudin B. Document clustering using concept space and cosine similarity measurement. 2009 Presented at: Computer Technology and Development, 2009. ICCTD '09. International Conference on; November 13, 2009; Kota Kinabalu, Malaysia p. 58-62.
64. cTAKES: clinical Text Analysis and Knowledge Extraction System. URL: http://ctakes.apache.org/ [accessed 2017-11-29] [WebCite Cache ID 6vLTlneFL]
65. Apache Spark. URL: https://spark.apache.org [accessed 2017-11-29] [WebCite Cache ID 6vLVIS1gd]
66. The University of Waikato. Weka Library URL: https://www.cs.waikato.ac.nz/ml/weka/ [accessed 2017-11-29] [WebCite Cache ID 6vLUwaBCP]
67. Hadoop. URL: http://hadoop.apache.org/ [accessed 2017-11-29] [WebCite Cache ID 6vLVBwzjO]

Abbreviations

ADE: adverse drug event
AUC: area under the curve
BigNN: big data neutral network
BoW: bag-of-words
EHR: electronic health record
EP: epoch
ITR: iteration
MWF: minimum word frequency
NLP: natural language processing
POS: part-of-speech
SVM: support vector machine
WS: window size
XML: extensible markup language

Edited by G Eysenbach; submitted 13.10.17; peer-reviewed by V Hristidis, R Segall, F Rohani; comments to author 03.11.17; revised version received 07.11.17; accepted 08.11.17; published 08.12.17.

Please cite as:
P Tafti A, Badger J, LaRose E, Shirzadi E, Mahnke A, Mayer J, Ye Z, Page D, Peissig P
Adverse Drug Event Discovery Using Biomedical Literature: A Big Data Neural Network Adventure
JMIR Med Inform 2017;5(4):e51
URL: http://medinform.jmir.org/2017/4/e51/
doi:10.2196/medinform.9170
PMID:29222076

Original Paper

Pain Self-Management for Veterans: Development and Pilot Test of a Stage-Based Mobile-Optimized Intervention

Sara S Johnson[1], PhD; Deborah A Levesque[1], PhD; Lynne E Broderick[1], MPH; Dustin G Bailey[1], BA; Robert D Kerns[2], PhD

[1]Pro-Change Behavior Systems, Inc, South Kingstown, RI, United States

[2]VA Connecticut Healthcare System, Yale University, Research Psychologist, Pain Research, Informatics, Multimorbidities and Education (PRIME) Center, Professor of Psychiatry, Neurology and Psychology, West Haven, CT, United States

Corresponding Author:
Sara S Johnson, PhD
Pro-Change Behavior Systems, Inc
1174 Kingstown Road
Suite 101
South Kingstown, RI, 02879
United States
Phone: 1 2032080481
Fax: 1 2032080481
Email: sjohnson@prochange.com

Abstract

Background: Chronic pain is a significant public health burden affecting more Americans than cardiovascular disease, diabetes, and cancer combined. Veterans are disproportionately affected by chronic pain. Among previously deployed soldiers and veterans, the prevalence of chronic pain is estimated between 44% and 60%.

Objective: The objective of this research was to develop and pilot-test *Health eRide: Your Journey to Managing Pain*, a mobile pain self-management program for chronic musculoskeletal pain for veterans. Based on the transtheoretical model of behavior change, the intervention is tailored to veterans' stage of change for adopting healthy strategies for pain self-management and their preferred strategies. It also addresses stress management and healthy sleep, two components of promising integrated treatments for veterans with pain and co-occurring conditions, including posttraumatic stress disorder (PTSD) and traumatic brain injury. In addition, Health eRide leverages gaming principles, text messaging (short message service, SMS), and social networking to increase engagement and retention.

Methods: Pilot test participants were 69 veterans recruited in-person and by mail at a Veterans Health Administration facility, by community outreach, and by a Web-based survey company. Participants completed a mobile-delivered baseline assessment and Health eRide intervention session. During the next 30 days, they had access to a Personal Activity Center with additional stage-matched activities and information and had the option of receiving tailored text messages. Pre-post assessments, administered at baseline and the 30-day follow-up, included measures of pain, pain impact, use of pain self-management strategies, PTSD, and percentage in the Action or Maintenance stage for adopting pain self-management, managing stress, and practicing healthy sleep habits. Global impressions of change and program acceptability and usability were also assessed at follow-up.

Results: Among the 44 veterans who completed the 30-day post assessment, there were statistically significant pre-post reductions in pain (P<.001) and pain impact (P<.001); there was some reduction in symptoms of PTSD (P=.05). There were significant pre-post increases in the percentage of participants in the Action or Maintenance stage for adopting pain self-management (P=.01) and for managing stress (P<.001) but not for practicing healthy sleep habits (P=.11). The global impressions of change measure showed that a majority had experienced some level of improvement. User ratings of acceptability were quite high; ratings of usability fell slightly below the mean for digital programs.

Conclusions: Preliminary data demonstrate the potential impact of the Health eRide program for chronic musculoskeletal pain for veterans. The results underscore that simultaneously addressing other behaviors may be a promising approach to managing pain and comorbid conditions. Additional formative research is required to complete development of the Health eRide program and to address areas of usability requiring improvement. A randomized trial with longer follow-up is needed to demonstrate the program's long-term effects on pain and pain self-management.

(JMIR Med Inform 2017;5(4):e40) doi:10.2196/medinform.7117

KEYWORDS

pain management; self-management; mobile health; mhealth

Introduction

Pain and Pain Self-Management

Chronic pain is a significant public health burden affecting more Americans than cardiovascular disease, diabetes, and cancer combined [1]. The economic toll of chronic pain is approximately US $635 billion annually. Veterans are disproportionately affected by chronic pain [2,3]. The prevalence of chronic pain among previously deployed soldiers and veterans is estimated between 44% and 60% [4,5], compared with 26% in a primary care sample [6]. Among veterans, pain is the most costly of all disorders treated in the Veterans Health Administration (VHA) facilities [7]. Chronic pain is particularly common among the veterans of Operations Iraqi Freedom (OIF), Enduring Freedom (OEF), and New Dawn. Furthermore, the co-occurrence of pain and posttraumatic stress disorder (PTSD), traumatic brain injury (TBI), and all three conditions (postdeployment multisymptom disorder, or PMD) is well documented [3,8,9] and can complicate and reduce the effectiveness of treatment of pain [10-12]. The proponents of integrated treatment for PMD or the co-occurrence of pain with either PTSD or TBI are advocating for innovative delivery of interventions that can address multiple conditions [8].

The ongoing personal, social, and economic burden of pain indicates that existing treatment approaches are insufficient. In addition, there is growing concern about the reliance on chronic opioid therapy for chronic pain [13], with mounting data questioning its efficacy and safety [14-18], particularly for veterans [19]. The 2011 Institute of Medicine Blueprint for Relieving Pain in America calls for a population-level pain management strategy; the promotion of self-management; reducing disparities among vulnerable subgroups; and the tailoring of pain care for each patient [1]. The need to increase the quality, variety, and accessibility of nondrug, evidence-based pain self-management skills is even more urgent for veterans, given that they are also disproportionately affected by the current opioid crisis in the United States [2,5].

There are numerous barriers to pain treatment for veterans—such as limited availability of therapists adequately trained in pain self-management [20], cost [20-23], and the distance or logistics of traveling to appointments [20,21]. Pain treatment is further hindered by limited or inadequate individual tailoring of treatment and an overreliance on ineffective and potentially risky treatments, including the use of opioid analgesics and surgical procedures [24]. Thus, veterans with chronic pain are at risk for a lifetime of increasingly progressive disability. The costs of that disability and its treatment could approach US $5 trillion [3].

Reviews have consistently demonstrated the effectiveness of exercise [25-27] and cognitive behavioral therapy (CBT) [28,29] for the treatment of pain. CBT encourages the use of cognitive (eg, coping self-statements) and behavioral (eg, activity pacing) pain coping skills. Interventions that increase reliance on those skills and adopt a biopsychosocial approach that acknowledges that biological, psychological, and social factors influence how pain is experienced and managed can significantly reduce pain, disability, and depressive symptoms [30].

Mobile technologies offer a promising approach to delivering pain self-management treatments incorporating CBT principles. Mobile-delivered interventions can reduce barriers related to access to treatment; they are convenient, enable a high degree of individual tailoring, and can be delivered with fidelity. At least 89% of adults in the United States have access to the Internet [31], and 79% own a smartphone [32]. Furthermore, among groups with historically less Internet access, the digital divide is shrinking. Whereas 44.0% of a sample of 266 veterans aged 65 years and older reported not having access to the Internet at home, nearly 50% had at least one close social tie whom they could ask to use a device, and 70% had at least one social tie whom they would ask for help accessing the Internet [33].

Research assessing the efficacy of mobile or Web-based pain self-management interventions or apps show, on average, positive preliminary results for pain severity, coping self-statements, and other outcomes [34,35]. However, a major problem with existing interventions is that they tend to neglect individual differences in motivation and readiness to adopt self-management strategies [36], have limited input from end users in the development calling into question their usability [37], fail to address other comorbid conditions [38], and are not based on evidence-based practices [39]. Another limitation is that no veteran-specific intervention could be identified.

Although mobile apps that promote self-management have the potential to speed the adoption of individualized, evidence-based, biopsychosocial treatments for pain [40], those developed to date have largely failed to deliver on that promise. A review of 195 mobile phone apps for pain management found serious limitations in those currently available: only 3% incorporated any evidence-based guidelines or principles from CBT [39]. None have been tested in rigorous clinical trials [39,40], and none developed specifically for veterans could be identified.

The primary objective of this research was to develop and conduct a pilot test of a theoretically grounded, mobile-optimized, Internet-based, interactive pain self-management program for veterans with chronic musculoskeletal pain. The program titled *Health eRide: Your Journey to Managing Pain* was designed to address the limitations of existing apps for pain self-management. The Health eRide intervention, developed specifically for veterans (1) relies on a participatory approach to design, eliciting veterans' input and feedback at each stage of the intervention's development; (2) integrates evidence-based practices for pain self-management; (3) is tailored to end users' readiness to adopt those best practices; and (4) helps to address

two comorbid conditions—PTSD and TBI—by including health behavior change messages that promote two core elements of promising integrated treatment for PMD: stress management and adoption of healthy sleep practices. In addition, the intervention leverages SMS text messaging (short messaging service, SMS), social networking, and gaming principles to increase engagement and retention. The pilot study reported here was conducted as a preliminary test of the program's potential impact on pain and other key outcomes among veterans experiencing pain.

Health eRide Intervention

Intervention Development

Intervention development was guided by the VHA's National Pain Management Strategy's recommendation to focus on innovative patient education programs, deliver cost-effective pain care, increase satisfaction with pain care, and ensure that veterans' needs are addressed [41]. It was also decided at the outset that the intervention would be tailored to veterans' readiness to self-manage pain, as well as their preferences regarding specific pain self-management strategies. The transtheoretical model (TTM) provided the theoretical framework. The TTM explains how individuals progress through a series of five stages of change: precontemplation (not intending to take action); contemplation (intending to take action in the next 6 months); preparation (intending to take action in the next 30 days); action (made the behavior change less than 6 months ago); or maintenance (made the behavior change more than 6 months ago) [42]. The other constructs by TTM—decisional balance, self-efficacy, and processes of change—are systematically related to stages in predictable ways [43-45]. The relationship between stage and these behavior change constructs provide an evidence-based framework for developing and delivering tailored feedback that is more likely to be remembered [46,47]; to be discussed with others [48]; to be considered personally relevant, interesting, and credible [48-50]; and to change behavior [48-50]. TTM-based interventions have been found effective across dozens of behaviors and populations [49,51,52].

Using a participatory design process, formative research elicited input from a panel of veteran advisors, experts, and end users to ensure that the program was perceived as meaningful, understandable, and useful; that its flow was easy to navigate and engaging; that the look and feel were attractive; and that the content was tailored to the veteran culture. Furthermore, input was sought on the most effective manner in which to integrate social networking, principles of gamification, and SMS text messaging. Throughout the development process, end-user interviews, focus groups, and usability testing were conducted to ensure that the program was accessible and acceptable. The content was written in plain language at a 7th grade reading level or less, and all content was reviewed with Health Literacy Advisor software distributed by Health Literacy Innovations, LLC.

Principles of Gamification

Efforts were made to maximize engagement and satisfaction with Health eRide by incorporating principles of gamification. The literature [53,54] and gamification experts stressed that gamification tactics must activate meaning, mastery, and autonomy to be effective. To increase meaning and personal relevance, the opening screens of the program ask users to identify their most important reason for managing pain. The options in the list (eg, get back to activities I love, feel more in control) had been generated by interview and focus group participants and veteran advisors. Users also have the option of uploading an image of their reason (eg, a picture of their children). Users are also asked to select an avatar to represent them throughout the program. They can select an avatar from a list provided or upload an image of their own.

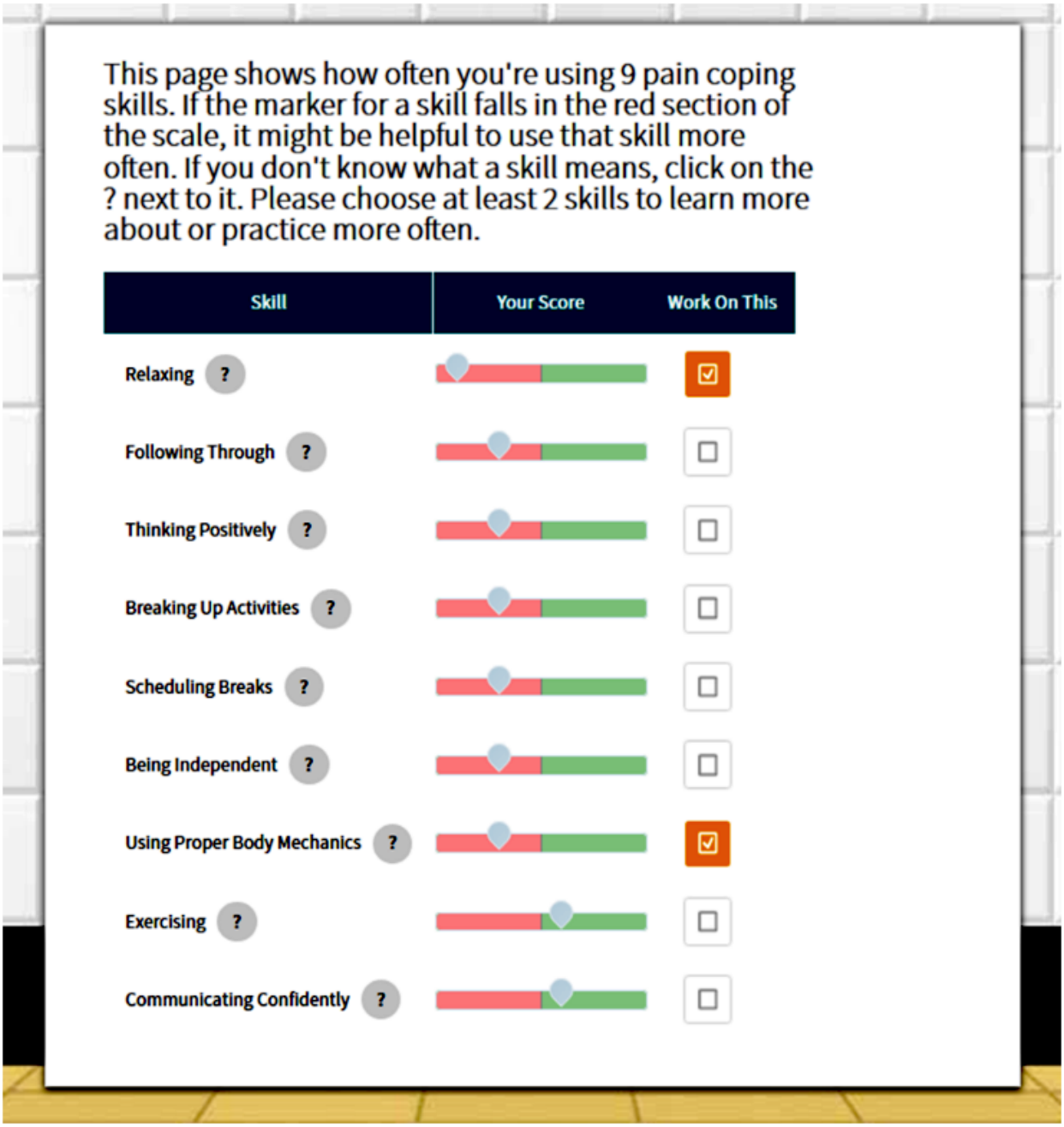
Mastery, which is derived from a sense of progressing to a goal or achieving something, was promoted in several ways. The Health eRide Personal Activity Center (PAC), described below, is structured as a subway map that the user must navigate to reach their final destination (ie, their main reason for managing their pain). Once an activity is completed, additional "stations" (ie, activities) become available, enabling the user to proceed closer to their final destination. The avatar moves down the subway map, and the user's progress is reflected by the accumulation of tickets in a ticket kiosk.

Program Flow

After inviting program users to select a primary reason for managing pain and to select an avatar, the Health eRide program delivers assessments of pain and stage of change for pain self-management, along with the Multidimensional Pain Readiness to Change Questionnaire (MPRCQ2) [55-57], which assesses readiness to use each of the nine strategies for pain self-management. Users receive feedback on their stage of change and a "report card" showing how often they use each of the coping strategies assessed in the MPRCQ2 (see Figure 1). They are asked to select at least two pain self-management strategies they would like to learn more about or practice more often. The program then administers TTM measures of decisional balance and self-efficacy for pain self-management as well as stage-matched guidance designed to facilitate progress to the next stage for using healthy strategies for pain self-management or to prevent relapse to an earlier stage.

In the second half of the session, participants receive brief assessments and stage-matched guidance targeting stress management and healthy sleep habits. Given the frequent co-occurrence of chronic pain and other conditions, especially PTSD [58] and TBI [59], additional assessments are administered to detect possible symptoms of these conditions. Participants screening positive for PTSD or TBI receive information on local and national resources.

Figure 1. Sample pain coping skills feedback from the Health eRide program.



Personal Activity Center

Once users complete the computer tailored intervention (CTI) session, they are brought to their PAC, also known as the Health eRide subway station. The subway station is a collection of 56 interactive activities designed to activate the processes of change that are most appropriate for the user based on his or her stage of change for each behavior. Users began with a "tour" of the station to highlight its features, including the participant's final destination—his or her most important reason for managing pain, identified at the beginning of the CTI session. The tour is designed to acclimate users to the program's principles of gamification, including unlocking new stations by completing activities and collecting "tickets" as they make stops at each station. Tickets are also used as an incentive for users to explore different stations, as they can collect additional "punches" on the tickets when they make extra stops (ie, complete additional interactive activities) in the stations.

Text Messages

At the beginning of the CTI session, users are presented with the opportunity to opt in to receive tailored text messages for each of the three targeted behaviors. Those opting in receive a text message asking them to "validate" their phone number to initiate the messages. Text message content and delivery schedules are matched to the stage of change for each behavior. Sample text messages include:

> *As a Veteran, you likely know many people who have or had pain. Think about one of them who could inspire you to manage your pain.*

Want to be more alert, make better decisions & fewer mistakes? Get a good night's sleep. It's not optional. #zzzs

Stress can make people more prone to pain. If you lower your stress, you can help lower your pain. See PAC activity Get the Facts [short-url].

Social Networking

Although full Facebook integration was not feasible for this prototype, users had access to a Health eRide Facebook page, which was regularly updated by the project team with relevant posts and content. Each screen of the Health eRide program, including the subway station PAC, included a link to the Facebook page. In one of the subway stop activities, Share Your Success Story, users are also presented with the opportunity to share their own story on the Facebook page.

Intervention Pilot Test

The remainder of this report describes a pilot test designed to assess the potential impact of the Health eRide program and its usability and acceptability among a small sample of veterans. In the pilot, eligible participants completed a Health eRide CTI session that included several study measures (eg, measures of pain and stage of change); additional study measures (demographics and military history) were appended to the end of the session. During the next 30 days, participants had access to a PAC with additional stage-matched activities and information and had the option of receiving tailored text messages. Follow-up assessments were administered 30 days following the CTI session.

Methods

Recruitment

Pilot test participants were 69 veterans not involved in the formative research. Pilot participants were recruited through in-person and mail recruitment at the Veterans Administration Connecticut Healthcare System (VACHS), community outreach and Facebook, and a Web-based survey company. Eligibility criteria included the following: age of 18 years or older; veteran status; having a chronic musculoskeletal pain rating of 4 or higher on a 0 to 10 numerical scale of pain intensity [60]; having had pain for more than 3 months; and not currently undergoing treatment with a psychologist, psychiatrist, or other mental health professional for a condition such as bipolar disorder, anxiety, or substance abuse.

In-Person and Mail Recruitment at the VACHS (n=29)

At the outset, a research assistant worked with pain clinic staff, nurses, and physicians at VACHS to identify potential participants and to promote the study at a community outreach table. In addition, a research assistant recruited potential participants from primary care waiting rooms. In both cases, the research assistant screened for eligibility and eligible participants were provided with a program link, user ID, and temporary password. Participants had the option of completing the baseline assessment and CTI session at the VA, on an iPad (Apple Inc) provided, or at home. were asked to call the VA facility during business hours to complete a phone screening with the research assistant. Eligible participants were provided with the program link and log-in credentials.

Community Outreach and Facebook (n=9)

The project team provided flyers to the local Operation Iraqi Freedom (OIF) and Operation Enduring Freedom (OEF) coordinator within the VA, as well as to local university and college veteran representatives; invited Veteran organizations to share recruitment information with the veterans they serve; hung flyers in grocery stores, coffee shops, veterans of foreign wars halls, and other settings; and reached out to personal contacts. Veterans who were interested in participating were asked to call the VA facility during business hours to complete a phone screening with the research assistant. Eligible participants were provided with a program URL and log-in credentials.

Other outreach activities included an 8-day national Facebook ad campaign targeting adults aged between 18 and 65, whose Facebook interests matched keywords, including Iraq and Afghanistan Veterans of America, Wounded Warrior Project, and back pain. Interested Facebook users were linked to an eligibility screener. Although the ad reached 42,811 Facebook users, and 945 of those users clicked through to the eligibility screener, none completed the Web-based eligibility screener.

Web-Based Survey Company (n=31)

The final recruitment channel was a Web-based survey company, Survey Sampling International (SSI). Panel members are individuals who agree to receive invitations matched to personal information they provide. SSI sent email invitations to panel members who had reported that they are veterans. Interested members completed a Web-based screener, and those meeting the eligibility criteria were provided with a link and log-in credentials for the Health eRide Program.

Participant Demographics and Military History

Participants' mean age was 50.3 years (SD 12.0); 81% (56/69) were male; 55% (38/69) were white non-Hispanic, 33% (23/69) black non-Hispanic, 9% (6/69) Hispanic, and 3% (2/69) "other"; 62 % (43/69) were married or cohabiting with a partner, 12% (8/69) were single and never married, and 26% (18/69) were separated, divorced, or widowed; 16% (11/69) had no education beyond high school, 35% (24/69) had attended some college, 41% (28/69) had a college degree, and 9% (6/69) had some postgraduate education. Participants had served an average of 8.7 years (SD 7.1) in the military; rank at discharge was enlisted for 50% (34/68) of the participants, senior enlisted for 44% (30/68) and officer for 6% (4/68). About half (48%, 33/69) reported that they had been deployed to Iraq, Afghanistan, the Gulf, Vietnam, and/or Korea, and 22% (15/69) reported that they had been deployed elsewhere.

Procedure

Participants completed a Health eRide CTI session, which included baseline measures, and were encouraged to complete at least two PAC activities. The PAC remained available for 30 days. For veterans opting to receive text messages, the program also delivered messages for 30 days. Upon completion of their first CTI session, participants received a US $25 gift card or,

for SSI participants, US $25 worth of "points" that they could exchange for rewards. Thirty days post baseline, participants were prompted via email to complete a brief follow-up assessment and acceptability survey. Nonrespondents received a reminder call from the VA facility research assistant. Upon completion of the follow-up assessment, participants received another US $25 incentive.

Measures

Questions assessing demographics, military history (eg, years of service and rank), and TBI [59] were administered at baseline only. Unless otherwise noted, the following measures were administered at baseline and 30-day follow-up.

Pain Intensity

Level of pain was assessed using the widely used 11-point numerical scale of pain intensity [60]. Four versions of the scale asked participants to rate their (1) level of pain right now, (2) usual level of pain in the last week, (3) best level of pain in the last week, and (4) worst level of pain in the last week [61]. All ratings were on a scale of 0 to 10, with 0=no pain and 10=worst pain. Provisional benchmarks for interpreting the clinical significance of change scores on numerical rating scales for pain suggest that reductions of ≥30% appear to reflect at least moderately important improvement.

Pain Impact

The Pain Impact Questionnaire (PIQ-6) [62] is a 6-item measure designed to measure level of pain and the impact of pain on work, leisure activities, and well-being. The measure has high internal consistency (Cronbach alpha=.94) and good convergent and discriminant validity. Weighted scores range from 40 to 78, with higher scores reflecting greater pain impact [62].

Pain Self-Management Skills

Pain self-management skills were assessed using the MPRCQ2 [56], a 26-item version of the 69-item MPRCQ [57]. Similar to the MPRCQ, the MPRCQ2 assesses readiness to use seven adaptive pain coping skills (exercise, task persistence, relaxation, cognitive control, activity pacing, assertive communication, and using proper body mechanics) and to stop using two maladaptive skills (pain contingent rest and asking for assistance). For adaptive skills, response options range from 1=I am not doing this now, and am not interested in ever doing it, to 7=I have been doing this for a long time (at least 6 months). For maladaptive skills, response options range from 1=I am doing this now and am not interested in ever stopping, to 7=I have not done this for a long time (at least 6 months). Two items assess each subscale, with the exception of cognitive control, which has a total of 10 items assessing five types of cognitive control (types of cognitive control were not examined). Scale scores are computed by taking the mean of the items representing each subscale. Other research has shown that MPRCQ2 subscale scores are highly correlated with subscale scores on the original MPRCQ, associated with readiness to change, and sensitive to change that occurs over the course of traditional treatment for pain. Unfortunately, in this study, two MPRCQ2 items were inadvertently omitted from the measure—one item from the cognitive control subscale and the other from the assertive communication subscale. The score for cognitive control is represented by the mean of the remaining 9 items; the score for assertive communication is represented by the score on the remaining single item.

Posttraumatic Stress Disorder

PTSD was measured using the PTSD Checklist—Military Version [63]. This 17-item measure asks how much respondents have been bothered in the past month by each of the 17 *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV) PTSD symptoms related to "stressful military experiences." Response options range from 1=not at all, to 5=extremely. A total symptom severity score (range=17-85) can be obtained by summing the 17 items. A severity score of 50 has been widely recommended as the cut-off suggestive of PTSD [63]. However, more recent research recommends cut scores as low as 31 [64].

Well-Being

Well-being was assessed using the Cantril Self-Anchoring Scale [65], which asks participants to imagine a ladder with steps numbered from 0 to 10, with the top representing the best possible life and the bottom representing the worst possible life, and to indicate where they feel their life falls currently and where it will fall in 5 years.

Stage of Change for Pain Self-Management

The stage of change measure for pain self-management was adapted from an algorithm developed in a previous work on pain self-management for patients with interstitial cystitis [66]. Participants were provided with a list of six effective self-management strategies (eg, exercising regularly, controlling negative thoughts about the pain) and asked about their readiness to use at least three of them to manage their pain. Patients who reported that they had no intention of doing so in the next 6 months were classified in the precontemplation stage; those who intended to do so in the next 6 months or next 30 days were classified in the contemplation or preparation stage, respectively. Those who had been meeting the action criteria for less than 6 months were in the action stage, and those who had been meeting criteria for more than 6 months were in maintenance.

Stage of Change for Stress Management

Readiness to practice stress management was assessed with a staging algorithm used previously to assess outcomes in a randomized trial of a computerized TTM intervention for stress management [67]. The question defines healthy stress management strategies and asks participants if they effectively practice them. [68]. Response options and scoring rules match those used for pain self-management, described above.

Stage of Change for Practicing Healthy Sleep Habits

Readiness to practice healthy sleep habits was assessed using a staging algorithm that provided a list of healthy sleep habits (eg, getting at least 7 hours of sleep a night, maintaining a regular bedtime and wake time, avoiding caffeine, alcohol, nicotine, spicy foods, and heavy meals within 4 hours of bedtime) and asked about the intention to engage in them regularly [69]. Response options and scoring rules match those used for pain self-management, described above.

Global Impressions of Change

The Patient Global Impression of Change Scale, administered only at follow-up, is recommended as a core outcome measure in studies of pain [60]. In this study, the scale included seven categorical responses to measure improvement or aggravation of pain. *Since beginning this program, how would you describe the change (if any) in activity limitations, symptoms, emotions, and overall quality of life related to your painful condition?* Response options ranged from 1=No change (or condition has gotten worse), to 7=A great deal better, and a considerable improvement that has made all the difference.

Program Usability

At follow-up only, program usability was assessed using the System Usability Scale (SUS) [70,71], a 10-item measure recommended by the Department of Health and Human Service usability.gov resource for assessing the usability of digital content [72]. Respondents were asked to score each of the 10 items (eg, "I felt very confident using the system") using responses ranging from 1=strongly agree to 5=strongly disagree. Some items were reversed scored. In this study, Cronbach alpha was .89. SUS items were summed and recalibrated to yield a total score ranging from 0 to 100. Across studies, the average SUS score was 68 [73].

Program Acceptability

Acceptability was assessed using 10 questions adapted from National Cancer Institute's Education Materials Review Form [74]. In this study, questions were positive statements regarding participants' perceptions of the program's appeal, suitability for veterans, and potential to impact change. Response options ranged from 1=strongly disagree, to 4=strongly agree. Cronbach alpha was .92 in this study. An overall acceptability score for Health eRide was computed as the mean of the 10 items. Additional open-ended questions assessed what participants liked most and liked least about the program, and how the program could be improved.

Analysis Plan

The first set of analyses assessed pre-post changes in pain, pain impact, pain coping strategies, PTSD, well-being, and measures of stage of change for pain self-management, stress management, and healthy sleep. Pre-post changes on continuous measures were examined using paired samples tests. Stage measures were dichotomized (pre-Action vs Action or Maintenance), and pre-post changes were examined using the McNemar chi-square test with continuity correction. The McNemar test is used for binary dependent variables in a within-subjects design when the same individuals are measured twice. Measures of effect size—Cohen for the continuous outcomes and odds ratios for the binary outcomes—were also computed. The formula for Cohen used here ([M−M]/SD) does not take into account the correlation between the pre- and postmeasures, yielding a more conservative—and accurate [75]—measure of effect size.

Descriptive statistics were computed for program usability and acceptability measures. It was decided at the outset that the criterion for establishing program usability would be a score >68, the average SUS score across studies [73]. The criterion for establishing program acceptability would be an overall mean acceptability score ≥3.

Results

A total of 69 participants completed an initial study session, which included the CTI and additional study measures. The session lasted an average of 39.3 min (SD 20.0 min). During the next 30 days, 81% (56/69) of the participants completed at least one PAC activity. On average, study participants completed an average of 9.4 PAC activities (SD 11.9). In all, 64% (44/69) opted to receive text messages and 30% (21/69) validated their phone number. During the course of the study, 5 participants texted "Stop" or turned the messages off manually through the Health eRide program.

At baseline, 10% (7/69) screened positive for a TBI. The mean score on the numerical rating scale assessing current pain was 5.8 (SD 2.0). The stage distribution for pain self-management was bimodal: 1% (1/69) of the participants were in the precontemplation stage for pain self-management; 15% (10/69) were in contemplation; 39% (27/69) preparation; 3% (2/69) action; and 42% (29/69) maintenance. Participants selected an average of 2.8 (SD 1.6) MPRCQ2 pain coping strategies to learn more about or work on during their CTI session. They were most likely to select exercise (47%, 32/68), relaxation (47%, 32/68), avoiding pain contingent rest (41%, 28/68), and cognitive control (35%, 24/68).

A total of 44 participants (64%) completed the 30-day follow-up assessment. There were no differences between respondents and nonrespondents on demographics, military history, positive screen for TBI, pain, stage of change for pain self-management, or any other study measures, with the exception of current well-being, assessed using the Cantril Self-Anchoring Scale [65]. Current well-being scores were significantly higher for respondents than for nonrespondents: 6.0 (SD 2.2) versus 4.3 (SD 2.1), respectively, $t_{67}=2.82$, $P=.006$).

Pre-Post Changes

Results, summarized in Tables 1 and 2, show that pre-post changes in the levels of pain and pain impact, as well as stage of change for pain self-management and stress management reached statistical significance; effect sizes were quite large.

Table 1. Pre-post changes on key measures.

Outcomes	Time 1 Mean (SD)	Time 2 Mean (SD)	t [a]	*P* value	Cohen *d*
Pain					
Pain now	5.8 (2.1)	5.0 (2.0)	3.325	.002	0.395
Usual pain past week	6.8 (1.6)	5.4 (1.9)	5.117	<.001	0.751
Best pain past week	4.9 (2.1)	4.0 (2.1)	3.253	.002	0.428
Worst pain past week	8.3 (1.4)	7.0 (1.7)	4.883	<.001	0.804
Pain impact	65.87 (5.4)	61.5 (7.2)	4.908	<.001	0.673
Pain coping skills					
Exercise	4.5 (1.6)	4.6 (1.5)	−2.273	.03	0.399
Task persistence	4.5 (1.7)	4.6 (1.6)	−0.271	.79	0.048
Relaxation	3.8 (1.9)	4.6 (1.7)	−2.835	.007	0.444
Cognitive control	4.1 (1.3)	5.3 (1.1)	−2.149	.04	0.376
Pacing	4.9 (1.8)	3.4 (1.4)	−1.445	.16	0.242
Avoiding pain contingent rest	3.5 (2.1)	3.5 (2.2)	0.222	.83	−0.047
Avoiding asking for assistance	3.7 (2.0)	4.9 (2.1)	0.446	.66	−0.098
Assertive communication	4.8 (2.4)	5.5 (2.1)	−0.252	.80	0.041
Use of proper body mechanics	4.7 (1.8)	4.6 (1.5)	−2.738	.009	0.457
Posttraumatic stress disorder	31.4 (21.5)	27.2 (21.8)	2.008	.05	0.192
Emotional well-being					
Present well-being	6.0 (2.2)	6.3 (1.8)	.966	.34	−0.139
Future well-being	7.2 (2.1)	7.4 (2.1)	.696	.49	−0.098

[a]Paired samples *t* test, degrees of freedom=43.

Table 2. Pre-post changes in percent in action/maintenance.

Target behavior	Time 1, %	Time 2, %	McNemar χ^{2a}	*P* value	Odds ratio
A or M stage-pain management	54.5	79.5	6.67	.010	6.500
A or M stage-stress management	50.0	88.6	13.47	<.001	18.000
A or M stage-healthy sleep	25.0	38.6	2.50	.113	4.000

[a]With continuity correction, degrees of freedom=1; N=44.

Using available benchmarks for interpreting the clinical significance of changes in pain intensity ratings (ie, a 30% reduction from pre to post), rates of at least moderately important improvement were 26% (11/43) for current pain, 32% (14/44) for usual pain in the past week, 34% (15/44) for best pain in the past week, and 23% (10/44) for worst pain in the past week. These rates of clinically significant improvement are comparable with those found in other studies of Web-based pain self-management programs (eg, 38% in a study of a Web-based acceptance and commitment therapy intervention [76] and 19% in a study of a 10-week interactive voice response–based CBT intervention [77]; in the latter study, the rate of clinically significant improvement among patients receiving a 10-week in-person CBT intervention was 33% [77]. Changes in pain impact scores correspond to reduction from severe impact to substantial impact and a drop below the national mean for chronic pain patients. The reduction in symptoms of PTSD approached significance (=.05). There were significant increases in four of the nine pain coping skills assessed with the MPRCQ2, which are as follows: exercise, relaxation, cognitive control, and use of proper body mechanics. Pre-post changes in perceptions of current and future well-being and stage of change for practicing healthy sleep habits were not statistically significant.

Patient Global Impression of Change Scale

When asked to report on their global impressions of change, 41% (18/44) of the respondents reported that they had experienced a slight but noticeable improvement, 11% (5/44) had experienced a definite improvement, and 16% (7/44) said that they had experienced considerable improvement in their condition. Only 32% (14/44) of the participants reported that they had not experienced any noticeable change in their condition or that the change did not make a difference.

Table 3. Mean system usability and acceptability scores (N=44).

Usability measure	Mean (SD[a])
System usability scale score[b]	65.4 (13.3)
Overall acceptability score[c]	3.2 (0.5)
Ten individual acceptability dimensions	
I liked the way the program looked.	3.3 (0.7)
I enjoyed using the program.	3.2 (0.5)
Questions were easy to understand.	3.2 (0.7)
Feedback was easy to understand.	3.3 (0.6)
Program was interesting.	3.3 (0.6)
Program was designed for Veterans.	2.8 (0.9)
Program gave sound advice.	3.2 (0.6)
Program gave me something new to think about.	3.3 (0.6)
Program gave me new ideas about managing pain.	3.3 (0.5)
Program could help me change behavior.	3.0 (0.6)

[a]SD: standard deviation.

[b]Usability criterion: mean system usability scale score ≥68.

[c]On a 4-point scale, acceptability criterion: mean overall acceptability score ≥3.

Program Usability and Acceptability

Program acceptability and usability ratings are presented in Table 3.

The mean usability score for the Health eRide program was 65.4 (SD 13.3), falling slightly short of the mean score of 68 found across other studies of digital materials [73]. The overall mean acceptability score was 3.2 (SD 0.5), exceeding the criterion score of 3.0 for program acceptability. The lowest mean rating was 2.8 for the statement, "The program was designed for Veterans."

In response to the question, "What did you like most about the program?" 95% (42/44) described elements they liked and the remainder (5%, 2/44) provided no response. Participants were most likely to comment that they like the information and content, and the ease of use. For example, participants wrote the following:

> *The program is very easy to use, large print, very intuitive, not a cumbersome program.*
>
> *It made me consider the things I have done to improve my quality of life with pain...exercise, knowing when to take it easy, sleep, eating better.*
>
> *All of it really but the steps the program gives is easy to follow in a pace u control at your own pace they [sic] some methods I used and others I am working on.*
>
> *It not only asked me about my pain and issues, but it also gave me solutions to resolve my issues.*

In response to the question, "What did you like least about the program?" 43% (19/44) said "nothing" or described elements they liked. The remainder (57%, 25/44) described elements they did not like. Respondents were most likely to comment on length of the program, confusion on how to answer some of the questions, confusion over the design of the program and the idea that the program did not necessarily provide users with "new" information:

> *The initial subway hub was confusing and the layout didn't help.*
>
> *Some of the questions were a little difficult to answer based on the answer choices.*
>
> *It seemed to take a lot of questions to get to a conclusion. After I go through everything I am not really sure how to find a particular piece of information that was provided.*
>
> *Some areas were a little confusing...needed to re-read directions, in order to understand what you were looking for.*

In response to the final open-ended question, "How could the program be improved?" 47.7% said, "Nothing" or "Don't know," or made a positive comment about the program—for example, "I think it's fine the way it is." The remainder (53.3%) offered a recommendation on how the program could be improved. Recommendations included making the program shorter, clarifying instructions and the wording of the questions, and making it more usable on mobile. Respondents also suggested adding audio, videos, or other features. For example, participants wrote the following:

> *Easier to drill down into the information. [M]ore concise way to get to the root of the problem and give the option for more info. It would also be nice if there was a notebook like feature where you could save parts that interest you for future reference.*

I didn't notice if there was an audio option for the program. This program was not good for mobile use. Might consider a mobile site.

Videos would be a good tool, seeing reactions of real people and how they manage pain the healthy way.

Discussion

Principal Findings

This research provides preliminary data on the potential impact, usability, and acceptability of Health eRide, a prototype of a TTM-based mobile intervention for pain self-management among veterans. The data are encouraging. After a single session, at 30 days' follow-up, participants reported statistically significant reductions in pain intensity and pain impact, and effect sizes were quite large. Benchmarks for interpreting clinical significance of reductions in pain intensity show that around one-fourth to one-third of the participants experienced at least moderately important improvement on the four measures of pain intensity examined. On the Patient Global Impression of Change scale, over one-fourth of the participants reported either definite or considerable improvement in their pain. Patients also showed significant pre-post changes in readiness to engage in pain self-management and stress management and on readiness to use the following four specific pain self-management strategies: exercise, relaxation, cognitive control, and use of proper body mechanics. Three of those strategies were among those that participants most often chose to focus on in their intervention sessions. Reductions in PTSD approached statistical significance (=.05). Whereas the sample's mean score on SUS fell short of the study's criterion score for establishing feasibility, the mean score on the acceptability measure exceeded the criterion score for establishing acceptability. Responses to open-ended questions show that some participants particularly appreciated the program's clarity and ease of use, whereas others found various components (eg, response options, the layout of the subways station) confusing. Additional usability and program refinement will be necessary to ensure ease of use for all participants. Responses to open-ended questions highlight a number of additional opportunities for improvement, including reducing session length (especially the number of measures) and including more videos. In subsequent implementations, additional efforts will be made to further customize the intervention materials to veterans. Reasons for relatively low validation of phone numbers among participants who opted to receive text messages will be explored.

The challenges to recruitment provide lessons for a subsequent randomized trial. First, the lack of follow-through on the screener on Facebook suggests some distrust of an unknown organization asking for contact information. This hypothesis is supported by the Web-based survey company's success in recruiting, given that respondents had a preexisting relationship with the organization. When recruiting from community sources, it will be critical to have the support and advocacy of an organization that serves veterans to help promote the program from the outset. Second, it may be best to conduct all eligibility screening online, with eligible participants segueing directly to the program log-in page. In some environments, a particularly promising approach may be to integrate the Health eRide program into clinical practice, with provider or clinic endorsement, and the provision of iPad or tablets to support universal Web-based screening and session completion in the waiting room.

Questions may be raised about the role incentives had on veteran's willingness to participate. In the pilot test, financial incentives (redeemable gift cards) were used to encourage participation. It is not uncommon to incentivize research participants [78], particularly because in this pilot study, they completed additional assessments that would not be included in the real-world implementation of Health eRide. Planned eventual dissemination channels for Health eRide include the Veteran's Administration and other veteran-service organizations (eg, Tricare); Veteran-centric social networking sites (eg, Rally Point); the app store; and community-based primary care, where there is a new emphasis on nonpharmacological approaches to managing chronic pain [79]. Previous research demonstrates that primary care provider referrals significantly increased adherence to a recommended behavior change intervention, particularly when accompanied by arranging follow-up [80]. Furthermore, in the longer-term clinical trial and in real-world implementations, nonmonetary incentives for participation in this could include the emotional and instrumental support from other participants via social networking; praise for participation provided by the program and by health care providers if the program is delivered in a clinical setting; the sense of mastery provided by progressing through the subway stops to the final destination in the Health eRide program; and, most importantly, the rewards of improved pain and pain self-management.

Limitations

There are several limitations to this research, including the self-selection bias introduced by the recruitment methods, small sample size, and brief follow-up. Given the bimodal distribution for pain self-management—39% (27/69) in the preparation stage, 42% (29/69) in the maintenance stage), it is safe to say that individuals in the precontemplation and contemplation stages were underrepresented in the study. Concerns are mitigated somewhat by the similarities in age, mean numeric rating of pain, racial and ethnic distribution to the sample recruited by Heapy et al [77]. Another limitation relates to the fact that participation in the 30-day follow-up was predicted by well-being at baseline, with respondents reporting significantly higher well-being than nonrespondents. This may have led to more favorable findings than if all participants had responded.

Future Work

Future work will include in the completion of the development of Health eRide to address the recommendations from pilot participants and lessons learned, to add other enhancements, and to program additional interactions with input from potential end users and experts in pain management, social networking, and gaming. A randomized trial with longer follow-up will be required to assess the efficacy of the Health eRide program. These preliminary data, however, suggest that Health eRide has

the potential to be an important component of an integrated evidence-based approach to pain care among veterans.

Acknowledgments

We gratefully acknowledge funding from the National Center for Advancing Translational Sciences (Grant R43TR000362).

Conflicts of Interest

Sara Johnson is co-owner and co-President and CEO of Pro-Change Behavior Systems, Inc, which may market the Health eRide program. Pro-Change may or may not directly or indirectly profit from this program. Deborah Levesque, Lynne Broderick, and Dustin Bailey are employees of Pro-Change Behavior Systems, Inc.

References

1. Institute of Medicine, Board on Health Sciences Policy, Committee on Advancing Pain Research, Care, and Education. Nap.edu. 2011. Relieving pain in America: a blueprint for transforming prevention, care, education, and research URL: https://www.nap.edu/catalog/13172/relieving-pain-in-america-a-blueprint-for-transforming-prevention-care [accessed 2017-02-10] [WebCite Cache ID 22553896]
2. Briggs JP. Nccih.nih. 2014. Exploring nondrug approaches to managing pain and related conditions in the military URL: https://nccih.nih.gov/about/offices/od/managing-pain-in-military [accessed 2017-09-19] [WebCite Cache ID 6tbDVqYn4]
3. Jonas WB, Schoomaker EB. Pain and opioids in the military: we must do better. JAMA Intern Med 2014 Aug;174(8):1402-1403 [FREE Full text] [doi: 10.1001/jamainternmed.2014.2114] [Medline: 24978149]
4. Toblin RL, Quartana PJ, Riviere LA, Walper KC, Hoge CW. Chronic pain and opioid use in US soldiers after combat deployment. JAMA Intern Med 2014 Aug;174(8):1400-1401. [doi: 10.1001/jamainternmed.2014.2726] [Medline: 24978399]
5. Childress S. Pbs. Veterans face greater risks amid opioid crisis URL: http://www.pbs.org/wgbh/frontline/article/veterans-face-greater-risks-amid-opioid-crisis/ [accessed 2017-10-04] [WebCite Cache ID 6txffSuEu]
6. Toblin RL, Mack KA, Perveen G, Paulozzi LJ. A population-based survey of chronic pain and its treatment with prescription drugs. Pain 2011 Jun;152(6):1249-1255. [doi: 10.1016/j.pain.2010.12.036] [Medline: 21397401]
7. Sinnott P, Wagner TH. Low back pain in VA users. Arch Intern Med 2009 Jul 27;169(14):1338-1339. [doi: 10.1001/archinternmed.2009.201] [Medline: 19636039]
8. Lew HL, Pogoda TK, Hsu PT, Cohen S, Amick MM, Baker E, et al. Impact of the "polytrauma clinical triad" on sleep disturbance in a department of veterans affairs outpatient rehabilitation setting. Am J Phys Med Rehabil 2010 Jun;89(6):437-445. [doi: 10.1097/PHM.0b013e3181ddd301] [Medline: 20489391]
9. Shipherd JC, Keyes M, Jovanovic T, Ready DJ, Baltzell D, Worley V, et al. Veterans seeking treatment for posttraumatic stress disorder: what about comorbid chronic pain? J Rehabil Res Dev 2007;44(2):153-166 [FREE Full text] [Medline: 17551870]
10. Gironda RJ, Clark ME, Ruff RL, Chait S, Craine M, Walker R, et al. Traumatic brain injury, polytrauma, and pain: challenges and treatment strategies for the polytrauma rehabilitation. Rehabil Psychol 2009 Aug;54(3):247-258. [doi: 10.1037/a0016906] [Medline: 19702423]
11. Otis JD, Keane TM, Kerns RD, Monson C, Scioli E. The development of an integrated treatment for veterans with comorbid chronic pain and posttraumatic stress disorder. Pain Med 2009 Oct;10(7):1300-1311. [doi: 10.1111/j.1526-4637.2009.00715.x] [Medline: 19818040]
12. Walker RL, Clark ME, Sanders SH. The "postdeployment multi-symptom disorder": an emerging syndrome in need of a new treatment paradigm. Psychol Serv 2010 Aug;7(3):136-147. [doi: 10.1037/a0019684]
13. Manchikanti L, Singh V, Datta S, Cohen SP, Hirsch JA, American Society of Interventional Pain Physicians. Comprehensive review of epidemiology, scope, and impact of spinal pain. Pain Physician 2009;12(4):E35-E70 [FREE Full text] [Medline: 19668291]
14. Jamison RN, Clark JD. Opioid medication management: clinician beware!. Anesthesiology 2010 Apr;112(4):777-778. [doi: 10.1097/ALN.0b013e3181cdefb4] [Medline: 20216398]
15. Martell BA, O'Connor PG, Kerns RD, Becker WC, Morales KH, Kosten TR, et al. Systematic review: opioid treatment for chronic back pain: prevalence, efficacy, and association with addiction. Ann Intern Med 2007 Jan 16;146(2):116-127. [Medline: 17227935]
16. Stein C, Reinecke H, Sorgatz H. Opioid use in chronic noncancer pain: guidelines revisited. Curr Opin Anaesthesiol 2010 Oct;23(5):598-601. [doi: 10.1097/ACO.0b013e32833c57a8] [Medline: 20585244]
17. Woodcock J. A difficult balance--pain management, drug safety, and the FDA. N Engl J Med 2009 Nov 26;361(22):2105-2107. [doi: 10.1056/NEJMp0908913] [Medline: 19940297]
18. Wu PC, Lang C, Hasson NK, Linder SH, Clark DJ. Opioid use in young veterans. J Opioid Manag 2010;6(2):133-139. [Medline: 20481178]

19. Seal KH, Shi Y, Cohen G, Cohen BE, Maguen S, Krebs EE, et al. Association of mental health disorders with prescription opioids and high-risk opioid use in US veterans of Iraq and Afghanistan. J Am Med Assoc 2012 Mar 07;307(9):940-947. [doi: 10.1001/jama.2012.234] [Medline: 22396516]
20. Williams DA, Kuper D, Segar M, Mohan N, Sheth M, Clauw DJ. Internet-enhanced management of fibromyalgia: a randomized controlled trial. Pain 2010 Dec;151(3):694-702 [FREE Full text] [doi: 10.1016/j.pain.2010.08.034] [Medline: 20855168]
21. Dobscha SK, Corson K, Perrin NA, Hanson GC, Leibowitz RQ, Doak MN, et al. Collaborative care for chronic pain in primary care: a cluster randomized trial. J Am Med Assoc 2009 Mar 25;301(12):1242-1252. [doi: 10.1001/jama.2009.377] [Medline: 19318652]
22. Lamb SE, Hansen Z, Lall R, Castelnuovo E, Withers EJ, Nichols V, Back Skills Training Trial investigators. Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. Lancet 2010 Mar 13;375(9718):916-923. [doi: 10.1016/S0140-6736(09)62164-4] [Medline: 20189241]
23. Kroenke K, Bair MJ, Damush TM, Wu J, Hoke S, Sutherland J, et al. Optimized antidepressant therapy and pain self-management in primary care patients with depression and musculoskeletal pain: a randomized controlled trial. J Am Med Assoc 2009 May 27;301(20):2099-2110 [FREE Full text] [doi: 10.1001/jama.2009.723] [Medline: 19470987]
24. Iprcc.nih. National pain strategy URL: https://iprcc.nih.gov/docs/DraftHHSNationalPainStrategy.pdf [accessed 2017-09-29] [WebCite Cache ID 6tq5SvBaZ]
25. Smeets RJ, Vlaeyen JW, Hidding A, Kester AD, van der Heijden GJ, van Geel AC, et al. Active rehabilitation for chronic low back pain: cognitive-behavioral, physical, or both? First direct post-treatment results from a randomized controlled trial [ISRCTN22714229]. BMC Musculoskelet Disord 2006 Jan 20;7:5. [Medline: 16426449]
26. Sherman KJ, Cherkin DC, Wellman RD, Cook AJ, Hawkes RJ, Delaney K, et al. A randomized trial comparing yoga, stretching, and a self-care book for chronic low back pain. Arch Intern Med 2011 Dec 12;171(22):2019-2026 [FREE Full text] [doi: 10.1001/archinternmed.2011.524] [Medline: 22025101]
27. Mailloux J, Finno M, Rainville J. Long-term exercise adherence in the elderly with chronic low back pain. Am J Phys Med Rehabil 2006 Feb;85(2):120-126. [Medline: 16428902]
28. Henschke N, Ostelo RW, van Tulder MW, Vlaeyen JW, Morley S, Assendelft WJ, et al. Behavioural treatment for chronic low-back pain. Cochrane Database Syst Rev 2010 Jul 7(7):CD002014. [doi: 10.1002/14651858.CD002014.pub3] [Medline: 20614428]
29. Hoffman BM, Papas RK, Chatkoff DK, Kerns RD. Meta-analysis of psychological interventions for chronic low back pain. Health Psychol 2007 Jan;26(1):1-9. [doi: 10.1037/0278-6133.26.1.1] [Medline: 17209691]
30. Nicholas MK, Asghari A, Corbett M, Smeets RJ, Wood BM, Overton S, et al. Is adherence to pain self-management strategies associated with improved pain, depression and disability in those with disabling chronic pain? Eur J Pain 2012 Jan;16(1):93-104. [doi: 10.1016/j.ejpain.2011.06.005] [Medline: 21705246]
31. Poushter J. Pewglobal. Smartphone ownership and Internet usage continues to climb in emerging economies URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2017-09-29] [WebCite Cache ID 6tq5ahVmo]
32. comScore. comScore reports January 2016 US smartphone subscriber market share URL: http://www.comscore.com/Insights/Rankings/comScore-Reports-January-2016-US-Smartphone-Subscriber-Market-Share [accessed 2017-09-29] [WebCite Cache ID 6tq5cwtWH]
33. Luger TM, Hogan TP, Richardson LM, Cioffari-Bailiff L, Harvey K, Houston TK. Older veteran digital disparities: examining the potential for solutions within social networks. J Med Internet Res 2016 Nov 23;18(11):e296 [FREE Full text] [doi: 10.2196/jmir.6385] [Medline: 27881361]
34. Macea DD, Gajos K, Daglia Calil YA, Fregni F. The efficacy of Web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis. J Pain 2010 Oct;11(10):917-929. [doi: 10.1016/j.jpain.2010.06.005] [Medline: 20650691]
35. Palermo TM, Eccleston C, Lewandowski AS, Williams AC, Morley S. Randomized controlled trials of psychological therapies for management of chronic pain in children and adolescents: an updated meta-analytic review. Pain 2010 Mar;148(3):387-397 [FREE Full text] [doi: 10.1016/j.pain.2009.10.004] [Medline: 19910118]
36. Ruehlman LS, Karoly P, Enders C. A randomized controlled evaluation of an online chronic pain self management program. Pain 2012 Feb;153(2):319-330 [FREE Full text] [doi: 10.1016/j.pain.2011.10.025] [Medline: 22133450]
37. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014 Jun;15(6):898-909. [doi: 10.1111/pme.12327] [Medline: 24422990]
38. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
39. Portelli P, Eldred C. A quality review of smartphone applications for the management of pain. Br J Pain 2016 Aug;10(3):135-140 [FREE Full text] [doi: 10.1177/2049463716638700] [Medline: 27583140]
40. Alexander JC, Joshi GP. Smartphone applications for chronic pain management: a critical appraisal. J Pain Res 2016 Sep 26;9:731-734 [FREE Full text] [doi: 10.2147/JPR.S119966] [Medline: 27713649]

41. Kerns RD, Philip EJ, Lee AW, Rosenberger PH. Implementation of the veterans health administration national pain management strategy. Transl Behav Med 2011 Dec;1(4):635-643 [FREE Full text] [doi: 10.1007/s13142-011-0094-3] [Medline: 24073088]
42. Prochaska JO, DiClemente CC. Stages of change in the modification of problem behaviors. Prog Behav Modif 1992;28:183-218. [Medline: 1620663]
43. DiClemente CC, Prochaska JO, Fairhurst SK, Velicer WF, Velasquez MM, Rossi JS. The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages of change. J Consult Clin Psychol 1991 Apr;59(2):295-304. [Medline: 2030191]
44. Hall KL, Rossi JS. Meta-analytic examination of the strong and weak principles across 48 health behaviors. Prev Med 2008 Mar;46(3):266-274. [doi: 10.1016/j.ypmed.2007.11.006] [Medline: 18242667]
45. Prochaska JO, Velicer WF, DiClemente CC, Fava J. Measuring processes of change: applications to the cessation of smoking. J Consult Clin Psychol 1988 Aug;56(4):520-528. [Medline: 3198809]
46. Campbell MK, DeVellis BM, Strecher VJ, Ammerman AS, DeVellis RF, Sandler RS. Improving dietary behavior: the effectiveness of tailored messages in primary care settings. Am J Public Health 1994 May;84(5):783-787. [Medline: 8179049]
47. Skinner CS, Campbell MK, Rimer BK, Curry S, Prochaska JO. How effective is tailored print communication? Ann Behav Med 1999;21(4):290-298. [Medline: 10721435]
48. Brug J, Glanz K, Van Assema P, Kok G, van Breukelen GJ. The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake. Health Educ Behav 1998 Aug;25(4):517-531. [Medline: 9690108]
49. Johnson SS, Driskell MM, Johnson JL, Dyment SJ, Prochaska JO, Prochaska JM, et al. Transtheoretical model intervention for adherence to lipid-lowering drugs. Dis Manag 2006 Apr;9(2):102-114. [doi: 10.1089/dis.2006.9.102] [Medline: 16620196]
50. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. Ann Behav Med 1999;21(4):276-283. [Medline: 10721433]
51. Johnson SS, Paiva AL, Cummins CO, Johnson JL, Dyment SJ, Wright JA, et al. Transtheoretical model-based multiple behavior intervention for weight management: effectiveness on a population basis. Prev Med 2008 Mar;46(3):238-246 [FREE Full text] [doi: 10.1016/j.ypmed.2007.09.010] [Medline: 18055007]
52. Prochaska JO. Treating entire populations for behavior risks for chronic diseases. Homeost Health Dis 2003;42:1-12.
53. McGonigal J. Reality Is Broken: Why Games Make Us Better and How They Can Change the World. New York: Penguin Group; 2011.
54. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
55. Nielson WR, Armstrong JM, Jensen MP, Kerns RD. Two brief versions of the multidimensional pain readiness to change questionnaire, version 2 (MPRCQ2). Clin J Pain 2009 Jan;25(1):48-57. [doi: 10.1097/AJP.0b013e3181817ab4] [Medline: 19158546]
56. Nielson WR, Jensen MP, Ehde DM, Kerns RD, Molton IR. Further development of the multidimensional pain readiness to change questionnaire: the MPRCQ2. J Pain 2008 Jun;9(6):552-565 [FREE Full text] [doi: 10.1016/j.jpain.2008.01.327] [Medline: 18337183]
57. Nielson WR, Jensen MP, Kerns RD. Initial development and validation of a multidimensional pain readiness to change questionnaire. J Pain 2003 Apr;4(3):148-158. [Medline: 14622712]
58. Dunn AS, Julian T, Formolo LR, Green BN, Chicoine DR. Preliminary analysis of posttraumatic stress disorder screening within specialty clinic setting for OIF/OEF veterans seeking care for neck or back pain. J Rehabil Res Dev 2011;48(5):493-502. [Medline: 21674400]
59. Schwab KA, Baker G, Ivins BJ, Sluss-Tiller M, Lux W, Warden D. Insights.ovid. The brief traumatic brain injury screen (BTBIS): investigating the validity of a self-report instrument for detecting traumatic brain injury (TBI) in troops returning from deployment in Afghanistan and Iraq URL: https://insights.ovid.com/neurology/neur/2006/03/001/brief-traumatic-brain-injury-screen-btbis/974/00006114 [accessed 2017-09-29] [WebCite Cache ID 6tq6EWdMS]
60. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
61. Jensen MP. Hypnosis for Chronic Pain Management: Therapist Guide. New York: Oxford University Press; 2011.
62. Becker J, Schwartz C, Saris-Baglama RN, Kosinski M, Bjorner JB. Using item response theory (IRT) for developing and evaluating the Pain Impact Questionnaire (PIQ-6™). Pain Med 2007 Oct;8(suppl 3):S129-S144. [doi: 10.1111/j.1526-4637.2007.00377.x]
63. Weathers F, Litz B, Herman DS, Huska JA, Keane TM. The PTSD checklist (PCL): reliability, validity, and diagnostic utility. 1993 Presented at: Annual Meeting of the International Society for Traumatic Stress Studies; 1993; San Antonio.
64. Yeager DE, Magruder KM, Knapp RG, Nicholas JS, Frueh BC. Performance characteristics of the posttraumatic stress disorder checklist and SPAN in Veterans Affairs primary care settings. Gen Hosp Psychiatry 2007;29(4):294-301. [doi: 10.1016/j.genhosppsych.2007.03.004] [Medline: 17591505]
65. Cantril H. The Pattern of Human Concerns. New Brunswick: Rutgers University Press; 1965.

66. Levesque D. IC hope self-management module: a staged approach to managing IC pain. 2013 Presented at: Interstitial Cystitis Foundation 2013 Patient Forum; 2013; Boston.
67. Evers KE, Prochaska JO, Johnson JL, Mauriello LM, Padula JA, Prochaska JM. A randomized clinical trial of a population- and transtheoretical model-based stress-management intervention. Health Psychol 2006 Jul;25(4):521-529. [doi: 10.1037/0278-6133.25.4.521] [Medline: 16846327]
68. Velicer WF, Prochaska JO, Fava JL, Norman GJ, Redding CA. Smoking cessation and stress management: applications of the transtheoretical model of behavior change. Homeost Health Dis Internet 1998 Jul;38:216-233 [FREE Full text]
69. Johnson SS, Castle PH. How do sleep, food, mood, and exercise relate to well-being? Am J Health Promot 2015 Apr;29(4):TAHP4-TAHP5. [Medline: 25884055]
70. Brooke J. SUS-a quick and dirty usability scale. Usability Eval Ind 1996;189(194):4-7.
71. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40.
72. Usability.gov. 2013. System usability scale (SUS) URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2017-10-03] [WebCite Cache ID 6tvyJsk3N]
73. Sauro J. MeasuringU. 2011. MeasuringU: measuring usability with the system usability scale (SUS) URL: https://measuringu.com/sus/ [accessed 2017-09-29] [WebCite Cache ID 6tq7jLNJL]
74. National Cancer Institute. Cancer.gov. 1989. Making health communication programs work: a planner's guide URL: https://www.cancer.gov/publications/health-communication/pink-book.pdf [accessed 2017-10-04] [WebCite Cache ID 6txk6qMLq]
75. Dunlap WP, Cortina JM, Vaslow JB, Burke MJ. Meta-analysis of experiments with matched groups or repeated measures designs. Psychol Methods 1996;1(2):170-177. [doi: 10.1037/1082-989X.1.2.170]
76. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KM. Internet-based guided self-help intervention for chronic pain based on acceptance and commitment therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
77. Heapy AA, Higgins DM, Goulet JL, LaChappelle KM, Driscoll MA, Czlapinski RA, et al. Interactive voice response-based self-management for chronic back pain: the COPES noninferiority randomized trial. J Am Med Assoc Intern Med 2017 Jun 01;177(6):765-773. [doi: 10.1001/jamainternmed.2017.0223] [Medline: 28384682]
78. Singer E, Couper MP. Do incentives exert undue influence on survey participation? Experimental evidence. J Empir Res Hum Res Ethics 2008 Sep;3(3):49-56 [FREE Full text] [doi: 10.1525/jer.2008.3.3.49] [Medline: 19385770]
79. Schneiderhan J, Clauw D, Schwenk TL. Primary care of patients with chronic pain. J Am Med Assoc 2017 Jun 20;317(23):2367-2368. [doi: 10.1001/jama.2017.5787] [Medline: 28520835]
80. Park ER, Gareen IF, Japuntich S, Lennes I, Hyland K, DeMello S, et al. Primary care provider-delivered smoking cessation interventions and smoking cessation among participants in the national lung screening trial. J Am Med Assoc Intern Med 2015 Sep;175(9):1509-1516 [FREE Full text] [doi: 10.1001/jamainternmed.2015.2391] [Medline: 26076313]

Abbreviations

CBT: cognitive behavioral therapy
CTI: computer tailored intervention
MPRCQ2: Multidimensional Pain Readiness to Change Questionnaire
OEF: Operations Enduring Freedom
OIF: Operations Iraqi Freedom
PAC: Personal Activity Center
PMD: postdeployment multisymptom disorder
PTSD: posttraumatic stress disorder
SMS: short message service
SD: standard deviation
SSI: Survey Sampling International
SUS: System Usability Scale
TBI: traumatic brain injury
TTM: transtheoretical model
VACHS: Veterans Administration Connecticut Healthcare System
VHA: Veterans Health Administration

Edited by G Eysenbach; submitted 07.12.16; peer-reviewed by S Baptista, A Paglialonga; comments to author 19.01.17; revised version received 16.06.17; accepted 04.08.17; published 17.10.17.

Please cite as:
Johnson SS, Levesque DA, Broderick LE, Bailey DG, Kerns RD
Pain Self-Management for Veterans: Development and Pilot Test of a Stage-Based Mobile-Optimized Intervention
JMIR Med Inform 2017;5(4):e40
URL: http://medinform.jmir.org/2017/4/e40/
doi:10.2196/medinform.7117
PMID:29042341

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org