JMIR Medical Informatics

Contents

Viewpoint

Clinical Note Creation, Binning, and Artificial Intelligence

Rodrigo Octávio Deliberato[1,2], MD, PhD; Leo Anthony Celi[2,3], MD, MPH, MS; David J Stone[2,4], MD

[1]Big Data Analytics Department, Hospital Israelita Albert Einstein, São Paulo, Brazil

[2]Harvard – MIT, Division of Health Science and Technology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, United States

[3]Beth Israel Deaconess Medical Center, Boston, MA, United States

[4]Departments of Anesthesiology and Neurosurgery, University of Virgina School of Medicine, Charlottesville, VA, United States

Corresponding Author:
Leo Anthony Celi, MD, MPH, MS
Harvard – MIT, Division of Health Science and Technology
Institute for Medical Engineering and Science
Massachusetts Institute of Technology
77 Massachusetts Ave, MIT E25-505
Cambridge, MA, MA 02139
United States
Phone: 1 617 253 7937
Fax: 1 617 258 07859
Email: lceli@mit.edu

Abstract

The creation of medical notes in software applications poses an intrinsic problem in workflow as the technology inherently intervenes in the processes of collecting and assembling information, as well as the production of a data-driven note that meets both individual and healthcare system requirements. In addition, the note writing applications in currently available electronic health records (EHRs) do not function to support decision making to any substantial degree. We suggest that artificial intelligence (AI) could be utilized to facilitate the workflows of the data collection and assembly processes, as well as to support the development of personalized, yet data-driven assessments and plans.

Introduction

Many doctors find the creation of the same note more onerous in an electronic health record (EHR) than on paper [1]. The following quote from a senior physician reflects the dissatisfaction doctors have with EHRs: " My experience with the EHR is that it is the biggest waste of time, interferes with patient care, forces the physician to collect thousands of pieces of useless information, and produces marginal improvements in quality ". For this and many other reasons, the quality of EHR documentation has ranged from suboptimal to dismal [2,3]. This paper explores and envisions how artificial intelligence (AI), which is increasingly transforming facets of daily living, could support the currently burdensome process of gathering and organizing the elements necessary for the creation of a clinical note.

Finding the Right Pieces

Part of the issue involves the user interface, where many users are not terribly facile with the keyboard and typing. It would probably be worthwhile for the creators of EHRs to design their user interfaces to be as similar as possible to the Internet-based applications, such as Web browsers, that even those who are unsophisticated with computers use every day. But the fundamental reason for this discomfort is that electronic note writers are not able to pull information seamlessly and freely from their own minds to create the contents of the kind of notes they wish to create. In contrast to the historic paper-based documentation workflow, the EHR user must painfully search through the bins of items buried in the software to extract the correct "pieces" of information necessary to complete the entry, requiring click after click after click in that process (Figure 1). While the freedom involved in creating paper notes might represent a positive, nostalgic memory, the healthcare system

is not going to abandon EHRs with all the manifold advantages that they represent and provide.

In the Lego system, the myriad individual pieces (or modules) are assembled together by the rules (or protocols) dictated by the snap connections to create the toy version of an engineered system [4]. In creating a note, the user identifies and captures the necessary data pieces, analyzes and reassembles the pieces to assess the clinical situation at the level of complexity required, and develops a plan of action, thereby recreating a kind of

clinical data system in itself each time a note is completed and entered [5]. But instead of rummaging around in a variety of bins for the right pieces, how could the de-binning ordeal be circumvented, and even improved, by a technical solution? We propose that a carefully engineered implementation of AI into the note creation software elements of the EHR would not only reduce the required rummaging through bins of pieces, but could assist in the assembly of those pieces into the desired output (ie, a useful, readable, and cogent note that meets all the necessary requirements for clinical documentation).

Figure 1. An assorted bin and 5 (mostly) color-coded bins of Lego toy pieces. The color-coded Legos may represent items that clearly and cleanly fall into a particular section of the note, depending on how the note is organized (ie, SOAP versus systems-oriented).

Analyzing and Assembling the Pieces

In the context of EHRs, how can the natural, direct brain-to-hand workflow of paper note creation process be digitally recreated to simulate the free and seamless flow of information that historically emanates from the clinician's brain directly onto paper? How could the obstructive middleman of technology be enhanced to support, rather than clog the process of clinical documentation? And could this be done in a fashion that makes utilization as intuitive as current Web browsers are to use? Furthermore, in addition to supplying the pieces, can this support also be applied to the assembly of the assessment and plan to assist in the production of a note that preserves the personal character—or the signature or "human"-ess—of the note writer? Optimally, in contrast to today's copied and pasted rote entries, the production of a note that is more interesting and easier to read than current electronic notes would also be a goal of this redesign process.

We will progressively need to introduce important note information from other sources (eg, personal device and

patient-entered data, population databases, even genomics) that supplement what is now available to the clinicians creating or reviewing the note at a later time [6]. In a previous publication, we described an engineered system that would support electronic note writing but did not specifically suggest how this might be done in a technical sense [7]. We suggest that the now increasingly familiar tools provided by AI provide a potential means by which to "de-bin" the process of data element selection and assist in the assembly of the data pieces with the goal of improved and more efficient electronic note creation.

AI has the potential to assist users in extracting the right information from the different information systems (ie, previous electronic notes and bedside monitors, and imaging, laboratory and pharmacy systems), assembling this information into the proper places in the note to assist in the formulation of the assessment with some bounds of certainty, and to analyze that assessment to develop a data-driven plan of action. There are many tools in the AI armamentarium—machine learning, natural language processing, computer vision, constraint satisfaction—but in essence, AI would power a learning

interface between the human user and digital health information system to produce a note that would be highly, and increasingly over use, similar to that note-writer's mental representation of what a clinical note should be.

We do realize that AI cannot analyze and repackage data until the latter has been incorporated into the system. The current history and physical examination, whether taken at the bedside or the office examining room, cannot be leveraged for note writing until they are so entered. Better, easier means for this must be devised: this might involve free text entry by voice recognition or keyboard, natural language processing of free text to enter structured data into the system, or new AI modalities as this exploding field develops.

Based on that current user input, as well as all available automatic data sources (eg, prior electronic notes, interfaced data like labs, and vital signs), AI would provide helpful suggestions to the user about what information is available and how it might influence the next course of action. AI could also function to emphasize or deemphasize certain elements of the record, based on previous results, external databases, and knowledge networks [8]. The technical strategy for providing these services could rest on a number of already available software solutions such as the tentative, but often very informative, textual suggestions that Google makes during searches. The careful use of autofill, especially for information types that tend to be repetitive, would cut down on excessive clicking and typing. Seamless but secure Internet connections integrated with the user interface would also facilitate decision support by allowing users to actively seek information not

intrinsically provided even in these "smarter" applications. In addition, AI would help format and populate the note based on what has been "learned" from a user's (and/or the patient's) prior entries. There are tradeoffs and risks to innovation, but the quality of AI will intrinsically improve as it is employed in this regard, and the physician user will be the ultimate filter of what is saved and incorporated into the EHR.

Creating an AI-enhanced Learning Healthcare System

At the heart of note writing is communicating important clinical events and decisions between different providers, and with the advent of patient portals, between providers and patients. AI is not the panacea to every problem in healthcare, but for a relatively repetitive and clearly defined task such as clinical note creation, it seems to provide a fairly ideal solution. It also bestows an opportunity to support an interdisciplinary care environment by learning from inter-specialty communication specifics and facilitating shared decision making by mining patient input and feedback. The final note would be the product of the user, but a user who is not exhausted by painful de-binning and endless clicking to insert the right data in the right places. An AI-enhanced system would boost the clinical workflow element of documentation, and maybe even inject some fun into the process of note writing. Such technology is upon us: 1 in 10 communications to the AI-powered personal assistant Amy (or Andrew) Ingram is a note of thanks, a testament to the 21st century computer passing the Turing test [9]. We certainly hope that an EHR company or some budding entrepreneur will take notice of this article and consider our idea in creating the next generation of EHRs.

Conflicts of Interest

None declared.

References

1. Hingle S. Electronic health records: an unfulfilled promise and a call to action. Ann Intern Med 2016 Dec 06;165(11):818-819. [doi: 10.7326/M16-1757] [Medline: 27595501]
2. Markel A. Copy and paste of electronic health records: a modern medical illness. Am J Med 2010 May;123(5):e9. [doi: 10.1016/j.amjmed.2009.10.012] [Medline: 20399309]
3. Hirschtick RE. A piece of my mind. John Lennon's elbow. JAMA 2012 Aug 01;308(5):463-464. [doi: 10.1001/jama.2012.8331] [Medline: 22851112]
4. Csete ME, Doyle JC. Reverse engineering of biological complexity. Science 2002 Mar 01;295(5560):1664-1669. [doi: 10.1126/science.1069981] [Medline: 11872830]
5. Celi LA, Csete M, Stone D. Optimal data systems: the future of clinical predictions and decision support. Curr Opin Crit Care 2014 Oct;20(5):573-580 [FREE Full text] [doi: 10.1097/MCC.0000000000000137] [Medline: 25137399]
6. Smoyer WE, Embi PJ, Moffatt-Bruce S. Creating local learning health systems: think globally, act locally. JAMA 2016 Dec 20;316(23):2481-2482. [doi: 10.1001/jama.2016.16459] [Medline: 27997662]
7. Celi LA, Marshall JD, Lai Y, Stone DJ. Disrupting electronic health records systems: the next generation. JMIR Med Inform 2015 Oct 23;3(4):e34 [FREE Full text] [doi: 10.2196/medinform.4192] [Medline: 26500106]
8. Celi LA, Zimolzak AJ, Stone DJ. Dynamic clinical data mining: search engine-based decision support. JMIR Med Inform 2014 Jun 23;2(1):e13 [FREE Full text] [doi: 10.2196/medinform.3110] [Medline: 25600664]
9. David RP. Big Think. 2017 Jun 11. Thanks, robot! Humans are showing kindness with their AI helpers URL: http://bigthink.com/david-ryan-polgar/thanks-robot-how-should-we-be-treating-our-ai-helpers [accessed 2017-06-06] [WebCite Cache ID 6r684hJbA]

Abbreviations

AI: artificial intelligence

XSL•FO

RenderX

EHR: electronic health record

XSL•FO
RenderX

Viewpoint

Health Information Technology (HIT) Adaptation: Refocusing on the Journey to Successful HIT Implementation

Po-Yin Yen[1,2], RN, PhD; Ann Scheck McAlearney[3], ScD; Cynthia J Sieck[3], PhD; Jennifer L Hefner[3], PhD; Timothy R Huerta[3], PhD

[1]Washington University in St Louis, Institute for Informatics, St Louis, MO, United States

[2]Goldfarb School of Nursing, BJC Healthcare, St Louis, MO, United States

[3]The Ohio State University, Department of Family Medicine, Columbus, OH, United States

Corresponding Author:
Po-Yin Yen, RN, PhD
Washington University in St Louis
Institute for Informatics
4444 Forest Park Ave, Suite 6318
St Louis, MO,
United States
Phone: 1 314 273 2213
Email: yenp@wustl.edu

Abstract

In past years, policies and regulations required hospitals to implement advanced capabilities of certified electronic health records (EHRs) in order to receive financial incentives. This has led to accelerated implementation of health information technologies (HIT) in health care settings. However, measures commonly used to evaluate the success of HIT implementation, such as HIT adoption, technology acceptance, and clinical quality, fail to account for complex sociotechnical variability across contexts and the different trajectories within organizations because of different implementation plans and timelines. We propose a new focus, HIT adaptation, to illuminate factors that facilitate or hinder the connection between use of the EHR and improved quality of care as well as to explore the trajectory of changes in the HIT implementation journey as it is impacted by frequent system upgrades and optimizations. Future research should develop instruments to evaluate the progress of HIT adaptation in both its longitudinal design and its focus on adaptation progress rather than on one cross-sectional outcome, allowing for more generalizability and knowledge transfer.

Introduction

Health information technology (HIT) is defined as "the application of information processing involving both computer hardware and software that deals with the storage, retrieval, sharing, and use of health care information, data, and knowledge for communication and decision making" [1]. During the past 10 years in the United States, several policies, such as the Health Information Technology for Economic and Clinical Health Act, have led to accelerated HIT adoption and implementation in health care settings, especially implementation of electronic health record (EHR) systems [2,3]. In addition, the Centers for Medicare and Medicaid Services established the EHR incentive program to promote the development of a robust HIT infrastructure, and as part of that effort, released Meaningful Use (MU) criteria in 2010. These criteria require hospitals to implement advanced capabilities of certified EHRs by certain dates in order to receive financial incentives. Other efforts focused on the creation of regional extension centers to facilitate the transition to EHR use through training. MU criteria consist of 3 stages [4]: stage 1, begun in 2011, has a focus on data capture and sharing; stage 2, begun in 2014, aims to improve clinical processes with health information exchange, ePrescription, and patient access; and stage 3, in 2017, was recently replaced by Advanced Care Information [5,6] due to criticism of the MU program [7,8].

Hospitals have been rapidly responding to these new policies and incentives with large-scale implementations of EHRs during the past few years. Adopting new technology requires the redesign of individual and collective workflows and results in

changes in both organizational structure and process [9-13]. Yet rapid adoption may hinder the interoperability of the EHR system [14,15]. To facilitate appropriate adoption and use, upgrades, redesign, and optimization are needed, including both minor and major changes in EHR infrastructures, functions, interfaces, and workflows. Further, recent studies have shown that there is a close relationship between the speed of adoption and patient safety concerns of clinicians, both across facilities and within different units [16-18]. EHR implementation could be a distraction from patient care with negative impact on patient outcomes [19] and has mixed association with quality improvement [20,21].

At the same time, studies suggest that unsuccessful implementation of HIT systems could be due to poorly designed HIT, poor use of HIT by clinicians, or socioorganizational factors such as goal conflicts, lack of time, or lack of support from colleagues [22]. However, these studies lack clarity in their measures [23]. This lack of differentiation between technological and human factors thus limits the ability to apply research findings to practice in technology implementation [24].

Given MU regulations, MU requirements have commonly been used as a means to assess HIT implementation success in order to promote essential HIT functionalities [4]. For example, MU stage 2 requires providers to have certain HIT functionalities (eg, computerized provider order entry, personal health record, medication reconciliation) in order to continue to participate in the EHR incentive programs [25]. However, this approach also creates a ceiling effect, hindering the advancement of innovative utilities. While the MU program may accelerate development and implementation of certain key functions, it also slows down other functionalities [26,27]. By focusing on achieving MU, we risk missing the big picture of health care system changes. Therefore, we propose that there is a need to improve our understanding of how to appropriately assess the performance and success of HIT implementation over time to allow us to generalize to other HIT implementation contexts.

Measuring Health Information Technology Implementation Success

Successful HIT implementation is commonly evaluated using measures such as HIT adoption, technology acceptance, and clinical quality. Yet this disparate array of measures fails to account for complex sociotechnical interactions, variability across contexts, and the different trajectories within organizations that exist because of different implementation plans and timelines. Appropriate measurement of HIT implementation thus needs to take into account this variability across organizations and over time but at the same time enable us to generalize the variation across HIT implementation studies in order to inform practice. As a result, the issue of consistent measurement becomes increasingly significant.

Current measures that exist in the literature include HIT adoption, HIT acceptance, and clinical quality measures (CQMs). The first common measure, HIT adoption, is defined by the EHR MU stages outlined by the Office of the National Coordinator and measures the rate of health care systems having chosen to invest resources toward EHR implementation. It is commonly reported as an adoption rate to reflect the percentage of health care organizations with specific EHR functionalities or capabilities that are meaningful for patient care. In 2013, 59% of hospitals reported at least a basic EHR system, but only 5.1% could meet the MU stage 2 criteria [2]. The expectation is that more meaningful use of an EHR system will ultimately result in improved care and more empowered clinicians. In addition, the Healthcare Information and Management Systems Society (HIMSS) measures EHR adoption through the Electronic Medical Record Adoption Model (EMRAM), which categorizes EHR capabilities into an 8-stage scale from stage 0 to stage 7 [28]. In 2015, HIMSS Analytics' Annual Study reported that 27% of hospitals are at stage 6 or above. Although it is helpful to recognize the EHR capabilities across organizations in the nation, it is unclear whether those functions are fully used by clinicians.

The second approach to measuring implementation success involves HIT acceptance, the extent of individual commitment to use the technology [29-33]. When assessing individual user acceptance, the technology acceptance model (TAM) [34,35] is a commonly applied and useful model, albeit with limitations [36]. TAM's predictive power in health care is lower than what has been found in other domains [24], and some recommend that the TAM should be integrated with other adoption theories [36], particularly those that include variables related to both human and social change processes [24].

CQMs [37] are another common metric used to assess the success of HIT [38]. However, HIT implementation appears to have little impact on care quality whether measured by patient mortality, adverse drug events, or readmission rates [39]. Although CQMs are helpful for assessing the extent to which HIT can be used to monitor the quality of health care services provided, this approach to measurement does not take into account organizational or human factors that could impact HIT implementation.

Measuring HIT adoption and acceptance alone provides only a limited understanding of HIT success. Both HIT adoption rates and TAM are helpful to understand the status of HIT implementation and acceptance, but they do not inform a strategic plan for promoting successful HIT implementation in a health care organization. CQM as a proxy for HIT success also fails to take into account the organizational context of implementation. In short, as HIT implementation is a process, not an outcome, understanding implementation success requires consideration of the sociotechnical environment in which it takes place.

Sociotechnical Theory: Improving Our Understanding of Health Information Technology Implementation

Sociotechnical theory positions people-focused (socio) elements, organizational and human, and information technology elements (technical) as interdependent parts of a system that cannot be studied in isolation and therefore must be evaluated together [40]. Sociotechnical theory has been discussed as a theoretical framework that is responsive to the tenets of complex adaptive systems (CAS) [41-44]. When viewed in concert, these 2 theoretical approaches support that interdependent interactions between people and technology within the workplace have both direct impacts, in the classical cause and effect sense, and

impacts through feedback, where these same people and technology attenuate, strengthen, distort, halt, or change valence over time [41,43,45].

Current sociotechnical evaluations involve assessing both the technology and the social contexts where the technology is implemented. A systematic review conducted on EHR implementations revealed that sociotechnical factors complicate HIT deployments [46]. Technical features of HIT interact with the social features of a health care work environment. Further, it has been demonstrated that the quality of the implementation process is just as important as the features and capabilities of the system being implemented [47-49].

We suggest grounding the theoretical framing of CAS that refers to adaptiveness as "the ability of components of a CAS to change their behavior as a result of interactions with the other components and the surroundings" [41]. In shifting the concept of adoption to adaptation, we frame sociotechnological change as occurring over time with system response characterized as the adaptiveness of a health care organization in the context of changes to HIT implementation [42,44]. For example, technical features are not static; rather they frequently change over time as new versions of the software are promulgated. As such,

adoption is not an end state; it is the application of an arbitrary end point to facilitate our understanding. From that perspective, understanding the adaptiveness, or HIT adaptation in this process, is thus significant in our understanding of HIT implementation success [50].

Health Information Technology Adaptation

Although sociotechnical theory and CAS have been used to explain complexity in health care [51], little has been discussed that uses adaptation as a measure to evaluate the success of HIT implementation over time. We thus propose a new focus: adaptation. Adaptation is conceptualized as "a process of modifying existing conditions in an effort to achieve alignment" [52] involving workflow redesign, user trainings, and technology maintenance [53]. In the context of HIT implementation, refocusing from adoption and acceptance to adaptation illuminates factors that facilitate or hinder the connection between use of the EHR and improved quality of care. Further, by shifting to adaptation, we refocus the question of HIT adoption to one that explores the trajectory of change as an explicit component of the way we measure these issues. Table 1 presents the definitions of adoption, adaptation, and acceptance as differentiated by Cooper and Zmud [53].

Table 1. Definitions of adoption, adaptation, and acceptance [54].

Concept	Definition
Adoption	A decision is reached to invest resources to accommodate the implementation effort.
Adaptation	The innovation is developed, installed, and maintained. Procedures are developed and revised. Members are trained both in the new procedures and in the innovation.
Acceptance	Organizational members are induced to commit to the innovation's usage.

MU criteria [54-56] and CQMs can be seen as verification and validation steps, respectively, for HIT implementation. In product or system design, evaluation is commonly done via verification and validation. Verification serves as quality control to assess whether a system is in compliance with regulations and specifications. On the other hand, validation is a quality assurance process that establishes evidence to ensure a system accomplishes what was intended. However, no measures have been proposed to assess HIT implementation performance between the steps of system verification and validation. We suggest that this period encompasses the HIT adaptation process, requiring its own measurement approach.

In Figure 1, we illustrate the current knowledge gap between meeting the MU criteria and achieving CQMs, linking this

conceptually to Donabedian's well-known structure-process-outcomes model, a quality assessment model presented to evaluate health services outcomes [57]. Considering HIT implementation in the context of the Donabedian model, structure refers to HIT resources, which are determined by MU criteria; process refers to clinicians' use or adaptation of HIT for their use; and outcomes refer to the effects of using HIT for the delivery of health care, as measured by CQMs. In practice, the HIT implementation journey will be impacted by frequent system upgrades and optimizations, leading to performance variability throughout the process. However, by including considerations of sociotechnical factors such as technology acceptance, communication and collaboration, work productivity, training and competency, leadership, and policy, the progress of HIT adaptation could be appropriately assessed.

Figure 1. Health information technology adaptation measures as the process evaluation.
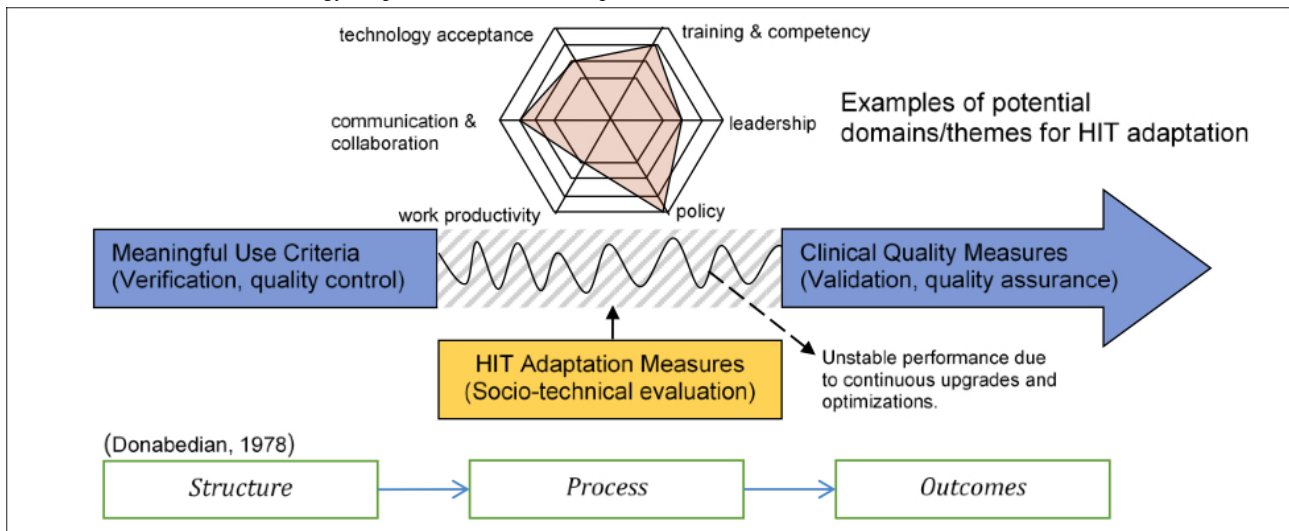
Theoretical Frameworks to Study Health Information Technology Adaptation

Two theoretical frameworks provide guidance for HIT adaptation research: the information technology (IT) implementation framework [58] and a new sociotechnical model [42]. First, the IT implementation framework [58] suggests that (1) IT use is complex, multidimensional, and influenced by a variety of factors at individual and organizational levels and (2) success in achieving change is enhanced by active participation of members from the target user group [58]. The new sociotechnical model [42] now aims to study HIT in complex adaptive health care systems and suggests investigating 8 dimensions: (1) hardware and software computing infrastructure; (2) clinical content; (3) human-computer interface; (4) people; (5) workflow and communication; (6) internal organizational policies, procedures, and culture; (7) external rules, regulations, and pressures; and (8) system measurement and monitoring [42]. Figure 2 illustrates our adapted model from the new sociotechnical model [42]. We do not include the seventh dimension, "external rules, regulations, and pressures," as we focus on factors within the organization.

Figure 2. Adapted from the sociotechnical model [43].
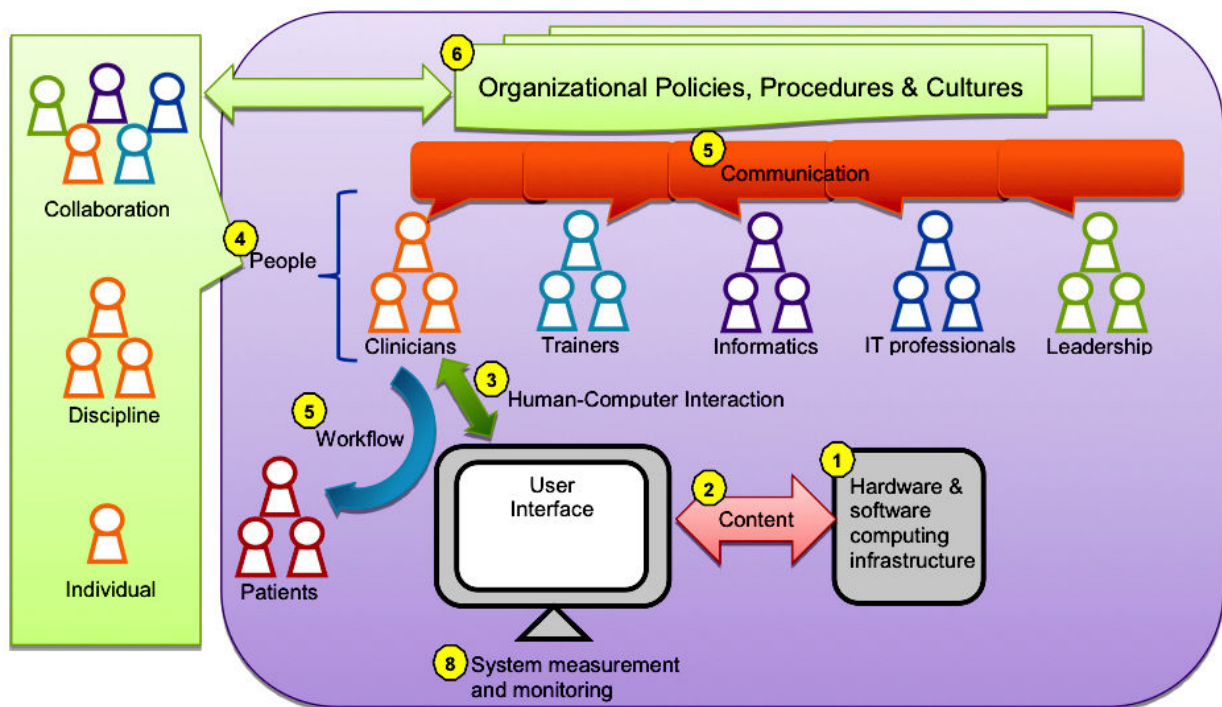
Recommendations for Future Health Information Technology Adaptation Research

We propose that HIT adaptation research should deploy multilevel and multidimensional evaluation to understand the HIT adaptation progress, drawing from both of these foundational theories. Specifically, HIT adaptation research should focus on developing fundamental and multidimensional facts that can inform the progress of HIT adaptation. Below we describe 4 directions that can drive future HIT adaptation research.

Develop Appropriate Process Measures

While the outcome measures (HIT adoption rate, acceptance, and CQMs) have been established, there is a need to develop process measures from individual and organizational perspectives and include multidimensional measures of adaptation to EHRs. These measures will need to incorporate factors such as communication channels, cultural conflict, interdisciplinary team dynamics, user satisfaction, work productivity, cost, and quality [38,59,60].

Consider the Culture and Context in Which Health Information Technology Is Implemented

Most HIT adoption or acceptance studies have used individuals or hospitals as the unit of analysis [39,61,62]. These findings are informative for identifying associated individual perceptions and experiences as well as hospital demographics. However, additional factors such as the culture of a discipline or a department, the interprofessional or multidisciplinary communication within or across departments, the training received, and workflow at the department level have not been discussed. In particular, while social support has been identified as one of the key factors for acceptance [63,64], no studies have been conducted at the department or unit level to study this factor.

Standardize the Definition and Methods for Sociotechnical Studies

Implementing a new technology into a complex environment is often disruptive, particularly in health care. Sociotechnical evaluations of HIT implementations are supported in both theory and empirically; however, little guidance exists in terms of how to conduct a sociotechnical evaluation [65]. Challenges in conducting sociotechnical evaluations include a lack of agreement on the components of the sociotechnical system, possible study designs, and data analysis strategies which may give light to both practical and conceptual challenges [65].

Study Adaptation Longitudinally and Multidimensionally

Processes are more important to study than outcomes because studying processes allows for generalizability and knowledge transfer beyond the clinical setting where the research was conducted [65]. Future studies need to employ longitudinal study designs with multiple data time periods to establish causal relationships [32,66,67]. In addition, the HIT evaluation toolkit proposed by the Agency for Healthcare Research and Quality emphasizes the advantages of conducting mixed methods studies to provide important dimensions in an evaluation study [68]. Thus, future HIT research studies should be designed as mixed methods sociotechnical evaluations focused on exploring the dynamic relationship between technology and social factors over time [65].

Conclusion

Measuring HIT adaptation can provide a more thorough understanding of the connection between HIT use and health care outcomes. Our ability to advance our understanding is predicated on good evaluation models, notably in the area of a health organization's overall performance. As the sociotechnical environment remains a confounding problem influencing our understanding of the generalizability of research findings about HIT implementation success, there is a need to integrate issues exacerbated by workarounds, poorly designed interfaces, suboptimal functionality, and the sheer complexity of systems that contribute to HIT adoption issues as well as consider the idiosyncrasies across contexts. However, existing evaluation models are not supportive of a greater understanding of the phenomenon itself. This paper is therefore presented to provide a new perspective to shift the focus from adoption to adaptation. Future research should develop instruments to evaluate the progress of HIT adaptation in both its longitudinal design and its focus on adaptation progress rather than on a single outcome, allowing for more generalizability and knowledge transfer.

Conflicts of Interest

None declared.

References

1. Thompson T, Brailer D. Health IT Strategic Framework. Washington: US Department of Health and Human Services; 2004.
2. Adler-Milstein J, DesRoches C, Furukawa M, Worzala C, Charles D, Kralovec P. More than half of US hospitals have at least a basic EHR but stage 2 criteria remain challenging for most. Health Aff (Millwood) 2014.
3. DesRoches CM, Charles D, Furukawa MF, Joshi MS, Kralovec P, Mostashari F, et al. Adoption of electronic health records grows rapidly, but fewer than half of US hospitals had at least a basic system in 2012. Health Aff (Millwood) 2013 Aug;32(8):1478-1485 [FREE Full text] [doi: 10.1377/hlthaff.2013.0308] [Medline: 23840052]

4. US Centers for Medicare & Medicaid Services. Electronic health records (EHR) incentive programs 2017. URL: https:/ /www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHRIncentivePrograms [accessed 2017-08-18] [WebCite Cache ID 6sp1G1AHt]

5. DeSalvo KB, Salavitt A. Moving toward improved care through information. 2016. URL: https://www.healthit.gov/buzz-blog/ from-the-onc-desk/moving-improved-care-information/ [accessed 2017-08-18] [WebCite Cache ID 6sp1KbyVJ]

6. Manchikanti L, Helm IS, Calodney AK, Hirsch JA. Merit-based incentive payment system: meaningful changes in the final rule brings cautious optimism. Pain Physician 2017;20(1):E1-E12 [FREE Full text] [Medline: 28072793]

7. Manchikanti L, Benyamin R, Falco F, Hirsch J. Metamorphosis of medicine in the United States: a carrot and stick policy of electronic medical records. Pain Physician 2014;17(6):e671-e680.

8. Manchikanti L, Hirsch JA. A case for restraint of explosive growth of health information technology: first, do no harm. Pain Physician 2015;18(3):E293-E298 [FREE Full text] [Medline: 26000676]

9. Ammenwerth E, Gräber S, Herrmann G, Bürkle T, König J. Evaluation of health information systems—problems and challenges. Int J Med Inform 2003 Sep;71(2-3):125-135. [Medline: 14519405]

10. Singh R, Singh A, Singh DR, Singh G. Improvement of workflow and processes to ease and enrich meaningful use of health information technology. Adv Med Educ Pract 2013;4:231-236 [FREE Full text] [doi: 10.2147/AMEP.S53307] [Medline: 24235855]

11. McAlearney A, Hefner J, Sieck C, Rizer M, Huerta T. Evidence-based management of ambulatory electronic health record system implementation: an assessment of conceptual support and qualitative evidence. Int J Med Inform 2014;83(7):484-494.

12. McAlearney AS, Hefner JL, Sieck C, Rizer M, Huerta TR. Fundamental issues in implementing an ambulatory care electronic health record. J Am Board Fam Med 2015;28(1):55-64 [FREE Full text] [doi: 10.3122/jabfm.2015.01.140078] [Medline: 25567823]

13. McAlearney AS, Song PH, Robbins J, Hirsch A, Jorina M, Kowalczyk N, et al. Moving from good to great in ambulatory electronic health record implementation. J Healthc Qual 2010;32(5):41-50. [doi: 10.1111/j.1945-1474.2010.00107.x] [Medline: 20854358]

14. Studeny J, Coustasse A. Personal health records: is rapid adoption hindering interoperability? Perspect Health Inf Manag 2014;11:1e [FREE Full text] [Medline: 25214822]

15. D'Amore JD, Mandel JC, Kreda DA, Swain A, Koromia GA, Sundareswaran S, et al. Are Meaningful Use Stage 2 certified EHRs ready for interoperability? Findings from the SMART C-CDA Collaborative. J Am Med Inform Assoc 2014;21(6):1060-1068 [FREE Full text] [doi: 10.1136/amiajnl-2014-002883] [Medline: 24970839]

16. Ford E, Pettit L, Silvera G, Huerta T. EMR implementation speed: patient safety culture paradox. J Healthc Inform Manage 2014;28(3):24-31.

17. Sakata K, Stephenson L, Mulanax A, Bierman J, McGrath K, Scholl G. Professional and interprofessional differences in electronic health records use and recognition of safety issues in critically ill patients. J Interprof Care 2016:1-7.

18. Lorenzi N, Novak L, Weiss J, Gadd C, Unertl K. Crossing the implementation chasm: a proposal for bold action. J Am Med Inform Assoc 2008;15(3):290-296.

19. Afonso A, Alfonso S, Morgan T. Short-term impact of meaningful use stage 1 implementation: a comparison of health outcomes in 2 primary care clinics. J Ambul Care Manage 2017.

20. Grinspan Z, Bao Y, Edwards A, Johnson P, Kaushal R, Kern L. Medicaid stage 1 meaningful use EHR incentive payments are associated with higher quality but not improvements in quality. Am J Med Qual 2016 Oct 13. [doi: 10.1177/1062860616673905] [Medline: 27738129]

21. Levine DM, Healey MJ, Wright A, Bates DW, Linder JA, Samal L. Changes in the quality of care during progress from stage 1 to stage 2 of Meaningful Use. J Am Med Inform Assoc 2017 Dec 01;24(2):394-397. [doi: 10.1093/jamia/ocw127] [Medline: 27567000]

22. Kaplan B. Evaluating informatics applications—some alternative approaches: theory, social interactionism, and call for methodological pluralism. Int J Med Inform 2001 Nov;64(1):39-56. [Medline: 11673101]

23. Wears RL, Berg M. Computer technology and clinical work: still waiting for Godot. JAMA 2005 Mar 09;293(10):1261-1263. [doi: 10.1001/jama.293.10.1261] [Medline: 15755949]

24. Ward R. The application of technology acceptance and diffusion of innovation models in healthcare informatics. Health Policy and Technology 2013;2:222-228.

25. US Centers for Medicare & Medicaid Services. Step 5: achieve meaningful use stage 2. 2014. URL: https://www.healthit.gov/ providers-professionals/step-5-achieve-meaningful-use-stage-2 [accessed 2017-08-18] [WebCite Cache ID 6sp2WB4bu]

26. Botta MD, Cutler DM. Meaningful use: floor or ceiling? Healthc (Amst) 2014 Mar;2(1):48-52. [doi: 10.1016/j.hjdsi.2013.12.011] [Medline: 26250089]

27. Nakamura MM, Harper MB, Castro AV, Yu FB, Jha AK. Impact of the meaningful use incentive program on electronic health record adoption by US children's hospitals. J Am Med Inform Assoc 2015 Mar;22(2):390-398. [doi: 10.1093/jamia/ocu045] [Medline: 25755126]

28. HIMSS Analytics. EMR adoption model. 2009. URL: http://www.himssanalytics.org/hc_providers/emr_adoption.asp [accessed 2017-08-18] [WebCite Cache ID 6sp2k3f5a]

29.   Phichitchaisopa N, Naenna T. Factors affecting the adoption of healthcare information technology. EXCLI J 2013;12:413-436 [FREE Full text] [Medline: 26417235]

30.   Venkatesh V, Sykes T, Zhang X. "Just what the doctor ordered:" a revised UTAUT for EMR system adoption and use by doctors. 2011 Presented at: Proceedings of the 44th Hawaii International Conference on System Sciences; 2011; Kauai.

31.   Patel V, Jamoom E, Hsiao C, Furukawa M, Buntin M. Variation in electronic health record adoption and readiness for meaningful use: 2008-2011. J Gen Intern Med 2013 Jul;28(7):957-964 [FREE Full text] [doi: 10.1007/s11606-012-2324-x] [Medline: 23371416]

32.   Steininger K, Stiglbauer B, Baumgartner B, Engleder B. Factors explaining physicians' acceptance of electronic health records. 2014 Presented at: 47th Hawaii International Conference on System Science; 2014; Waikoloa.

33.   Liu Z, Min Q, Ji S. A comprehensive review of research in IT adoption. 2008 Presented at: 4th International Conference on Wireless Communications, Networking and Mobile Computing; 2008; Dalian.

34.   Davis F. User acceptance of information technologyystem characteristics, user perceptions, and behavioral impacts. Int J Man Mach Stud 1993;38(3):475-487.

35.   Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):318-340.

36.   Gangwar H, Date H, Raoot A. Review of IT adoption: insights from recent technologies. J Enterprise Inform Manag 2014;27(4):488-502.

37.   Center for Medicare and Medicaid Services. Clinical quality measures. 2014. URL: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/2014_ClinicalQualityMeasures.html [accessed 2017-08-18] [WebCite Cache ID 6sp3iQexe]

38.   Spetz J, Burgess JF, Phibbs CS. The effect of health information technology implementation in Veterans Health Administration hospitals on patient outcomes. Healthc (Amst) 2014 Mar;2(1):40-47. [doi: 10.1016/j.hjdsi.2013.12.009] [Medline: 26250088]

39.   Agha L. The effects of health information technology on the costs and quality of medical care. J Health Econ 2014 Mar;34:19-30 [FREE Full text] [doi: 10.1016/j.jhealeco.2013.12.005] [Medline: 24463141]

40.   Robertson A, Cresswell K, Takian A, Petrakaki D, Crowe S, Cornford T. Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. Brit Med J 2010:341.

41.   Wang Y, Gu J, Zhou Z. Complex adaptive socio-technical systems theory view of ubiquitous computing systems research. 2014 Presented at: Ubiquitous Computing and Ambient Intelligence Personalisation and User Adapted Services; 2014; Belfast p. 216-223.

42.   Sittig D, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Safety Healthc 2010;19:68-74.

43.   Ellis B, Herbert S. Complex adaptive systems (CAS): an overview of key elements, characteristics and application to management theory. Inform Prim Care 2011;19(1):33-37.

44.   Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care—an interactive sociotechnical analysis. J Am Med Inform Assoc 2007;14(5):542-549 [FREE Full text] [doi: 10.1197/jamia.M2384] [Medline: 17600093]

45.   Ellis B. Complexity in practice: understanding primary care as a complex adaptive system. Inform Prim Care 2010;18(2):135-140.

46.   Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012;36(1):241-277.

47.   Patel V, Kannampallil T. Human factors and health information technology: current challenges and future directions. Yearb Med Inform 2014;9(1):58-66.

48.   Coiera E. Putting the technical back into socio-technical systems research. Int J Med Inform 2007;76:S98-S103.

49.   Balka E, Tolar M, Coates S, Whitehouse S. Socio-technical issues and challenges in implementing safe patient handovers: insights from ethnographic case studies. Int J Med Inform 2013;82(12):e345-e357.

50.   Ludwick D, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009;78(1):22-31.

51.   Kannampallil T, Schauer G, Cohen T, Patel V. Considering complexity in healthcare systems. J Biomed Inform 2011;44(6):943-937.

52.   Majchrzak A, Rice R, Malhotra A, King N. Technology adaptation: the case of a computer-supported inter-organizational virtual team. MIS Quarterly 2000;24(4):569-600.

53.   Cooper R, Zmud R. Information technology implementation research: a technological diffusion approach. Manag Sci 1990;36:123-139.

54.   Center for Medicare and Medicaid Services. EHR Incentive Programs. 2014. URL: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Stage_2.html [accessed 2017-08-18] [WebCite Cache ID 6sp4YJfxD]

XSL•FO

RenderX

55.   Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. New Engl J Med
       2010;363(6):501-504.
56.   Marcotte L, Seidman J, Trudel K, Berwick D, Blumenthal D, Mostashari F. Achieving meaningful use of health information
       technology: a guide for physicians to the EHR incentive programs. Arch Intern Med 2012;172(9):731-736.
57.   Donabedian A. The quality of medical care. Science 1978 May 26;200(4344):856-864. [Medline: 417400]
58.   Kukafka R, Johnson S, Linfante A, Allegrante J. Grounding a new information technology implementation framework in
       behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform 2003;36(3):218-227.
59.   eHealth Observatory and School of Health Information Science University of Victoria. Benefits evaluation indicators,
       technical report. 2012. URL: https://www.infoway-inforoute.ca/en/component/edocman/
       450-benefits-evaluation-indicators-technical-report-version-2-0/view-document?Itemid=101 [accessed 2017-08-18] [WebCite
       Cache ID 6sp5635Ep]
60.   Agency for Healthcare Research and Quality. Health IT evaluation measures. 2014. URL: https://healthit.ahrq.gov/
       health-it-tools-and-resources/evaluation-resources/health-it-evaluation-toolkit-and-evaluation-measures-quick-reference
       [accessed 2017-08-18] [WebCite Cache ID 6sp58tchb]
61.   Jamoom E, Patel V, Furukawa M, King J. EHR adopters vs non-adopters: impacts of, barriers to, and federal initiatives for
       EHR adoption. Healthcare 2014;2:33-39.
62.   Price M, Singer A, Kim J. Adopting electronic medical records: are they just electronic paper records? Canadian Fam
       Physician 2013;59(7):e322-e329.
63.   Ash J, Sittig D, Guappone K, Dykstra R, Richardson J, Wright A. Recommended practices for computerized clinical decision
       support and knowledge management in community settings: a qualitative study. BMC 2012;12:6.
64.   Hsiao J, Chang H, Chen R. A study of factors affecting acceptance of hospital information systems: a nursing perspective.
       J Nurs Res 2011;19(2):150-160.
65.   Cresswell KM, Sheikh A. Undertaking sociotechnical evaluations of health information technologies. Inform Prim Care
       2014;21(2):78-83 [FREE Full text] [doi: 10.14236/jhi.v21i2.54] [Medline: 24841408]
66.   Hameed M, Counsell S, Swift S. A conceptual model for the process of IT innovation adoption in organizations. J Engineer
       Technol Manag 2012;29(3):358-390.
67.   Spaulding A, Gamm L, Kim J, Menser T. Multiproject interdependencies in health systems management: a longitudinal
       qualitative study. Health Care Manag Rev 2014;39(1):31-40.
68.   Agency for Healthcare Research and Quality. Health Information Technology Evaluation Toolkit. 2009. URL: https:/
       /healthit.ahrq.gov/sites/default/files/docs/page/Evaluation%20Toolkit%20Revised%20Version.pdf [accessed 2017-08-18]
       [WebCite Cache ID 6sp5r81OY]

Abbreviations

CAS: complex adaptive system
CQM: clinical quality measure
EMRAM: electronic medical record adoption model
HER: electronic health record
HIMSS: Healthcare Information and Management Systems Society
HIT: health information technology
IT: information technology
MU: Meaningful Use
TAM: technology acceptance model

[XSL•FO](#)
RenderX

Original Paper

DynAMo: A Modular Platform for Monitoring Process, Outcome, and Algorithm-Based Treatment Planning in Psychotherapy

Tim Kaiser[1], MSc; Anton Rupert Laireiter[1], PhD

Psychotherapy Research Group, Department of Psychology, University of Salzburg, Salzburg, Austria

Corresponding Author:
Tim Kaiser, MSc
Psychotherapy Research Group
Department of Psychology
University of Salzburg
2nd Fl.
Hellbrunnerstasse 34
Salzburg, 5020
Austria
Phone: 43 69917305064
Fax: 43 662804451
Email: Tim.Kaiser@sbg.ac.at

Abstract

Background: In recent years, the assessment of mental disorders has become more and more personalized. Modern advancements such as Internet-enabled mobile phones and increased computing capacity make it possible to tap sources of information that have long been unavailable to mental health practitioners.

Objective: Software packages that combine algorithm-based treatment planning, process monitoring, and outcome monitoring are scarce. The objective of this study was to assess whether the DynAMo Web application can fill this gap by providing a software solution that can be used by both researchers to conduct state-of-the-art psychotherapy process research and clinicians to plan treatments and monitor psychotherapeutic processes.

Methods: In this paper, we report on the current state of a Web application that can be used for assessing the temporal structure of mental disorders using information on their temporal and synchronous associations. A treatment planning algorithm automatically interprets the data and delivers priority scores of symptoms to practitioners. The application is also capable of monitoring psychotherapeutic processes during therapy and of monitoring treatment outcomes. This application was developed using the R programming language (R Core Team, Vienna) and the Shiny Web application framework (RStudio, Inc, Boston). It is made entirely from open-source software packages and thus is easily extensible.

Results: The capabilities of the proposed application are demonstrated. Case illustrations are provided to exemplify its usefulness in clinical practice.

Conclusions: With the broad availability of Internet-enabled mobile phones and similar devices, collecting data on psychopathology and psychotherapeutic processes has become easier than ever. The proposed application is a valuable tool for capturing, processing, and visualizing these data. The combination of dynamic assessment and process- and outcome monitoring has the potential to improve the efficacy and effectiveness of psychotherapy.

Introduction

Background

One of the major strategic objectives of the National Institute of Mental Health is to develop ways to tailor existing and new interventions to optimize outcomes and to foster personalized interventions and strategies for sequencing or combining existing and novel interventions [1]. In practice, this can be accomplished by monitoring individual trajectories of change instead of assuming similar treatment responses for every patient. Promising advances have been made in psychotherapy research

toward adhering to this goal, leading to the propagation of scientifically informed clinical practice. In the last 20 years, various process and outcome monitoring systems have been developed [2]. One of these well-established outcome monitoring systems is the Outcome Questionnaire-45.2 (OQ-45.2) [3,4]. Using weekly post-session assessments, this system screens for therapeutic change in the domains of reduced symptom distress, interpersonal functioning, and social role. Additionally, potential for risk factors such as suicidal tendencies, substance abuse, and violence is screened for. The OQ system includes a software application called OQ Analyst. This application makes it possible to administer and evaluate routine outcome questionnaires in psychotherapy; it also enables clinicians to check if the outcomes of a current patient are satisfying, or if a revision of the treatment plan is necessary. If a patient's outcome is not within the expected range, the software warns the clinician.

The Partners for Change Outcome Management System (PCOMS) [5] also uses outcome ratings that are administered before every session. Furthermore, the quality of the therapeutic alliance is rated after sessions using the Session Rating Scale.

These systems mostly provide mental health practitioners with valuable information on the outcomes of their clinical interventions and on the risk that may worsen a patient's condition. By incorporating this information, practitioners may correct their treatment plan when the patient's condition appears to be "off track." In addition, patients get motivated by noticing that their improvements are actually measurable. Thus, it is not surprising that systematic monitoring of relevant variables has been found to not only prevent negative treatment outcomes but also to enhance positive ones [6].

A successful psychotherapy continues to show positive outcomes after most of the therapy sessions and also between sessions. Over time, patients begin to form mental representations of their psychotherapy that can be activated between therapy sessions. Following this assumption, Orlinsky et al [7] were the first to propose the "representations of a patient's psychotherapy between two sessions" as a focus for psychotherapy research, building the foundation of what later became the concept of the intersession process. The intersession process encompasses thoughts, feelings, and behaviors concerning a patient's current psychotherapy, including the therapist. They occur between two therapy sessions; can be of varying emotional quality, intensity, and frequency; and include memories of words and feelings toward the dyadic partner, applying techniques learned in the psychotherapy or doing therapeutic homework. The intersession process has been operationalized and can be measured for a specific period of time between two sessions using the Intersession Experience Questionnaire (IEQ) [8]. The positive intersession experience of patients has been positively linked to therapy outcome variables in various studies [9], and it has been found to be predictive of therapy outcome using weekly retrospective measures [10-12].

More fine-grained monitoring systems such as the Synergetic Navigation System (SNS) [13] not only measure outcomes but focus on data concerning the intersession process. Ubiquitous Internet access enables such systems to draw their data from the Ecological Momentary Assessment (EMA). This approach bears detailed information that routine outcome monitoring is by design unable to detect, thereby making it possible to target interventions or to recognize the specific areas of psychotherapy that need more attention. If, for example, a patient reports a drop in perceived therapeutic relationship quality in between sessions, the therapist should focus on improving this quality in the next session. For patients, there are clear benefits as well. Their perception of problematic thoughts and feelings is trained, possibly leading to a more mindful processing of their daily experience. Problematic thoughts and feelings are validated during feedback sessions with their therapists, which may also improve the quality of the therapeutic relationship. An in-depth discussion of the advantages and possible caveats of this approach were discussed by the authors of the SNS [13].

Dynamic Modeling of Psychopathology

Another area of psychotherapy in which monitoring systems for mental health practitioners may be important but are not yet as widespread, is diagnosis. Standard methods for clinical diagnosis of mental disorders are cross-sectional. Though diagnostic criteria for most disorders require a manifestation of symptoms over a certain period of time, ranging from weeks to several months, patients are generally assessed retrospectively at one specific point in time. Results are then typically compared with those of other patients or a specific "cutoff" score. These diagnoses are useful for classification, but they offer only limited guidance for planning interventions. Thus, individualized case formulations of psychopathology are frequently used in most psychotherapeutic orientations [14-16], offering a more complex view on individual cases and allowing practitioners to choose specific interventions.

Methodologically, this bears some problems. These methods only try to approximate temporal associations and causalities using subjective retrospective data. Recently, attempts have been made to base diagnostics on data collected in real time using EMA. For example, Fisher [17] was able to build highly individualized disorder models that were used for prescriptive treatment decisions. This approach follows four steps. First, an inventory of test items for measuring all relevant aspects of a disorder is compiled. This typically includes self-report symptom scales. Then, these items are administered using intensive repeated measurement. In practice, daily questionnaires assessing symptom severity are used. The assessment frequency can be increased to several times a day. This step results in multivariate time series data on a patient's psychopathology. To determine the synchronous associations of symptoms, factor-analysis methods are applied. The "P-technique" factor analysis is the most common approach for this step [18]. In the final step, the time-dependent relationships are assessed using multivariate time series analysis methods such as vector autoregressive (VAR) modeling. Following this diagnostic approach, Fisher thereupon presented first attempts to plan therapeutic interventions [19] based on individualized assessment.

Objective

Until now, only commercial process and outcome monitoring applications exist. Their implementation can be costly, especially

XSL•FO

RenderX

for private practices. Commercial applications cannot be extended with new functionalities by third persons. Instead, new features have to be requested from the developing company, which in turn can take considerable amount of time, depending on how much development effort is put into the requested functionalities. This makes it hard for researchers to fit the applications to their needs. Free software released under an open-source license makes the source code of the application publicly available, guaranteeing easy extension by developers who are interested in participating.

Hence, our objective was to develop an application containing the latest advancements in psychotherapy process research and dynamic assessment of psychopathology using open-source software. To accomplish this objective, we created the DynAMo (short for "dynamic assessment and modeling") Web application. The results of this development process are presented in this paper to inform the researchers and clinicians of the results of this process.

Methods

The DynAMo Web Application

The presented application is based on community-driven, open-source software packages. One of its main uses is treatment planning based on the diagnostic approach that has been described previously. Treatment planning begins before the actual psychotherapy starts; hence, this function is directed at persons with mental health problems intending to seek treatment. The treatment planning function is based on an algorithm that generates patient-specific models of psychopathology from data collected in real time. These patient-specific models can be considered as an important step to a methodologically more sound and a more individualized view of psychopathology and, in practice, to treatment approaches that become more effective by being tailored to a single patient's needs. The application can also be used for monitoring processes and outcomes in psychotherapy. The DynAMo application consists of multiple modules that can run in combination but also independently. The collection of data is accomplished by a data assessment module. The treatment planning algorithm uses the collected data to generate actionable information for targeting interventions. The practitioner interface is used by clinicians to access and inspect the collected data. Researchers can use this interface to examine data collected in psychotherapy process studies.

Data Assessment Module

The questionnaires can be designed freely, using user interface elements from the Shiny Web framework [20]. Using the mirtCAT (computerized adaptive testing with multidimensional

item response theory) package for Gnu R [21], this module of the DynAMo Web application is able to automatically send Web links to personal questionnaires at preset times. At this time, this is possible either via email or text message. Messages include a URL that leads to the questionnaire page. An example item is depicted in Figure 1. Every item has to be answered and is completed by clicking the *Next* button. After completing a questionnaire, the collected data are saved to the server. All data transmitted to and from the application are encrypted using Transport Layer Security 1.2. Data collected in the assessment are stored without any person-related data. In the configuration file, a patient code chosen by the therapist can be entered for later identification. This patient code is used for naming the database entries, so that the therapist can identify his patients when loading data. Other patient data have to be stored externally.

General Approach to Data Collection Using DynAMo

Before a person with intention to treat can begin dynamic assessment, he/she has to meet with his/her clinician for initial diagnostic screening and an introduction to the assessment system. This is necessary to determine which symptoms should be included in the daily measures. Basically, items for the main diagnosis and any comorbidity should be included. Because of the novelty of the dynamic assessment method, there are no recommended scales. However, every self-report measure of psychopathology that is reliable and valid can be used. For many disorders, short questionnaires exist, and in the current Diagnostic and Statistical Manual of Mental Disorders, 5th edition, the American Psychiatric Association is offering a number of disorder-specific measures; for example, for various anxiety disorders, posttraumatic stress disorder, depression, acute stress symptoms, and dissociative experiences [22]. Additionally, the mirtCAT module used in this application allows to administer computer-adaptive questionnaires following item response theory (IRT). However, DynAMo currently has no option to administer psychological tests following IRT.

All items to be selected for daily assessments are then compiled to one questionnaire and included in the patient's configuration file. If daily questionnaires are used, assessments should be scheduled for the evening. Thereby, the patient can retrospectively estimate his moods and symptoms experienced during the past day. If multiple assessments per day are planned, it is recommended to schedule them with equal temporal distances. This can be achieved by asking patients about their regular sleeping and waking times and splitting the resulting time window into equal parts. Generally, multiple assessments per day should be preferred, because higher assessment frequencies ensure a more fine-grained dataset that includes daily variation.

Figure 1. Data assessment interface on a mobile phone.

Treatment Planning Algorithm

The key feature of the DynAMo application is the algorithm used for estimating the impact of treatment when a specific symptom is targeted. This information is distilled from the time series data collected from patients before starting their treatment. The algorithm provides priority scores for each assessed symptom. The algorithm analyzes data from all patients currently participating in a dynamic assessment procedure. If a model was generated successfully, notifications can be sent out to practitioners or researchers via email. These notifications can optionally include a table of symptom priority scores, so it is optional to use the practitioner interface.

Though the fundamental structure of the treatment planning algorithm was proposed by Fisher [17], it has been refined and automatized for use in DynAMo. Practitioners can use the practitioner interface to check if a model has already been identified. Every time this is checked, the algorithm will run, trying to identify a disorder model with satisfactory fit indices with the data that have been collected. The number of measurements required depends on the number of items administered and has great interindividual variation. Though in the authors' experience 40 to 60 measurements are sufficient for some patients, this number was set to about 120 by other researchers [19]. In terms of time, exposure assessments can take from 2 to 6 weeks. The steps of the algorithm are explained in the following sections. For exemplary R code, including example data, see Multimedia Appendix 1.

Exploratory P-Factor Analysis

The first step in assessing the correlational structure in a patient's multivariate symptom time series is factor analysis. This is done to identify the latent dimensional structure of a patient's disorder. A maximum likelihood factor analysis is conducted with the collected time series data. Oblimin rotation is applied, because the latent dimensions of psychopathology are expected to show intercorrelation. Three models are generated simultaneously, assuming two, three, or four latent factors. Models including more than four factors are theoretically conceivable but have not yet been implemented in this application. If goodness-of-fit statistics reach satisfactory levels for any of these models, the algorithm proceeds to the next step. Goodness-of-fit measures for all factor analysis steps have been chosen according to the cutoff criteria proposed by Hu and Bentler [23]. These criteria have been determined in simulations and were found to minimize the risk of both an over- and underestimation of model fit. If more than one model reaches satisfactory goodness-of-fit measures, the model with the smallest number of factors is selected.

Confirmatory P-Factor Analysis

A factor loading matrix is extracted from the exploratory model. This matrix gets converted to a structural equation system that can be tested using the *lavaan* [24] package for Gnu R. Only items with factor loadings greater than .30 are included in this equation system. This step was introduced to increase statistic rigor in the algorithm. If the exploratory model is confirmed as indicated by fit measures, the algorithm proceeds to the next step.

Vector Autoregressive (VAR) Modeling

Vector autoregressive models are fit using the *vars* package [25] for Gnu R, following an approach first used in econometry by Lütkepohl [26]. Its use has been increasingly common in psychiatry and psychotherapy [27-30]. An application of this approach to VAR modeling to psychotherapeutic process data, including an in-depth description and basic R and Statistical Analysis System code, was provided by Ramseyer et al [30].

Factor scores for every point in assessment are generated by multiplying the raw data matrix with a weighting matrix obtained by confirmatory factor analysis (CFA). This results in a multivariate time series of factor scores. To this time series,

VAR modeling is applied to determine the associations between the extracted factors. The parameters obtained in this step are:

1. Autoregressive parameters. This is a regression parameter that describes how strongly the value of a factor at one point in time (time $t-1$) is associated with the value of the same factor at a later point in time (t). Factors with high autoregressive parameters are relatively stable over time.

2. Cross-regressive parameters. This describes the intensity of the association of a certain factor at one point of time with another factor measured later on.

3. Linear trends. Positive or linear trends are more likely to be observed in longer time series. They indicate that the mean of the respective factor changes over time.

4. Synchronous associations. This refers to the correlation between two factors at one point in time, indicated by a correlation coefficient.

5. Whereas only auto- and cross-regressive parameters are directly relevant for the next steps, the risk to overestimate them is reduced by including the variance explained by linear trends in the model.

Factor Scores

The relative amount of variance (expressed in percentage) explained by each factor in the confirmatory factor model is multiplied with the relative amount of variance explained by this factor's auto- and cross-regressive parameters in the VAR model. When all factors are scored, the raw scores are standardized by dividing their scores by the score of the highest-scoring factor. This is done to determine the relevance of a patient's latent dimensions of psychopathology. If one factor explains 60% of variance and the other only 10%, then the symptoms associated with the first factor can be considered predominant. Also, if a factor strongly influences other factors, treating symptoms associated with it will more likely have beneficial effects of symptoms associated with other factors.

Symptom Scores

Symptom ratings are averaged and standardized by dividing each symptom's mean rating by the highest symptom mean. Means are then multiplied by the factor scores the items belong to and their loadings on this factor. Symptom scores are then standardized just like factor scores. These scores now contain a lot of information relevant for treatment planning. First, high-scoring items indicate that treating them first will result into the strongest decrease of subjective distress, because they are more likely to belong to a high-scoring factor. Second, because the item scores contain information on time-lagged associations, treating them will most likely affect other

symptoms as well, as they are more likely to explain a greater amount of variance in the VAR model.

Results

Clinical Example for the Treatment Planning Algorithm

In order to exemplify the treatment planning algorithm, data recorded from a 30-year-old male patient currently in treatment for social anxiety disorder were used to illustrate the steps that have been described in the previous section. The patient completed a 10-item dimensional scale measuring social anxiety symptom severity [31] three times a day, at 8:00 AM, 2:00 PM, and 8:00 PM, respectively, for 5 weeks, resulting in 103 data points. The items administered are listed in Table 1.

Step 1: Exploratory Factor Analysis (EFA)

The algorithm is programmed to find an exploratory factor model using the least number of factors while still fulfilling the fit criteria. In this case, a two-factor model was found. The factor analysis was conducted as described in the previous section. Table 2 shows the factor loadings determined. Due to oblimin rotation, correlations between factors were allowed in the model. Factors were correlated with $r=.56$.

This factor solution suggested that the latent structure of this patient's disorder consists of one factor mainly driven by anxiety and to some extent, by fear and avoidance, whereas items loading on the second factor described not only physical symptoms such as a racing heart and muscle tension, but also distraction and avoidance. In exploratory models, the algorithm checks the Tucker-Lewis-Index (TLI) and the root mean squared residual (RMS). Both measures were adequate for this model (TLI=.958, RMS=.056).

Step 2: Confirmatory Factor Analysis

Model terms for CFA are determined by the algorithm according to the EFA model structure. Only items with factor loadings greater than .30 are included in the CFA model. In this case, one item was excluded from the model due to insufficient factor loadings. The structure that is to be confirmed can be represented by the following two terms:

Factor 1 = Item 1 + Item 2 + Item 6

Factor 2 = Item 1 + Item 3 + Item 4 + Item 5 + Item 6 + Item 7 + Item 8 + Item 9

Factor loadings from the CFA model can be found in Table 3. Also, as in EFA, correlations between the factors were allowed in this model. In the CFA model, factor scores were correlated with $r=.243$.

XSL•FO

RenderX

Table 1. List of items in the social anxiety questionnaire administered by the example patient.

Item number	Item text, prefixed by "Since the last assessment, I have..."
1	Felt moments of sudden terror, fear, or fright in social situations.
2	Felt anxious, worried, or nervous about social situations.
3	Had thoughts of being rejected, humiliated, embarrassed, ridiculed or offending others.
4	Felt a racing heart, sweaty, trouble breathing, faint, or shaky in social situations.
5	Felt tense muscles, felt on edge or restless, or had trouble relaxing in social situations.
6	Avoided, or did not approach or enter, social situations.
7	Left social situations early or participated only minimally (eg, said little or avoided eye contact)
8	Spent a lot of time preparing what to say or how to act in social situations.
9	Distracted myself to avoid thinking about social situations.
10	Needed help to cope with social situations (eg, with alcohol, medications, or superstitious objects).

Table 2. Factor loadings in the exploratory model. Loadings smaller than .10 were omitted.

Item number	Factor 1	Factor 2
1	.31	.40
2	.97	-
3	-	.58
4	-	.57
5	-	.65
6	.34	.40
7	-	.43
8	-	.49
9	-	.56
10	-	-

Table 3. Factor loadings resulting from confirmatory factor analysis.

Item number	Factor 1	Factor 2
1	.569	.129
2	.810	-
3	-	.624
4	-	.519
5	-	.657
6	.626	.094
7	-	.368
8	-	.670
9	-	.437

The TLI and the standardized root mean squared residual (SRMR) is checked for confirmatory models. Both measures were adequate for this model (TLI=.959, SRMR=.05). Thus, the model found with EFA was confirmed with increased statistical rigor, and the algorithm could proceed.

Step 3: Vector Autoregressive (VAR) Model

For this step, factor scores are extracted from the time series by multiplying the raw scores with the factor loading matrix. In this example, this resulted in a time series of 103 points of measurement for each factor. From these time series, a VAR model is computed to determine time-lagged associations between factor scores. The number of lags is determined by comparing the Akaike information criterion (AIC) for models with one to five lags, choosing the number that leads to the lowest AIC value. In this case, four lags were chosen. Regression models for both factors explained a statistically

XSL•FO
RenderX

significant amount of variance (R^2=.294, $F_{9,89}$=5.527, P=5.011^{-6}) for the first factor and (R^2=.314, $F_{9,89}$=5.979, P=1.63^{-6}) for the second factor. Figure 2 shows a graphical depiction of the model. The VAR model shows that both factors show time-lagged associations with factor 1. From this model, it can be concluded that symptoms associated with the first factor have significant influence on the severity of symptoms associated with the second factor. In this example, an increased level of anxiety, worry, or nervosity about social situations as well as sudden fright and terror in social situations and avoidance were associated with the first factor. If these symptoms increase, symptoms associated with the second factor (an increased level of somatic symptoms such as muscle tension, a racing heart and sweat, or more cognitive symptoms such as preparation for social situations or thoughts about being ridiculed or humiliated) are more likely to increase. Because there are significant vector regressive parameters from the three lags and the patient completed a symptoms questionnaire three times a day, this influence was measurable from measurements up to 24 hours ago. The first factor also has a significant auto-regressive component, meaning that symptoms associated with the first factor are more stable over time.

Figure 2. Vector autoregressive model, including auto- and cross-regressive parameters (one-headed arrows) and a synchronous association (double-headed arrows). Only statistically significant parameters are shown. Explained variance is indicated by R-squared values.
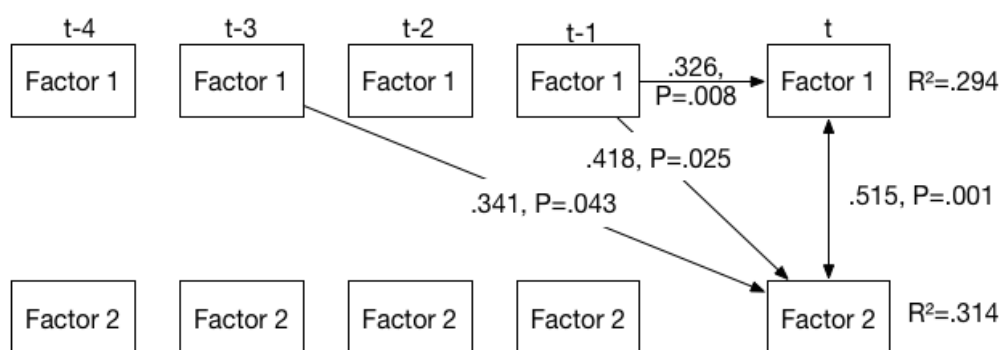
Step 4: Factor Scores

After the time-lagged associations between factor values have been determined, the algorithm proceeds to determine the relevance of the two factors for the patient's psychopathology. First, the squared loadings of each item in the CFA model are divided by the number of factors and summed up. The result is the explained amount of variance for the respective factor. This value represents the amount of variance a factor explains at one point in time. The same is done with the auto- and cross-regressive parameters in the VAR model, leading to a value indicating the amount of explained variance between several points in time. In this case, the first factor explained 34.27% of within-time variance and 22.96% between-time variance. The second factor explained 47.27% within-time variance and 6.1% of between-time variance. Both types of variance are then multiplied and normalized by dividing them by the largest resulting value. The two factor scores resulting from this step were 1 and .367.

Step 5: Symptom Scores

In the last step, means for each item from the social anxiety scale are calculated and normalized by dividing them by the largest item mean. Then, each item's factor loading is multiplied with the factor score. If an item loads on more than one factor, this is done for every factor the item loads upon. The resulting value is multiplied with the item's normalized mean. Now, every item has a symptom score containing information about their average severity, their contribution to each factor, and the relevance of the factor the item contributes to. Finally, these item scores are normalized by dividing them by the highest item score and multiplying them by 100. The result is a priority rating for each item. Table 4 shows the items and their symptom scores.

Table 4. Symptom scores resulting from the treatment planning algorithm, sorted by maximum to minimum priority.

Item text	Normalized symptom score
Felt anxious, worried, or nervous about social situations.	100
Spent a lot of time preparing what to say or how to act in social situations.	95.43
Felt a racing heart, sweaty, trouble breathing, faint, or shaky in social situations.	44.13
Avoided, or did not approach or enter, social situations.	41.58
Felt moments of sudden terror, fear, or fright in social situations.	36.78
Had thoughts of being rejected, humiliated, embarrassed, ridiculed or offending others.	29.86
Felt tense muscles, felt on edge or restless, or had trouble relaxing in social situations.	28.61
Left social situations early or participated only minimally (eg, said little, avoided eye contact)	23.80
Distracted myself to avoid thinking about social situations.	21.21

The result obtained from applying the treatment planning algorithm to the patient's data suggests that the treatment targets the first two symptoms listed in table 4: anxiety, worry and nervous feelings about social situations and the excess amount of time that the patient spends preparing for social situations. This could be done by combining a relaxation exercise with imaginal exposure techniques. Self-control desensitization [32] is a well-validated approach that would cover this combination. Note that, after treatment has proceeded for some time and the recording of symptom scores is continued, a new model might be found by the algorithm, prioritizing different symptoms and thus, suggesting a change in the treatment plan.

Practitioner Interface

The practitioner interface was developed using the Shiny Web framework. This allows combining the powerful statistical computing and plotting capabilities of the R programming language with a well-developed Web framework. The application consists of two modules that can run independently: the practitioner interface designed for convenient data interpretation and the assessment interface that delivers questionnaires to patients. Practitioners can review patient data using a convenient and intuitive Web interface.

Treatment Planning Information

The summary of item scores produced by the treatment planning algorithm can be inspected in the "Treatment Planning" tab. A table (similar to Table 1) with all symptoms that have been assessed is shown, including a standardized priority score that results from the treatment planning algorithm.

Time-Series Inspection

Individual time series for every item assessed by DynAMo can be inspected via the practitioner interface. As depicted in Figure 2, different auxiliary plots are available:

- A local regression curve graph with 95% CI for easy interpretation of shifts in means.

- A plot of the *dynamic complexity* of the time series. Dynamic complexity is composed of the intensity of fluctuation and the degree of distribution of values in a moving window of a time series [33]. Typically, this window has a width of 5 to 7 data points. Flat curves with a few different values result in low dynamic complexity, whereas curves that oscillate strongly and include a variety of different values result in high complexity. Local peaks of dynamic complexity indicate critical instabilities of a system, which are likely to be accompanied by sudden changes in the system's components, so-called *phase transitions* [34]. Applied to psychotherapy, critical instabilities often precede symptom changes if "boundary conditions" such as a positive therapeutic relationship are fulfilled [35]. Hence, this plot offers additional information that can be interpreted alongside the raw data curve. In this application, the dynamic complexity is rescaled to fit the theoretical maximum and minimum of the raw plot.

- As illustrated in Figure 3, the application is able to generate the so-called *recurrence plots* [36] to further ease the interpretation of time series for the trained user. Recurrence plots visualize the Euclidean distance between points in a time series so that the recurring patterns in time series data become more obvious. This is achieved by plotting the Euclidean distance in time x time diagrams. Similar to dynamic complexity, recurrence plots are used to identify phases of critical instability. With recurrence plots, rare or far-from-normal states in a time series can be identified, which are indicative of occurring phase transitions. Recurrence plots are optional and not plotted by default in this application. However, their use in the monitoring of psychotherapy processes is quite common in comparable applications [13].

Summarized, when using DynAMo, practitioners can extract a large amount of information on their patients by analyzing their time series. No complicated mathematical operations have to be carried out by the practitioners themselves because all of the information contained in a time series is represented graphically.

XSL•FO

RenderX

Figure 3. Time series plot showing answers to an intersession process item. A line smoothed by local regression scatterplot smoothing (LOESS) is added for easy interpretation of long-term change, including the 95% CI. The red line represents the measure of dynamic complexity. The x-axis represents the date of measurement (day and month).
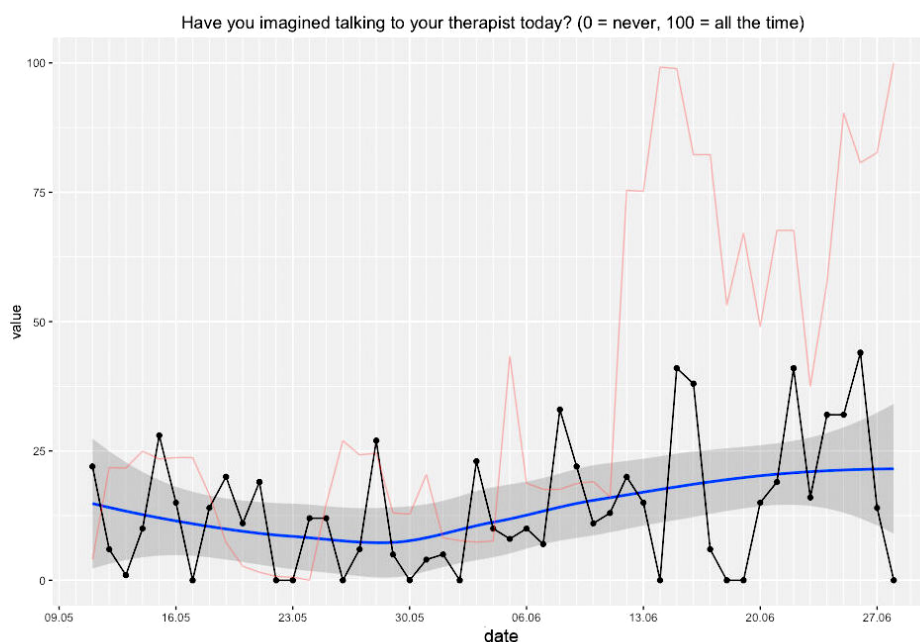
Clinical Example

This example illustrates process monitoring of a 22-year-old female patient suffering from bulimia nervosa, currently undergoing Rogerian person-centered psychotherapy.

The psychotherapy process was measured daily, using a short form of the German version of the IEQ [8]. This questionnaire measures therapy-related cognitions, emotions, and behaviors. The *Reflecting treatment* scale measures thoughts about the patient's behavior toward psychotherapy. The scale named *Relationship fantasies* indicates the frequency of thoughts involving the therapist. The "Problem solving" scale contains items remembering therapy contents and applying them in the patients' daily routine. Also, *Therapy-related emotions* are measured, reflecting positive and negative emotions toward the current therapy.

Weekly therapy outcome was measured on Sundays, using a 27-item short form of the Symptom Checklist-90 (SCL-90) in German [37,38]. Weekly pre-post change scores were calculated.

A depicted in Figure 4, the patient experienced two local peaks of dynamic complexity in her "Problem Solving" scale. The first peak, measured in the first week of treatment, was followed by a symptom reduction of 15%. One week later, another 5% of symptom reduction was observed. The second major peak was measured in the third and fourth week of treatment. Initial weekly treatment outcome was a 2% increase in symptom severity; this was followed by an 11% decrease in the following

week. Time-lagged change in outcome measures is a common pattern observed when associating them with periods of complexity [33]. In the fifth and sixth week of treatment, complexity scores dropped and no peaks occurred. Also, a 34% increase in symptoms was observed, followed by another slight increase.

The observed peaks of complexity can be interpreted as ongoing processes of change, which can be useful in anticipating changes. Practitioners observing these peaks would be well-advised to ensure that the patient experiences stability and a positive therapeutic relationship in her therapy sessions.

However, the increase in symptoms in the fifth week cannot be explained by inspecting only one curve. Additional information can be drawn from other curves. As illustrated in Figure 6, the outcome measure on June 4 was associated with a significant decrease in positive treatment-related emotions (eg, *relief, hope, secure*) and an increase in negative emotions (eg, *anxious, frustration, sadness, hurt*). This information can be used for clarifying the negative outcome reported by the patient.

This relatively simple example shows how therapy process data can inform clinicians about their patients' thoughts and feelings toward their therapy. Information on periods of change the patient is going through as well as negative evaluations of therapy progress would not be accessible to the therapist, possibly leading to missed opportunities to course correction of a therapy that went "off track."

Figure 4. Recurrence plot for the time series illustrated in Figure 2. Darker, red-colored areas of the plot indicate low Euclidean distance between the respective points of the time series whereas brighter, more yellow areas indicate greater Euclidean distance. Greater distance also implies that the time series currently describes a period that has not occurred before. Note that the periods of increased dynamic complexity in the time series depicted in Figure 2 are reflected in the recurrence plot. Both "time" axes refer to points of measurement in the time series.
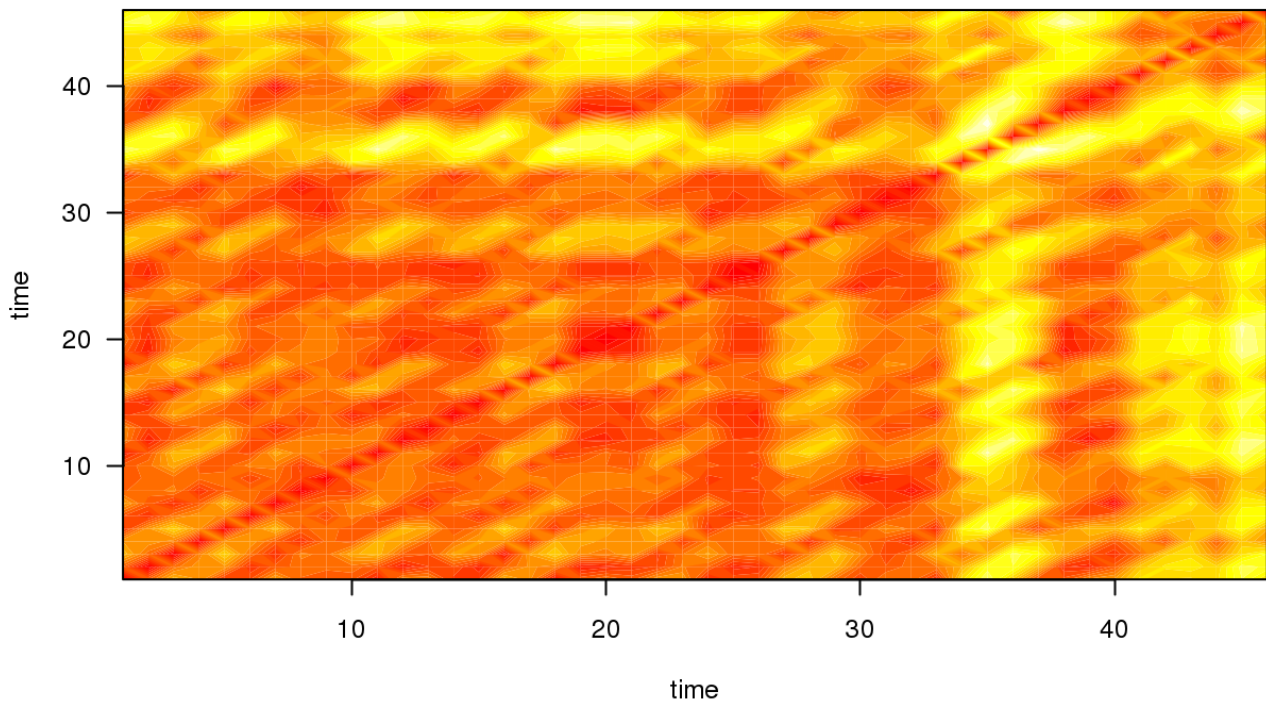
Figure 5. Time series plot showing mean scores of the "Problem Solving" scale of the short intersession questionnaire. A line smoothed by local regression scatterplot smoothing (LOESS), including the 95% CI is added for easy interpretation of long-term change. The red line represents the measure of dynamic complexity that was rescaled from 0 (minimum complexity) to 100 (maximum complexity). Percentage values indicate weekly treatment outcome in percent symptom change.
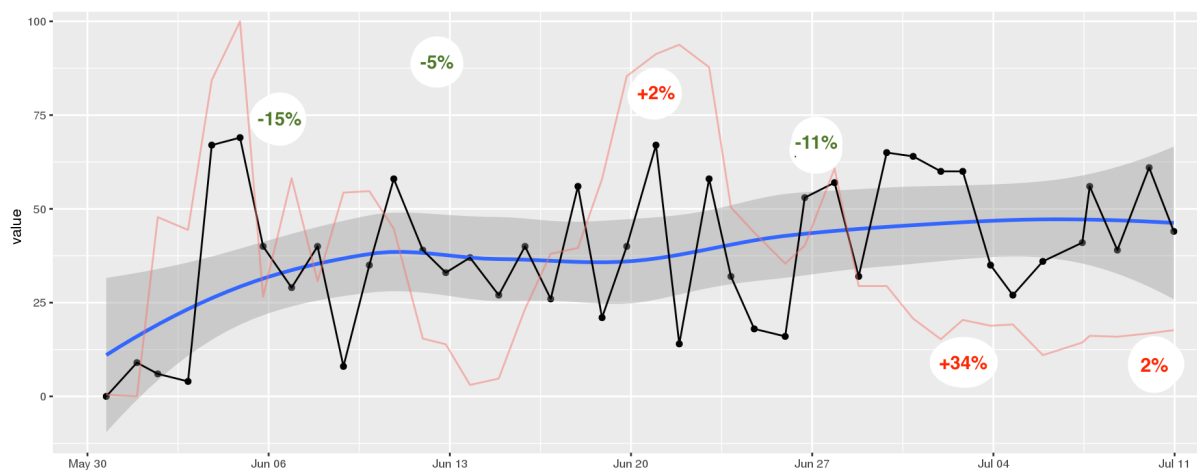
Figure 6. Time series plot showing the (a) the "positive treatment-related emotion" and (b) the "negative treatment-related emotion" scores. The red box marks a significant drop in positive treatment-related emotions, as well as an increase in negative ones.
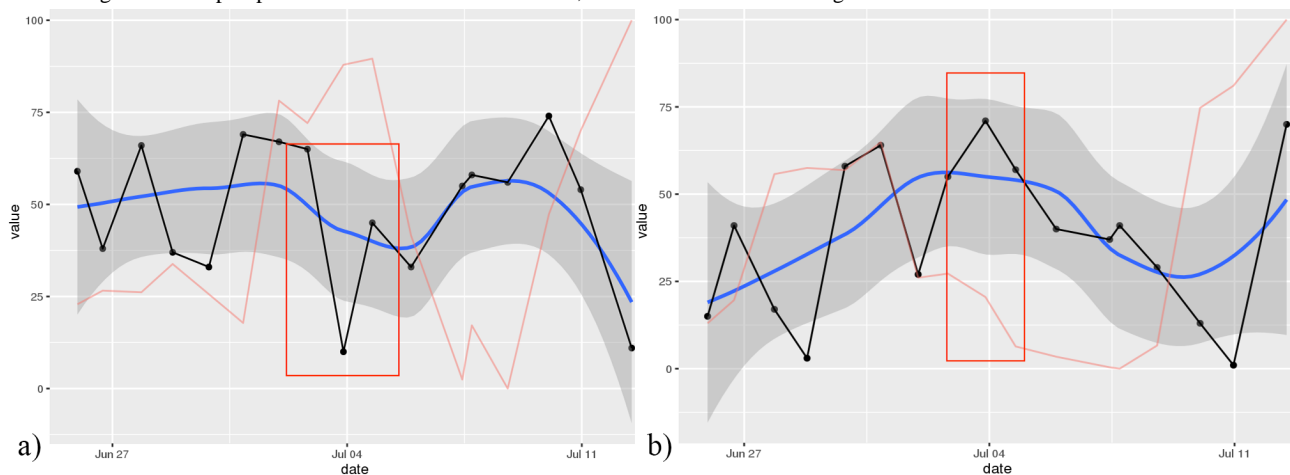
Discussion

Limitations and Challenges

In his recent discussion of challenges in implementing psychotherapy monitoring systems, Boswell [39] identified four core obstacles for mental health providers: financial burden, time burden, different needs of different stakeholders, and turnover. We consider the software presented in this paper an attempt to tackle these obstacles. First of all, because it was developed as an open-source project, the DynAMo software package can be provided without licensing fees.

Practitioners seeking to implement process- and outcome monitoring applications face the challenge of structural changes in their day-to-day routine. The application also tries to keep time and energy expenditure as low as possible for clinicians. We developed the software keeping in mind that it will be used by mental health practitioners, paying great attention to a user-friendly design. A first study of usability of the practitioner and treatment planning interfaces in clinical practice is currently starting. We acknowledge that there is no process monitoring application that is suitable for everyone.

Despite the easy-to-use Web interface, therapists need to be trained in the use of the DynAMo software so that they can extract useful information from patient data. Training should include tutorials and exercises on time series interpretation, including an understanding of the dynamic systems approach to change, which should include interpretation of dynamic complexity and the recognition phase transitions. This way, practitioners can easily grasp information from their patient's trajectories and learn how to recognize periods of change. Training should also include several clinical examples so that practitioners understand how to learn from their patients' data. Another crucial skill is offering feedback to patients so that they learn to view process and outcome monitoring as a part of their treatment. A training program including these different elements is currently developed in our research group.

Another challenge lies in data collection. Psychotherapy patients could find it difficult to answer daily questionnaires, especially if suffering from more severe disorders. Compliance rates for process monitoring are reduced by delayed starting of

questionnaires, early termination, or not filling out a questionnaire at all. A feasibility study by Schiepek [40] showed an average compliance rate of 78.3% and average missing data rates of about 13% when using a 42-item daily questionnaire in an inpatient setting. Similar values were found when assessing adherence to daily mobile phone-based assessment of a short depression scale [41]. Although there is no published data for outpatients yet, modern Internet-enabled devices such as mobile phones, tablet computers, or laptops greatly facilitate data collection. Internal pre-tests of the DynAMo application showed similar compliance rates (80%) and missing data rates (10%) for outpatients. These rates can be considered satisfactory as they do not reduce data quality to a large extent. On a more general level, a study by Torous [42] could show that about 76% (n=100) of persons in the age group of 18 to 60 years are interested in monitoring their mental health with mobile applications.

Therapists' acceptance of technology such as process monitoring applications can depend on several factors. Confirming the issues brought up by Boswell [39], research on technology acceptance [43,44] could show that expected performance of the software is a strong predictor of the intention to use it. Thus, clinical practitioners could feel that the proposed software performs not well enough for routine use. In future usability studies, we will carefully review the data provided by therapists.

Also, the effort to use the software predicts this intention, stressing the importance of usability and training sessions. Practitioners could however see training sessions as a burden, consuming excess time; hence, it will be necessary to design the software so that training sessions can be reduced to a minimum.

Influence by coworkers has been found to increase usage, especially for users with limited experience. It can be concluded that training besides training sessions and regular meetings of users could foster the adoption of this technology. In these meetings, users can also discuss experiences and possible issues using this software.

The intent of this paper was to present a series of tools for psychotherapy process research to the community. Thus, empirical data was only used for illustrating the software's

features, offering only a limited view on the possibilities of this software package. It will be the focus of future empirical studies to obtain larger data sets, including usability data provided by clinical practitioners.

Outlook

It was the goal of this project to provide the psychotherapy research community with a set of tools to study the processes and mechanisms of change in psychotherapy with the high temporal resolution needed to get ecologically valid results and without depending on costly alternatives. We presented a newly developed software for psychotherapy process monitoring and treatment planning in mental health settings. Whereas parts of the software are still under development, the base set of features is complete, and it can now be considered ready for application in empirical research and clinical practice. The DynAMo software should be viewed as an evolving toolset and the full source code is to be released under an open-source software license at a future date via a public project hosting platform such as GitHub, inviting other developers and researchers to participate.

The treatment planning algorithm makes it possible to tailor therapeutic interventions to individual patients, appreciating the great complexity of psychopathology and psychotherapy. With the DynAMo application, the therapy process can be monitored, so that important periods of change are transparent to the therapist. This includes identifying periods of change in ongoing therapies, both from a linear point of view using smoothed graphs and from a dynamic systems point of view, using dynamic complexity plots. The functions of this application cover the period before starting psychotherapy, psychotherapy itself, and they can also be used as a means of sustaining change and preventing relapse by monitoring symptoms after completion for a certain period of time. Thus, the presented application is not only a research tool but also a tool for enhancing psychotherapy with new technologies.

Next steps in the development process of the DynAMo application include an administrator's interface that features easy creation of patient configurations and editing of assessment items using a graphical user interface. Another feature is the possibility of adding short, free-form text items that can be viewed as annotations in time-series graphs. This way, patients can report on meaningful events in a more detailed way. Also, "Traffic-light"-style notifications for certain critical items will be introduced in the future. These notifications can warn therapists of possible treatment drop-out, self-harming behavior, or other critical incidents and also inform them about beneficial developments such as increases in working alliance quality or successes while applying behaviors learned in therapy. Both therapists' and patients' experience will be recorded and examined, so that the application can adapt to their needs.

Conflicts of Interest

None declared.

Multimedia Appendix 1

This example script will demonstrate the treatment planning algorithm in the paper "A Modular Platform for Monitoring Process, Outcome and Algorithm-Based Treatment Planning in Psychotherapy". All steps are illustrated using a simulated data set with items on social anxiety included in the ZIP file. An additional text file with item texts is included as well.

[ZIP File (Zip Archive), 5KB - medinform_v5i3e20_app1.zip]

References

1. NIMH. 2015. Strategic Objective 3: Strive for Prevention and Cures URL: https://www.nimh.nih.gov/about/strategic-planning-reports/strategic-objective-3.shtml [accessed 2016-10-10] [WebCite Cache ID 6l9Zewplv]
2. Drapeau M. 10 tools for progress monitoring in psychotherapy. Integrating Science and Practice 2012 Nov;2(2):1-45 [FREE Full text]
3. Lambert MJ. Progress feedback and the OQ-system: the past and the future. Psychotherapy (Chic) 2015 Dec;52(4):381-390. [doi: 10.1037/pst0000027] [Medline: 26641368]
4. Lambert MJ. Helping clinicians to use and learn from research-based systems: the OQ-analyst. Psychotherapy (Chic) 2012 Jun;49(2):109-114. [doi: 10.1037/a0027110] [Medline: 22642518]
5. Miller SD, Duncan BL, Sorrell R, Brown GS. The partners for change outcome management system. J Clin Psychol 2005 Feb;61(2):199-208. [doi: 10.1002/jclp.20111] [Medline: 15609362]
6. Shimokawa K, Lambert MJ, Smart DW. Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. J Consult Clin Psychol 2010 Jun;78(3):298-311. [doi: 10.1037/a0019247] [Medline: 20515206]

7.   Orlinsky DE, Geller JD, Tarragona M, Farber B. Patients' representations of psychotherapy: a new focus for psychodynamic research. J Consult Clin Psychol 1993 Aug;61(4):596-610. [Medline: 8370855]

8.   Hartmann A, Orlinsky D, Zeeck A. The structure of intersession experience in psychotherapy and its relation to the therapeutic alliance. J Clin Psychol 2011 Oct;67(10):1044-1063. [doi: 10.1002/jclp.20826] [Medline: 21800294]

9.   Stewart S, Schröder T. Emotional homework: a systematic literature review of patients' intersession experiences. J Psychother Integr 2015;25(3):236-252. [doi: 10.1037/a0039639]

10. Hartmann A, Zeeck A, Herzog W, Wild B, de Zwaan M, Herpertz S, et al. The intersession process in psychotherapy for Anorexia Nervosa: characteristics and relation to outcome. J Clin Psychol 2016 Sep;72(9):861-879. [doi: 10.1002/jclp.22293] [Medline: 27199179]

11. Hartmann A, Orlinsky D, Weber S, Sandholz A, Zeeck A. Session and intersession experience related to treatment outcome in bulimia nervosa. Psychotherapy (Chic) 2010 Sep;47(3):355-370. [doi: 10.1037/a0021166] [Medline: 22402092]

12. Zeeck A, Hartmann A. Relating therapeutic process to outcome: are there predictors for the short-term course in anorexic patients? Eur Eat Disord Rev 2005 Jul 11;13(4):245-254. [doi: 10.1002/erv.646]

13. Schiepek G, Eckert H, Aas B, Wallot S, Wallot A. Integrative Psychotherapy: A Feedback-Driven Dynamic Systems Approach. Göttingen: Hogrefe Publishing; 2015.

14. McWilliams N. Psychoanalytic Case Formulation. New York City: Guilford Press; 1999.

15. Schiepek G. Systemische Diagnostik in der Klinischen Psychologie. Weinheim: Psychologie Verlag Union; 1986.

16. Beck AT. Cognitive therapy and the emotional disorders. New York: International Universities Press; 1976.

17. Fisher AJ. Toward a dynamic model of psychological assessment: implications for personalized care. J Consult Clin Psychol 2015 Aug;83(4):825-836. [doi: 10.1037/ccp0000026] [Medline: 26009781]

18. Lee IA, Little TD. P-technique factor analysis. In: Laursen B, Little TD, Card NA, editors. Handbook of developmental research methods. New York: The Guilford Press; Jan 2012:350-363.

19. Fisher AJ, Boswell JF. Enhancing the personalization of psychotherapy with dynamic assessment and modeling. Assessment 2016 Aug;23(4):496-506. [doi: 10.1177/1073191116638735] [Medline: 26975466]

20. Shiny RStudio. A web application framework for R URL: http://shiny.rstudio.com/ [accessed 2016-10-12] [WebCite Cache ID 6lCxkWUzz]

21. Chalmers RP. Generating adaptive and non-adaptive test interfaces for multidimensional item response theory Applications. J Stat Softw 2016;71(5):1-38. [doi: 10.18637/jss.v071.i05]

22. Section III: Emerging measures and models. In: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Arlington: American Psychiatric Association; 2013.

23. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6(1):1-55. [doi: 10.1080/10705519909540118]

24. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]

25. Pfaff B. VAR, SVAR and SVEC models: implementation within R package vars. J Stat Softw 2008;27(4):1-32. [doi: 10.18637/jss.v027.i04]

26. Lütkepohl H. New introduction to multiple time series analysis. Heidelberg: Springer Science & Business Media; 2005.

27. Tschacher W, Zorn P, Thommen M, Müller D, Roder V. Time series analysis of therapy process in groups of clients with personality disorders. Eur Psychiatry 2007 Mar;22(Suppl 1):S60. [doi: 10.1016/j.eurpsy.2007.01.232]

28. Tschacher W, Haemmig R, Jacobshagen N. Time series modeling of heroin and morphine drug action. Psychopharmacology (Berl) 2003 Jan;165(2):188-193. [doi: 10.1007/s00213-002-1271-3] [Medline: 12404073]

29. Tschacher W, Baur N, Grawe K. Temporal interaction of process variables in psychotherapy. Psychother Res 2000;10(3):296-309. [doi: 10.1093/ptr/10.3.296]

30. Ramseyer F, Kupper Z, Caspar F, Znoj H, Tschacher W. Time-series panel analysis (TSPA): multivariate modeling of temporal associations in psychotherapy process. J Consult Clin Psychol 2014 Oct;82(5):828-838. [doi: 10.1037/a0037168] [Medline: 24932566]

31. Beesdo-Baum K, Klotsche J, Knappe S, Craske MG, Lebeau RT, Hoyer J, et al. Psychometric properties of the dimensional anxiety scales for DSM-V in an unselected sample of German treatment seeking patients. Depress Anxiety 2012 Dec;29(12):1014-1024. [doi: 10.1002/da.21994] [Medline: 22933460]

32. Goldfried MR. Systematic desensitization as training in self-control. J Consult Clin Psychol 1971 Oct;37(2):228-234. [doi: 10.1037/h0031974]

33. Schiepek G, Strunk G. The identification of critical fluctuations and phase transitions in short term and coarse-grained time series-a method for the real-time monitoring of human change processes. Biol Cybern 2010 Mar;102(3):197-207. [doi: 10.1007/s00422-009-0362-1] [Medline: 20084517]

34. Haken H. Chance and Necessity. In: Synergetics: Introduction and Advanced Topics. Heidelberg: Springer-Verlag; 2004:147-189.

35. Schiepek GK, Tominschek I, Heinzel S. Self-organization in psychotherapy: testing the synergetic model of change processes. Front Psychol 2014 Oct 02;5:1089 [FREE Full text] [doi: 10.3389/fpsyg.2014.01089] [Medline: 25324801]

36. Marwan N, Carmen Romano M, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems. Phys Rep 2007 Jan;438(5-6):237-329. [doi: 10.1016/j.physrep.2006.11.001]

37. Hardt J, Egle UT, Kappis B, Hessel A, Brähler E. [Symptom Checklist SCL-27]. Psychother Psychosom Med Psychol 2004 May;54(5):214-223. [doi: 10.1055/s-2003-814786] [Medline: 15106055]

38. Derogatis LR. SCL-90-R: Symptom Checklist-90-R: Administration, Scoring, and Procedures Manual. New York: National Computer Systems, Inc; 1994.

39. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. Psychother Res 2015;25(1):6-19. [doi: 10.1080/10503307.2013.817696] [Medline: 23885809]

40. Schiepek G, Aichhorn W, Gruber M, Strunk G, Bachler E, Aas B. Real-time monitoring of psychotherapeutic processes: concept and compliance. Front Psychol 2016 May 03;7:604 [FREE Full text] [doi: 10.3389/fpsyg.2016.00604] [Medline: 27199837]

41. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015 Mar 24;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]

42. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan 21;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]

43. Venkatesh V, Thong J, Xu X. Consumer acceptance and use of information technologyxtending the unified theory of acceptance and use of technology. MIS Quarterly 2012 Mar;36(1):157-178.

44. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003 Sep;27(3):425-478.

Abbreviations

AIC: Akaike information criterion
CFA: confirmatory factor analysis
DynAMo: dynamic assessment and modeling
EFA: exploratory factor analysis
EMA: Ecological Momentary Assessment
IEQ: Intersession Experience Questionnaire
IRT: item response theory
mirtCAT: computerized adaptive testing with multidimensional item response theory
OQ: Outcome Questionnaire
PCOMS: Partners for Change Outcome Management System
RMS: root mean squared residuals
SNS: Synergetic Navigation System
SRMR: standardized root mean squared residual
TLI: Tucker-Lewis-Index
VAR: vector autoregressive

XSL•FO
RenderX

Original Paper

Optimizing the Use of Electronic Health Records to Identify High-Risk Psychosocial Determinants of Health

Nicolas Michel Oreskovic[1,2], MPH, MD; Jennifer Maniates[3], MA; Jeffrey Weilburg[2,4], MD; Garry Choy[2,5], MD

[1]Internal Medicine and Pediatrics, Massachusetts General Hospital, Boston, MA, United States

[2]Harvard Medical School, Boston, MA, United States

[3]Integrated Care Management Program, Massachusetts General Hospital, Boston, MA, United States

[4]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[5]Department of Radiology, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Nicolas Michel Oreskovic, MPH, MD
Internal Medicine and Pediatrics
Massachusetts General Hospital
125 Nashua Street
Boston, MA, 02114
United States
Phone: 1 617 726 0593
Email: noreskovic@mgh.harvard.edu

Abstract

Background: Care coordination programs have traditionally focused on medically complex patients, identifying patients that qualify by analyzing formatted clinical data and claims data. However, not all clinically relevant data reside in claims and formatted data. Recently, there has been increasing interest in including patients with complex psychosocial determinants of health in care coordination programs. Psychosocial risk factors, including social determinants of health, mental health disorders, and substance abuse disorders, are less amenable to rapid and systematic data analyses, as these data are often not collected or stored as formatted data, and due to US Health Insurance Portability and Accountability Act (HIPAA) regulations are often not available as claims data.

Objective: The objective of our study was to develop a systematic approach using word recognition software to identifying psychosocial risk factors within any part of a patient's electronic health record (EHR).

Methods: We used QPID (Queriable Patient Inference Dossier), an ontology-driven word recognition software, to scan adult patients' EHRs to identify terms predicting a high-risk patient suitable to be followed in a care coordination program in Massachusetts, USA. Search terms identified high-risk conditions in patients known to be enrolled in a care coordination program, and were then tested against control patients. We calculated precision, recall, and balanced F-measure for the search terms.

Results: We identified 22 EHR-available search terms to define psychosocial high-risk status; the presence of 9 or more of these terms predicted that a patient would meet inclusion criteria for a care coordination program. Precision was .80, recall .98, and balanced F-measure .88 for the identified terms. For adult patients insured by Medicaid and enrolled in the program, a mean of 14 terms (interquartile range [IQR] 11-18) were present as identified by the search tool, ranging from 2 to 22 terms. For patients enrolled in the program but not insured by Medicaid, a mean of 6 terms (IQR 3-8) were present as identified by the search tool, ranging from 1 to 21.

Conclusions: Selected informatics tools such as word recognition software can be leveraged to improve health care delivery, such as an EHR-based protocol that identifies psychosocially complex patients eligible for enrollment in a care coordination program.

XSL•FO
RenderX

Introduction

An increasing number of states in the United States are transitioning from fee-for-service care to establishing accountable care organizations (ACOs) for patients enrolled in Medicaid, a health care program for people with limited resources, in efforts to improve patient outcomes and control health care costs. Since 2012, 14 states have developed Medicaid ACOs, with Massachusetts launching a pilot version in December 2016 [1]. With the prospect of both Medicaid and Medicare patients enrolled in ACOs in large health care networks across many states, such as Massachusetts, accountability for risk and quality will increasingly be assumed by health care networks and participating providers. For many patients enrolled in Medicaid ACOs, managing risk and improving outcome markers will require understanding factors other than traditional medical complexity [2]. Patients enrolled in Medicaid can often have a variety of upstream social factors that can influence their health, such as housing and employment instability and food insecurity, collectively known as social determinants of health, as well as mental health conditions and substance abuse. These psychosocial factors can shape one's ability to obtain health needs and adhere to health recommendations, and can have a substantial impact on health outcomes [3]. The ability of health care networks participating in Medicaid ACOs to identify those patients with psychosocial drivers with the highest utilization will become increasingly important as networks seek to contain escalating health care costs and appropriately manage pooled risk [4].

While various approaches to identifying medical complexity from an electronic health record (EHR) have been developed and are being employed by health care networks across the United States, there is less certainty about how to identify and grade psychosocial complexity from an EHR [5,6]. As with patients with a high degree of medical complexity, patients with high psychosocial complexity may likewise use and consume substantial health care resources and be challenging to manage clinically [7,8]. Accordingly, there may be value in developing an EHR-based data mining tool for identifying patients with increased psychosocial complexity. Once identified, such patients could be enrolled in a care coordination program that manages complex patients and focuses on decreasing health care utilization and containing health care costs. Care coordination programs have traditionally cared for medically complex patients and have developed various approaches to identifying patients who qualify as high risk [9]. Unlike medical complexity, psychosocial complexity may be more difficult to identify. Medically complex patients are typically identified using International Classification of Diseases codes from claims data or EHR-based algorithms that use structured fields in the medical chart (diagnosis codes, problem lists, medications, or laboratory studies). Privacy laws around mental health and substance abuse, along with the lack of formatted fields for many of the risk factors underlying psychosocial risk, make identifying patients with high psychosocial complexity more challenging. The data necessary to populate the risk categories are often unavailable or suboptimal for population-level screening. Furthermore, when compared with more automated

and search technology-enabled approaches, individual chart review is impractical given its time-consuming and often subjective nature of identifying patients with high complexity.

Given the known limitations and challenges of using available data to identify patients with increased psychosocial risk, we sought to develop an EHR-based tool that could identify patients with increased psychosocial risk. We used the analytics platform QPID (Queriable Patient Inference Dossier; developed at Massachusetts General Hospital and QPID Health Inc, Boston, MA, USA) to search the EHR for key terms predictive of psychosocial risk. QPID is a health intelligence platform incorporating an EHR search engine with a scalable library of US Health Insurance Portability and Accountability Act (HIPAA) -compliant search queries, and a programmable ontology-driven system for application and query development [10]. The engine searches all the data residing within a patient's EHR, including inpatient and outpatient notes, radiology reports, and laboratory data, and can be used to extract detailed information from a single patient's EHR or can be run against an entire patient census.

QPID consumes both structured and unstructured data from the EHR. The unstructured data are in free-text form from the medical record in native format. Both forms of data are extracted, transformed, and loaded into the QPID system, which then performs natural language processing, term indexing, and data aggregation to find and combine medically relevant entities for patients and populations. The natural language processing involves negation detection and date detection, among other techniques. A querying language is overlaid on this processed data to access and visualize data as needed. Medical concepts are clustered through structured ontologies and machine learning techniques, and both open-source ontologies and proprietary clinical knowledge mappings are used. The terms can be mapped to Medical Subject Headings (MeSH); however, custom mapping of medical concepts is often necessary to supplement existing ontologies, especially in the space of psychosocial factors, given that general ontologies often only include biomedically relevant concepts and may lack the nuance to capture social aspects of a patient's well-being.

We hypothesized that, using programmable word recognition software, we could identify patients with high psychosocial complexity at risk for increased health care utilization by using only data available in a patient's EHR.

Methods

Study Population

The study included patients receiving care at Massachusetts General Hospital, a major academic medical center located in Boston, MA, USA. We analyzed EHRs of 132 patients covered by Medicare using QPID to determine the validity of the 22 search terms that we identified. We tested the algorithm on 120 patients enrolled in a care coordination program with documented risk profiles and known psychosocial complexity. Of these 120 index patients, 60 were enrolled in a Medicaid insurance program and 60 were not enrolled in Medicaid. The Medicare patients served as real-world controls against the

Medicaid patients, with known higher rates of psychosocial comorbidity. The Impact Pro score—a medical risk-predictive modeling score based on medical and pharmacy claims data and medical diagnoses information—was available for all patients enrolled in the care coordination program. An additional 12 healthy patients not enrolled in a care coordination program or Medicaid, of whom 6 were adults and 6 were children, served as true-negative controls.

QPID

We used QPID to search patients' EHRs for terms associated with underlying clinical conditions and social risk factors. A list of 54 terms belonging to 4 psychosocial domains (mental health, substance use, social determinants, and legal history) was generated, from which we ultimately identified 22 terms as being sufficiently sensitive and specific to the clinical or social marker being queried (see below). As part of the search term algorithm development process, we removed certain terms that were sensitive but not specific, as summarized below. A blinded manual chart review without knowledge of the search term results was conducted for every study patient by 1 of the study investigators (NO) with expertise in care coordination, with a clinical determination based on clinical judgment for each patient on whether they required care coordination to help manage their psychosocial complexities. The chart review served as the reference standard for assessing psychosocial complexity and ensured that the 22 search terms correctly identified documented psychosocial risk and distinguished psychosocial from medical risk. Using the chart review and QPID result, we created a contingency table and assigned each patient to 1 of 4 categories: true positive, true negative, false positive, or false negative.

Sensitive But Not Specific Search Terms

We designed several search queries (terms) to be sensitive markers (correctly identified patients who were at risk) but nonspecific (also identified patients without risk in whose chart the search term was present but not assigned to the index patient). A relatively more sensitive than specific search term was better suited for screening health records. Several scenarios produced false positives, including lexical variations such as polysemy (a term or abbreviation with multiple meanings), negation, preformatted text, and misallocation. In one example of polysemy, the search term "AA" was a useful and effective marker for identifying alcohol abuse by correctly identifying patients where AA was used as an abbreviation for Alcoholics Anonymous in the EHR, but infrequently also incorrectly identified charts where AA was used as shorthand to signify unrelated categories, including clinical information (amino acid) and demographic information (African American). Negation, a common finding and false-positive source, existed where the search term was listed in a patient note as not being present. Preformatted text was another scenario that produced sensitive

but nonspecific terms, where, for example, the term "depression" incorrectly identified all screening questionnaires and preformatted notes in the EHR that included the word depression, even when a patient reported not being depressed. Misallocation was another scenario resulting in false positives, where data in the EHR describing the reported condition of a friend or relative were incorrectly assigned to the index patient, as with the term "arrested;" an example of this was the mention in the EHR of a patient's son being arrested.

Statistical Approach

We used descriptive statistics to calculate the number of times each term was present within a patient's EHR, and report the mean, interquartile range (IQR), and range for each search term by patient group. We created the list of search terms in the final algorithm by including only terms where the IQRs for index and control patients did not overlap. We compared results between Medicaid-enrolled patients and non-Medicaid-enrolled patients (controls), the latter of which included both non-Medicaid index patients enrolled in the care coordination program and true-negative patients—that is, non-Medicaid patients not enrolled in the care coordination program. Using contingency table results, we calculated the accuracy, precision, recall, and balanced F-measure for the 22-term algorithm's ability to correctly detect and assign psychosocial complexity.

Results

Table 1 describes the study population, providing a summary of demographics and clinical information. Mean Impact Pro scores were not statistically different between Medicaid and non-Medicaid patients enrolled in the care coordination program (3.2 vs 6.2, P=.9).

We identified 22 search terms that correctly predicted increased psychosocial risk with a high degree of specificity: anxiety, depressed, sad, angry, neurovegetative, schizoaffective, substance, abuse, addict, aa, sober, cocaine, heroin, crack, mushrooms, prison, jail, homeless, shelter, stamps, stolen, and tox.

Among the 60 patients enrolled in Medicaid, the mean number of terms per patient was 14.1 (IQR 11-18, range 2-22). Among the 72 control patients not enrolled in Medicaid, the mean number of terms per patient was 6.0 (IQR 3-8, range 1-21). Among the true-negative patients, the mean number of terms per patient among pediatric patients was 2.7 (range 2-3), and among adult patients it was 2.0 (range 1-3). As Figure 1 shows, in blind testing, the 22-search term-based analysis achieved an overall 91% accuracy, 80% precision, 98% recall, and balanced F-measure of 88%.

The 22-search term-based analysis performed well among both Medicare-enrolled and Medicaid-enrolled patients, as well as patients not enrolled in a care coordination program (Figure 2).

Table 1. Study population characteristics.

Characteristics	Index patients (care coordination program enrollees)		True negatives (non-Medicaid, noncare coordination program)	
	Medicaid (n=60)	Non-Medicaid (n=60)	Adults (n=6)	Children (n=6)
Age in years, mean	41	66	64	6
Male, n (%)	32 (53)	27 (45)	3 (50)	4 (67)
Race/ethnicity, n %				
White	54 (90)	50 (83)	6 (100)	6 (100)
Black	2 (3)	1 (2)	0	0
Hispanic/Latino	3 (5)	4 (7)	0	0
Other	1 (2)	5 (8)	0	0
Impact Pro[a] score, median	3.2	6.2	N/A[b]	N/A
QPID[c] terms, mean	14.1	6.0	2.0	2.7

[a]A medical risk-predictive modeling score that uses medical and pharmacy claims data, laboratory results, and medical diagnoses information to predict patients at risk for future severe health problems.

[b]N/A: not available.

[c]QPID: Queriable Patient Inference Dossier.

Figure 1. Performance evaluation using a contingency table of the 22-term QPID (Queriable Patient Inference Dossier) algorithm for identifying psychosocial complexity. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

Figure 2. Performance evaluation using contingency tables of the 22-term QPID (Queriable Patient Inference Dossier) algorithm by insurance payer and care coordination status. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

Patients not enrolled in a care coordination program (true-negative controls), n=12

	Correct		Not correct	
Identified as high risk	TP	0	FP	0
Not identified as high risk	FN	0	TN	10

Medicaid patients enrolled in care coordination program (index patients), n=60

	Correct		Not correct	
Identified as high risk	TP	42	FP	5
Not identified as high risk	FN	1	TN	12

Accuracy=54/60=90%
Precision=42/47=89%
Recall=42/43=98%
Balanced F-measure=93%

Medicare patients enrolled in care coordination program (real-world controls), n=60

	Correct		Not correct	
Identified as high risk	TP	9	FP	8
Not identified as high risk	FN	0	TN	43

Accuracy=52/60=87%
Precision=9/17=53%
Recall=9/9=100%
Balanced F-measure=69%

Discussion

In this paper, we describe the use of a word recognition software program to develop a search term algorithm that accurately identifies Medicaid-enrolled patients with elevated psychosocial risk as distinct from medical risk. While methods exist for assessing and quantifying medical risk using existing medical taxonomy and medical insurance claims data, respectively, psychosocial risk, in contrast, is less well defined in medical claims data and not as robustly classified by medical nomenclature, making it harder to identify using existing datasets. With the expansion of Medicaid ACOs across the United States, and the known prevalence of psychosocial complexity among patients enrolled in Medicaid insurance programs, there will be increased pressure to identify increased psychosocial risk among Medicaid populations for population health management, as well as increasing demand for clinical decision support systems with the capacity to identify patient-attributable psychosocial risk concepts on an individual patient level [11]. Our novel approach offers the ability to use a patient's EHR as a way to identify important psychosocial risk factors potentially driving or contributing to health care utilization and costs, and medical outcomes, among patients enrolled in Medicaid. Moreover, by running our model on patients followed in a care coordination program that manages patients with known medical and psychosocial complexity, we were able to use the algorithm to disentangle medical and psychosocial risk and identify those patients with active psychosocial complexity. In so doing, our findings also underscore the importance of understanding and accounting for psychosocial risk, and provide a mechanism through which providers and health care networks can assess and manage their risk pool by quantifying and triaging psychosocial risk.

Setting the positive criteria as having 9 or more terms present in the EHR as identified by our search tool allowed us to identify patients with a moderate to high burden of active psychosocial complexity, while excluding patients with an existing but low psychosocial complexity or patients with several false-positive markers. Creating an algorithm that assigns the outcome status based on a count of EHR-identified categories rather than on raw term counts avoids creating an algorithm that includes patients who may have a single domain of psychosocial complexity that is frequently documented (eg, a patient whose only health problem is severe anxiety requiring frequent health care visits) or a patient without any psychosocial complexity who has multiple false-positive data returns (eg, an elderly woman who has been administered multiple depression screens over the years; an adult patient with a remote history of child abuse frequently documented in the EHR). Another decision when building the algorithm was to not use date search parameters, given known limitations with how data are entered into and notes are formatted in the EHR (eg, old text sections frequently being carried over into new notes; pretyped templates containing false-positive terms).

Our study has several limitations worth noting. First, our study was a retrospective chart review, and did not prospectively predict outcomes or utilization. Second, we did not compare our findings with utilization data. Large categories of health care utilization data, including mental health data, are not available due to HIPAA requirements, making a valid cost analysis of psychosocial risk difficult to perform. Third, our

reference standard for psychosocial complexity was inclusion in a care coordination program with documented psychosocial complexity requiring social work and mental health services. While possibly subjective and difficult to systematize, the advantage of using patients with known psychosocial complexity who receive services is that this approach uses real-world examples and results in an algorithm that can identify patients who can benefit from such services. Fourth, we used search terms as proxies for identifying clinical concepts, an approach that leverages the power of natural language processing software to search unformatted text for data retrieval; nevertheless, terms and concepts are not necessarily the same, and a clinical concept may be present even when search terms are not. Fifth, for the

methods we describe in this paper to be scalable, the technology will require additional functional enhancements. We ran each patient's data through QPID individually and manually counted the number of identified search items; in order for the approach we describe to be useful for large health care networks, one would need the ability to batch run a list of patients, and the software should automatically return term tallies for each patient.

Despite these limitations, this study provides an important step forward for population health management by outlining a new method for identifying the important role that social determinants and mental health play in health outcomes, and offers a promising new approach to stratifying this risk burden on a population level.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Massachusetts Executive Office of health and Human Services. 1115 waiver proposal information. Boston, MA: The Commonwealth of Massachusetts; 2017. URL: http://www.mass.gov/eohhs/gov/departments/masshealth/1115-waiver-proposal-information.html [accessed 2017-08-08] [WebCite Cache ID 6sZUBZigk]

2. Sheff A, Park ER, Neagle M, Oreskovic NM. The patient perspective: utilizing focus groups to inform care coordination for high-risk medicaid populations. BMC Res Notes 2017 Jul 25;10(1):315 [FREE Full text] [doi: 10.1186/s13104-017-2638-1] [Medline: 28743288]

3. Centers for Disease Control and Prevention. Social determinants of health: know what affects health. Atlanta, GA: CDC; 2017 Jul 28. URL: https://www.cdc.gov/socialdeterminants/ [accessed 2017-08-09] [WebCite Cache ID 6sax5teFk]

4. Hsu J, Price M, Vogeli C, Brand R, Chernew ME, Chaguturu SK, et al. Bending the spending curve by altering care delivery patterns: the role of care management within a pioneer ACO. Health Aff (Millwood) 2017 May 01;36(5):876-884. [doi: 10.1377/hlthaff.2016.0922] [Medline: 28461355]

5. Kotay A, Huang JL, Jordan WB, Korin E. Exploring family and social context through the electronic health record: physicians' experiences. Fam Syst Health 2016 Jun;34(2):92-103. [doi: 10.1037/fsh0000190] [Medline: 27149050]

6. Adler NE, Stead WW. Patients in context--EHR capture of social and behavioral determinants of health. N Engl J Med 2015 Feb 19;372(8):698-701. [doi: 10.1056/NEJMp1413945] [Medline: 25693009]

7. Chang H, Boyd CM, Leff B, Lemke KW, Bodycombe DP, Weiner JP. Identifying consistent high-cost users in a health plan: comparison of alternative prediction models. Med Care 2016 Sep;54(9):852-859. [doi: 10.1097/MLR.0000000000000566] [Medline: 27326548]

8. Coughlin TA, Long SK. Health care spending and service use among high-cost Medicaid beneficiaries, 2002-2004. Inquiry 2009;46(4):405-417. [Medline: 20184167]

9. Hong CS, Siegel AL, Ferris TG. Caring for high-need, high-cost patients: what makes for a successful care management program? Issue Brief (Commonw Fund) 2014 Aug;19:1-19. [Medline: 25115035]

10. QPID. Boston, MA: QPID Health; 2015. URL: http://www.qpidhealth.com/ [accessed 2017-08-08] [WebCite Cache ID 6sZTr4ZwQ]

11. Crawford M, McGinnis T, Auerbach J, Golden K. Population health in Medicaid delivery system reforms. Hamilton, NJ: Center for Health Care Strategies, Inc; 2015 Mar. URL: https://www.chcs.org/media/CHCS_PopulationHealth_IssueBrief.pdf [accessed 2017-08-08] [WebCite Cache ID 6sZUVOWCw]

Abbreviations

ACO: accountable care organization
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
IQR: interquartile range

XSL·FO
RenderX

MeSH: Medical Subject Headings
QPID: Queriable Patient Inference Dossier

XSL•FO
RenderX

Original Paper

System-Wide Inpatient Portal Implementation: Survey of Health Care Team Perceptions

Jennifer L Hefner[1], MPH, PhD; Cynthia J Sieck[1], MPH, PhD; Daniel M Walker[1], MPH, PhD; Timothy R Huerta[1,2], MS, PhD; Ann Scheck McAlearney[1,3], ScD, MS

[1]Department of Family Medicine, The College of Medicine, The Ohio State University, Columbus, OH, United States

[2]Department of Biomedical Informatics, The College of Medicine, The Ohio State University, Columbus, OH, United States

[3]Division of Health Services Management and Policy, The College of Public Health, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Jennifer L Hefner, MPH, PhD
Department of Family Medicine
The College of Medicine
The Ohio State University
2231 N. High St
Columbus, OH, 43201
United States
Phone: 1 6146856322
Fax: 1 6142932715
Email: Jennifer.Hefner@osumc.edu

Abstract

Background: Inpatient portals, a new type of patient portal tailored specifically to the hospital setting, can allow patients to access up-to-date health information and exchange secure communications with their care team. As such, inpatient portals present an opportunity for patients to increase engagement in their care during a time of acute crisis that emphasizes focus on a patient's health. While there is a large body of research on patient portals in the outpatient setting, questions are being raised specifically about inpatient portals, such as how they will be incorporated into the flow of patient care in hectic, stressed, team-based hospital settings.

Objective: Our aim is to improve understanding about hospital care team members' perceptions of the value of an interactive patient portal for admitted patients, as well as to ascertain staff orientation toward this new technology.

Methods: Throughout the course of 2016, an inpatient portal, MyChart Bedside (MCB) was implemented across a five-hospital health system. The portal is a tablet-based app that includes a daily schedule, lab/test results, secure messaging with the care team, a place to take notes, and access to educational materials. Within a month of initial rollout, hospital care team members completed a 5-minute, anonymous online survey to assess attitudes and perceptions about MCB use and staff training for the new technology.

Results: Throughout the health system, 686 staff members completed the survey: 193 physicians (23.6%), 439 nurses (53.7%), and 186 support staff (22.7%). Questions about the importance of MCB, self-efficacy in using MCB with patients, and feelings about sufficient training and resources showed that an average of 40-60% of respondents in each group reported a positive orientation toward the MCB technology and training received. This positive orientation was highest among support staff, lower among nurses, and lowest for physicians (all differences by staff role were statistically significant at $P<.001$). Additionally, 62.0% of respondents reported "not enough" training.

Conclusions: Despite the robust training effort, similar to that used in previous health information technology implementations at this health system, hospital care team members reported only a moderately positive orientation toward MCB and its potential, and the majority wanted more training. We propose that due to the unique elements of the inpatient portal—interactive features used by patients and providers requiring explanation and collaboration—traditional training approaches may be insufficient. Introduction of the inpatient portal as a new collaborative tool may thus require new methods of training to support enhanced engagement between patients and their care team.

XSL•FO
RenderX

Introduction

Patient portals—a class of electronic personal health records (PHRs) tethered to an electronic health record (EHR)—allow patients to view lab and medication information, schedule appointments, and exchange secure messages with providers [1]. Growth in the availability of patient portals has been almost exclusively limited to the ambulatory environment, with studies linking portal use to improved self-management of chronic conditions [2-6] as well as providing evidence of their potential to improve health and lower costs [7-9]. Inpatient portals are emerging as a new type of patient portal tailored specifically to the hospital setting and offer the opportunity for patients to increase engagement with the portal during a time of acute crisis that emphasizes focus on a patient's health.

As portals are interactive, provide up-to-date health information for patients, and enable secure communications with their care team, questions are being raised specifically about inpatient portals such as how they will be incorporated into the flow of patient care in hectic, stressed, team-based hospital settings [9]. Research on inpatient portals, however, is scant, with fewer than 10 studies published that have examined patient use and acceptance in small-scale implementations of inpatient portal technologies [10-13]. Further, while initial studies have reported generally positive findings related to inpatient portal use, the technologies studied have not included interactive elements such as secure messaging with the care team. A recent case study of inpatient portal use at five different academic medical centers, for instance, found variation in the availability of portal features, emphasizing the need to study these novel, interactive elements of inpatient portals [9].

MyChart Bedside (MCB), an inpatient portal, is a tablet-based app patients can use to access their data while admitted at an Epic-equipped hospital and includes interactive functionalities. MCB was developed by Epic—a proprietary software company whose EHR has been adopted by hospitals serving more than 50% of US patients—to provide patients and their families and caregivers access to information customized to the inpatient setting. It includes a daily schedule, lab/test results, secure messaging with the care team, a place to take notes, and access to educational materials. Recent implementation of MCB across a large Midwestern multihospital health system provided the opportunity to survey staff during the initial implementation phase to explore the perceived value of a patient portal for admitted patients from the clinician perspective, as well as to ascertain staff attitudes to deployment of this new technology. This study adds to what we expect will be a growing literature that identifies the unique dynamics associated with inpatient use of patient portals.

Methods

Study Setting

Throughout 2016, MCB was implemented across all units of a five-hospital tertiary care academic medical center in a large metropolitan city, with nearly 1400 inpatient beds and over 5000 providers. The MCB implementation was accompanied by a training and engagement plan that included identifying and training unit "champions" who received dedicated time to devote to this role, delivering information sessions on each unit to orient staff to the technology and tablet provisioning plan, having information technology staff available on the units during the initial "go-live", and providing access to online documentation detailing tablet provisioning procedures and e-learning modules focused on MCB.

Survey Process

We surveyed hospital staff across the health system to assess attitudes and perceptions about MCB use and their training to use the new technology. Specifically, within a month of initial MCB implementation, hospital care team members received a recruitment email with a link to the survey. Over the implementation timeframe, this email was sent to all 5000 providers through unit-specific listservs. This protocol was approved by the study site's institutional review board.

Survey Instrument

The anonymous, online survey instrument took about 5 minutes to complete. Questions included the respondent's role within the academic medical center (physician, nurse, unit clerical associate [UCA], patient care assistant [PCA]), and a series of questions about the respondent's orientation toward and training with the technology, such as the importance of MCB, self-efficacy for using MCB, and feelings about sufficient training and resources (5-point response categories from "Strongly Disagree" to "Strongly Agree"). In addition, the seven features of MCB were listed—Dining on Demand, Education, Secure Messaging, Medication List, Problem List, Schedule, and Description of Care Team—with respondents asked to rate "the features of MCB according to how much you expect that patients will use them" (5-point response categories: "Not at all" to "A lot") and "the features of MCB according to how much you expect that patients will benefit from them" (5-point response categories: "Not at all" to "Extremely").

Survey Analysis

Descriptive statistics were calculated for each survey question and when appropriate, cross tabulations and chi square tests of a statistically significant difference between categories were conducted to compare response choices between survey groups. For these analyses, 5-point response categories listed in the above section were also collapsed into a binary variable equal to 1 for the top two response categories (eg, "Agree/Strongly agree"). Additionally, the UCA and PCA respondent groups were collapsed into one job category due to small numbers and the fact that they played similar roles in the inpatient medical

care hierarchy. This job category is referred to as "clinical support staff" below.

Results

Across the health system, 686 staff members completed the survey: 193 physicians (23.6%), 439 nurses (53.7%), and 186 clinical support staff (22.7%). Table 1 presents responses to questions about respondents' orientation toward MCB. We found that the questions about the importance of MCB, self-efficacy in using MCB, and feelings about sufficient training and resources showed an average of 40-60% of respondents in each group reporting a positive orientation toward the MCB technology and training received. This positive orientation was highest among support staff, lower among nurses, and lowest for physicians (all differences by staff role were statistically significant at $P<.001$). On average, 62.0% (425/686) of respondents reported "not enough" training. Among physicians, 79.9% (154/193) responded they had lacked sufficient training

compared with 61.7% (271/439) of nurses and 46.2% (86/186) of support staff.

When asked about the MCB features patients would be likely to use most often and how much patients would benefit, respondents reportedly valued the features differently (Table 2). Dining on Demand was the feature respondents reported patients would most likely use and benefit from, with more than two thirds of respondents reporting patients would use electronic meal ordering "A lot/Often" (473/686, 68.9%) and would benefit "Very much/Extremely" (449/686, 65.5%). Next, almost half of respondents reported the Medication List and the Schedule as features patients were likely to both use and benefit from. Secure Messaging was less frequently endorsed, with low rates of likely use and benefit: 16.5% (113/686) and 24.9% (171/), respectively. Notably, there was a large discrepancy between perceptions of use and benefit in the Education feature, with 22.4% (154/686) reporting likely patient use and 50.9% (349/686) reporting potential patient benefit.

Table 1. Hospital staff perspectives on MCB technology and training[a].

	Strongly agree/Agree, %			
	All (n=686)	Physicians (n=193)	Nurses (n=439)	Clinical support staff (n=186)
I am aware of the reasons this health system is implementing MCB.	75.2	57.8	78.5	85.2
I feel the health system is promoting use of MCB.	70.6	56.9	73.2	76.5
It is important to provide access to MCB to patients in this hospital.	63.2	48.3	63.4	76.4
I believe that patients will benefit from MCB.	57.0	41.5	55.9	73.9
I understand responsibilities within the care team on my unit for responding to MCB questions.	56.6	33.3	57.6	75.0
I can play an important role in helping patients manage their health through MCB.	47.3	36.1	46.8	58.6
I am interested in helping patients manage their health through MCB.	48.6	38.4	45.4	65.6
There are sufficient resources on my unit to effectively incorporate MCB.	42.4	19.7	43.7	59.5
I have the tools I need to help my patients use MCB.	40.2	22.5	39.3	58.6

[a]For all statements, differences between groups were statistically significant at $P<.001$.

Table 2. Hospital staff ratings of MCB features patients are likely to use most often and how much patients will benefit from them.

	Patients will use "A lot/Often", %	Patient will benefit "Very/Extremely", %
Dining on Demand	68.9	65.5
Education	22.5	50.9
Secure Messaging	16.6	25.0
Medication List	41.4	43.9
Problem List	21.9	29.7
Schedule	37.7	43.0
Description of Care Team	27.3	37.7

Discussion

Principal Considerations

Results from this early implementation survey revealed staff had a moderately positive orientation toward the MCB tool and

its potential, and this varied by job role. Clinical support staff (PCAs and UCAs) was most positively oriented toward the technology, while nurses and physicians were less convinced that MCB was an important tool. Further, physicians were less confident than the other groups about both their role with the technology and whether they had sufficient training to feel

XSL•FO

RenderX

comfortable incorporating MCB in their workflow. As physicians at this institution were less involved in the training and implementation than other groups, this may account for their less positive attitude toward the technology. The primary physician use of MCB is secure message communication, a new feature in the hospital setting, thus their lack of comfort with MCB suggests the need for physician training focused on this feature.

Our findings about less positive nurse feelings are less clear. Although MCB provisioning procedures differ by unit across the health system, the nurses and support staff are all involved in the distribution, use, and collection of MCB. Increasing engagement of nurses involved in direct patient care has been highlighted as an important element of portal use in the inpatient setting [9]. Given that nurses can be expected to interact with patients frequently using the portal, whether by responding to questions when they are in the patient's room or via secure messages, our findings suggest that additional focus on improving nurse perceptions of this tool may be important.

Limitations

This study has several notable limitations. First, given the survey was anonymous, we do not have any information about nonrespondents. There could have been nonresponse bias related to satisfaction with this new technology. For example, those with more negative attitudes may have been more likely to not respond. If this is the case, then the true level of negative feelings is even lower. It is also possible that demographic factors such as age and gender, tenure at the organization, or experience in the field may play roles influencing attitudes toward inpatient portals. This short paper is the first reporting results from a program of research for this study team on the implementation and use of an inpatient portal across a large medical center. Interviews with staff and providers are ongoing and will provide crucial information about the facilitators and barriers to improving providers' attitudes toward, and increasing their confidence using, this new technology.

Conclusions

For this implementation of MCB, the medical center engaged in a robust staff education effort similar to that used in previous health information technology (HIT) implementations. This general approach, however, may not account for unique features of an inpatient portal compared to other hospital-focused HIT tools. First, the inpatient portal includes features utilized by patients, not just the care team, and these features may require additional explanation to support their appropriate use. Second, the inpatient portal introduces the ability to communicate via secure messaging and represents a new avenue for collaboration between the patient and the care team not previously available in the inpatient setting. Research in the outpatient context suggests that this type of collaboration is particularly challenging for both patients and providers because it requires new rules by which each party engages [14]. Introduction of the inpatient portal as a new collaborative tool may thus require new methods of training to support enhanced engagement between patients and their care team.

Conflicts of Interest

None declared.

References

1. Wald JS, Middleton B, Bloom A, Walmsley D, Gleason M, Nelson E, et al. A patient-controlled journal for an electronic medical record: issues and challenges. Stud Health Technol Inform 2004;107(Pt 2):1166-1170. [Medline: 15360996]
2. Hess R, Bryce C, Paone S. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. Telemed J E Health 2007;13(5):509-517. [doi: 10.1089/tmj.2006.0089published]
3. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
4. Halamka J, Mandl K, Tang P. Early experiences with personal health records. J Am Med Inform Assoc 2008;15(1):1-7. [doi: 10.1197/jamia.M2562]
5. Kaelber D, Pan EC. The value of personal health record (PHR) systems. In: AMIA Annual Symposium Proceedings. 2008 Jan 01 Presented at: AMIA Annual Symposium; Nov. 8-12, 2008; Washington, DC p. 343-347.
6. Emont S. Measuring the Impact of Patient Portals: What the Literature Tells Us. Oakland, CA: California HealthCare Foundation; 2011. URL: http://www.chcf.org/publications/2011/05/measuring-impact-patient-portals [accessed 2017-09-05] [WebCite Cache ID 6tG7xSVlk]
7. Dentzer S. Rx for the 'blockbuster drug' of patient engagement. Health Aff (Millwood) 2013 Feb;32(2):202 [FREE Full text] [doi: 10.1377/hlthaff.2013.0037] [Medline: 23381509]
8. Vawdrey DK, Wilcox LG, Collins SA, Bakken S, Feiner S, Boyer A, et al. AMIA Annual Symposium Proceedings. 2011 Oct 22 Presented at: AMIA Annual Symposium; Oct. 22-26, 2011; Washington, DC.

XSL·FO

RenderX

9.   Collins S, Rozenblum R, Leung W, Morrison CR, Stade DL, McNally K, et al. Acute care patient portals: a qualitative study of stakeholder perspectives on current practices. J Am Med Inform Assoc 2017 Apr 01;24(e1):e9-e17. [doi: 10.1093/jamia/ocw081] [Medline: 27357830]

10.   Prey J, Woollen J, Wilcox L, Sackeim AD, Hripcsak G, Bakken S, et al. Patient engagement in the inpatient setting: a systematic review. J Am Med Inform Assoc 2014;21(4):742-750 [FREE Full text] [doi: 10.1136/amiajnl-2013-002141] [Medline: 24272163]

11.   Dalal A, Dykes P, Collins S, Lehmann LS, Ohashi K, Rozenblum R, et al. A web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: a preliminary evaluation. J Am Med Inform Assoc 2016 Jan;23(1):80-87. [doi: 10.1093/jamia/ocv093] [Medline: 26239859]

12.   Kelly M, Hoonakker P, Dean S. Using an inpatient portal to engage families in pediatric hospital care. J Am Med Inform Assoc 2017 Jan;24(1):153-161. [doi: 10.1093/jamia/ocw070] [Medline: 27301746]

13.   O'Leary K, Lohman M, Culver E, Killarney A, Randy SGJ, Liebovitz D. The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. J Am Med Inform Assoc 2016 Jan;23(1):159-165. [doi: 10.1093/jamia/ocv058] [Medline: 26078412]

14.   Sieck C, Hefner J, Schnierle J, Florian H, Agarwal A, Rundell K, et al. The rules of engagement: Perspectives on secure messaging from experienced ambulatory patient portal users. JMIR Med Inform 2017 Jul 04;5(3):e13 [FREE Full text] [doi: 10.2196/medinform.7516] [Medline: 28676467]

Abbreviations

EHR: electronic health record
HIT: health information technology
MCB: MyChart Bedside
PCA: patient care associate
PHR: personal health record
UCA: unit clerical associate

XSL•FO
RenderX

Original Paper

Triaging Patient Complaints: Monte Carlo Cross-Validation of Six Machine Learning Classifiers

Adel Elmessiry[1], PhD; William O Cooper[2], MD, MPH; Thomas F Catron[2], PhD; Jan Karrass[2], PhD; Zhe Zhang[3], PhD; Munindar P Singh[1], PhD

[1]North Carolina State University, Department of Computer Science, Raleigh, NC, United States
[2]Vanderbilt University Medical Center, Nashville, TN, United States
[3]IBM, Research Triangle Park, NC, United States

Corresponding Author:
Adel Elmessiry, PhD
North Carolina State University
Department of Computer Science
890 Oval Drive 3320 Engineering Building II
Raleigh, NC, 27606
United States
Phone: 1 919 515 2858
Fax: 1 919 515 7896
Email: adel_elmessiry@ncsu.edu

Abstract

Background: Unsolicited patient complaints can be a useful service recovery tool for health care organizations. Some patient complaints contain information that may necessitate further action on the part of the health care organization and/or the health care professional. Current approaches depend on the manual processing of patient complaints, which can be costly, slow, and challenging in terms of scalability.

Objective: The aim of this study was to evaluate automatic patient triage, which can potentially improve response time and provide much-needed scale, thereby enhancing opportunities to encourage physicians to self-regulate.

Methods: We implemented a comparison of several well-known machine learning classifiers to detect whether a complaint was associated with a physician or his/her medical practice. We compared these classifiers using a real-life dataset containing 14,335 patient complaints associated with 768 physicians that was extracted from patient complaints collected by the Patient Advocacy Reporting System developed at Vanderbilt University and associated institutions. We conducted a 10-splits Monte Carlo cross-validation to validate our results.

Results: We achieved an accuracy of 82% and F-score of 81% in correctly classifying patient complaints with sensitivity and specificity of 0.76 and 0.87, respectively.

Conclusions: We demonstrate that natural language processing methods based on modeling patient complaint text can be effective in identifying those patient complaints requiring physician action.

Introduction

Patient complaints are an important source of information for health care organizations for improving the patient experience. Patients are uniquely positioned to make observations about the care they receive, particularly when they complain when health care professionals or organizations fail to meet their expectations. When patients and family members share their observations, organizations can engage in service recovery, the process of "making right" what went wrong for patients and families [1]. Most patient complaints can be resolved at the point of service and require no additional action. However, when a patient expresses dissatisfaction about some aspect of a physician's practice, it is important to share that information with the physician so that he or she can reflect on the situation and potentially develop strategies for preventing the recurrence

of the events that engendered the initial dissatisfaction [2]. When patterns develop, reviewing patient complaints offers insight into sources of potential continuing patient dissatisfaction that can be addressed by the medical professional and/or the organization. Some patient and family complaints require immediate response by the organization and/or the health care professional because review and response is required by law, regulation, or policy [3] (eg, sexual boundary violation, drug or alcohol impairment in the workplace).

Many health care organizations receive thousands of unsolicited patient complaints a year [4-6]. Manual review of these complaints by trained coders has been shown to be reliable and valid [7], but it is time consuming and may occur some weeks or months after the complaint is received. In addition, scalability of human coding presents logistical and time challenges. Thus, there is a need to triage patient complaints to identify complaints that should be shared with the involved physician(s). We describe a study in which we implemented several well-known machine learning classifiers to optimally detect patient complaints about physicians' practices using data from the Patient Advocacy Reporting System (PARS), a national program that draws data from multiple hospitals' patient complaint reporting systems to identify professionalism concerns and malpractice risk among health care professionals [7].

Problem, Challenges, and Approach in Brief

We posit that complaint text can be used to discern the relevance of a complaint to a physician and thus correctly and efficiently identify which complaints should be shared. The goal of this study is to determine whether a given patient complaint can be shared with the physician with the same level of accuracy that is achieved with existing manual approaches.

Our problem is challenging due to the following factors:

1. Physician practice-related complaints are not always easy to characterize. The vocabulary used to describe physician-related complaints overlaps with that used for other types of complaints (eg, billing), partly because of the common effect of the medical setting.
2. Achieving optimal accuracy is a delicate balance. Failing to detect a physician-related complaint results in continued patient dissatisfaction that could have been successfully addressed. On the other hand, false positive instances, where unrelated complaints are shared with physicians, would result in wasted time and effort.
3. Text for a single complaint may gather multiple perspectives, including the patient and the patient's family, friends, and care providers. These parties have different and possibly conflicting objectives. In most cases, patient advocates record patient complaints in the system using patient words without rewording or paraphrasing. The advocate may insert their impression, such as "the patient was angry" or "the patient was shouting." Patient advocates may add subsequent actions and responses to the patient complaint. In other cases, the complaint process begins when a patient writes a letter to the medical center, in which case the advocates would take snippets from the actual letter. There can be paraphrasing depending on the organization and the individual advocate.

Our approach involves extracting common features from physician-related patient complaints that have already been correctly classified as such by a team of human coders. Those common features are then applied to a second group of patient complaints in order to classify them as either physician-related or non-physician-related complaints.

For our comparisons, we (1) implement a framework that employs six well-known classifiers and (2) experiment with two methods of feature extraction from complaint text.

Related Work

The bulk of the textual artifacts in health care can be found in two main sources: clinical and nonclinical. Clinical textual artifacts are largely entries in the medical chart, comments on the case, or physician notes. Medical chart notes tend to be consciously made and well structured, whereas case comments and physician notes focus on treatment (including diagnoses) of the patient. Nonclinical textual artifacts include unsolicited patient feedback and often revolve around complaints. The text is variable, may contain abbreviations, and may extend beyond the actual treatment or diagnosis.

Previous research has focused on clinical textual artifacts [8]. Recent research demonstrates the possibility to apply natural language processing (NLP) on electronic medical records to identify postoperative complications [9]. Bejan and Denny [10] showed how to identify treatment relationships in clinical text using a supervised learning system that is able to predict whether or not a whether or not a treatment relation exists between any two medical concepts mentioned in the clinical notes exists between any two medical concepts mentioned in the clinical notes.

Cui et al [11] explored a large number of consumer health questions. For each question, they selected a smaller set of the most relevant concepts adopting the idea of the term frequency-inverse document frequency (TF-IDF) metric. Instead of computing the TF-IDF based on the terms, they used concept unique identifiers. Their results indicate that we can infer more information from patient comments than commonly thought. However, questions are short and limited, whereas patient complaints are rich and elaborate.

Sakai et al [12] concluded that how risk assessment and classification is configured is often a decisive intervention in the reorganization of the work process in emergency services. They demonstrated the textual analysis of feedback provided by nurses can expose the sentiment and feelings of the emergency workers and help improve the outcomes.

Temporal information in discharge summaries has been successfully used [13] to classify encounters, enabling the placement of data within the structure to provide a foundational representation on which further reasoning, including the addition of domain knowledge, can be accomplished.

Additional research [14] extended the clinical Text Analysis and Knowledge Extraction System (cTAKES) with a simplified feature extraction, and the development of both rule and machine learning-based document classifiers. The resulting system, the Yale cTAKES Extensions (YTEX), can help classify radiology

reports containing findings suggestive of hepatic decompensation. A recent systematic literature review of 85 articles focusing on the secondary use of structured patient records showed that electronic health record data structuring methods are often described ambiguously and may lack clear definition as such [15].

Complaints

For the objective of this research, we group complaints into two main categories, as described subsequently.

Complaints Involving a Physician

These are complaints can be inferred to be, and are, associated with a physician's practice:

> Dr XXX seemed more concerned with getting to her next patient than to listening to what I had to say.

> After the procedure she asked Dr XXX if he would be speaking with her dad. He said no, he tells the family and they can tell the pt [patient]. The daughter does not feel it was her place to discuss with her dad that he has terminal cancer..

> The patient asked the doctor to give her an x-ray, but he refused. Two days later, the patient went to the emergency room and an x-ray showed that her arm was broken.

> Obviously, Dr XXX did not review his medical chart

> Dr XXX rushed through the appointment.

> I arrived early for my appointment but had to wait almost 2 hours to be seen. This happens every time I see Dr XXX.

Complaints Not Involving a Physician or His/Her Practice

These are complaints that concern billing or requesting information (or are not a complaint at all). They normally do not require medical escalation and can be typically handled by the staff:

> Patient has contacted our office multiple times to get assistance with getting her CPAP machine repaired. She stated that we had not given her home health company the needed information.

> The ER triage RN "treated her husband like garbage." [The inpatient] RN "the attitude queen would not call the doctor for a sleeping medication" and that the service coordinator was "rude and stated the manager of the unit refused to speak to her."

> Mrs X was scheduled for an appointment in the North office on October 20. She was told that her appointment would be in the East location. Mrs X's son traveled a couple of hours to bring his mother to her appointment. When they arrived for her appointment, there was no one in the East office so they left and went home.

> She sat in the ER last night from 7:45 pm to 8:20 pm without being triaged. Patient states she has asthma and she was having a severe allergic reaction. Patient states a young male RN told her she would be seen next but the other triage RN called seven people before her.

Human Coders

Each unsolicited patient complaint report in our dataset had previously been reviewed by a trained research assistant and identified as either containing a complaint about a physician or not. These 15 research assistants received extensive training on the classification protocol and met internally developed reliability standards [7]. The standard of reliability was an alpha of 0.80 or higher [16,17]. The interrater agreement reliability between pairs of research assistants ranged from 0.70 to 0.95, with a median alpha of 0.86. The intercoder agreement was high due to the extensive training the coders underwent on the PARS classification.

Methods

No single term or attribute signifies whether or not a patient complaint involves a physician and/or his or her medical practice. Therefore, we approached the problem by clustering text into one of two clusters. Documents are commonly represented as a sparse vector over the entire feature set consisting of all distinct terms over all documents. Two major drawbacks are (1) high dimensionality (ie, a large number of features) and (2) feature sparsity (ie, features appearing in only a few documents) [18].

Accordingly, we implemented a framework that consisted of the following steps: (1) preprocessed the documents to remove common stop words and numbers and to perform stemming (eg, the stem "argu" would replace "argue," "argued," "argues," "arguing," and "argus"); (2) ran Monte Carlo cross-validation [19] using 10 splits and for each we randomly sampled an 80% training and 20% testing dataset from our corpus (approximately 11,468 training complaints and 2867 testing complaints), extracted features through generating sparse representation of the documents based on TF or TF-IDF, reduced features by removing sparse terms, and trained a model to predict the labels; (3) computed the mean accuracy, sensitivity, and specificity for each classifier; and (4) selected the best-performing classifier.

Feature Extraction

The first step was to map patient complaints to a set of representative features. Wilcox and Hripcsak [20] showed that domain knowledge representation can vary between task-specific and representation-specific knowledge. Medical knowledge is specific to the conditions being identified and essential for clinical report classification. As in our case, Wilcox and Hripcsak emphasized attribute or feature extraction. Generating medically relevant features requires an understanding of the medical report or the underlying meaning of the text. Our approach associates medical relevance with feature relevance to the document.

We compared two methods for feature extraction, namely TF and TF-IDF. TF-IDF seeks to emphasize the importance of a word to a document in a collection or corpus [21]. In information classification and retrieval, TF-IDF is widely used [22]. The

idea is simply to multiply the TF with IDF computed with respect to the entire corpus as shown in Equation 1:

(1) TF-IDF(t)=tf(t,d)×log(N/n$_t$)

where tf (t,d) counts the frequency by which term t appears in document d, N is the total number of documents in the corpus, and n_t is the number of documents in which the term t appears.

The idea of incorporating IDF is to reduce the weight on words that occur frequently in each document, but are not sufficiently selective. For example, the words "her" and "nurse" would occur too commonly in patient complaints to be useful for retrieval or selection.

We adopted TF-IDF for feature extraction as follows: (1) generated a vocabulary of unique terms, (2) generated term frequency per document, (3) generated inverse document weight per term, and (4) replaced the frequency with the TF-IDF weights using Equation 1. The result was a sparse vector representation of the document.

Feature Reduction

Feature reduction aims at reducing the number of features while maintaining the underlying meaning of the document. A smaller number of representative features can maintain a comparable level of prediction performance while reducing noise and unnecessary processing. Both TF and TF-IDF generated a large number of features, the majority of which are not relevant in predicting whether a complaint involves a physician. To reduce the number of features, we removed sparse features. We applied a similar definition of the term sparsity described in Saif et al [23], which can be defined as the ratio of the number of documents in which this term appears and the total number of documents in the corpus, as shown in Equation 2:

(2) Sparsity=n$_t$/N

where n_t is the number of documents in which the term t appears and N is the total number of documents in the corpus. A term with 0.90 sparsity means the term appears in at least 90% of the documents, whereas a term with 0.99 sparsity appears in at least 99% of the documents.

We repeated the Monte Carlo cross-validation training and prediction while varying the sparsity from 0.90 to 0.99 to assess the minimum number of features to select and still maintain the desired prediction performance. A reduced number of representative features is desirable because it reduces the size of the model while maintaining the accuracy. The following example shows some selected word stem features organized into four groups for illustration purposes: (1) financial account, charge, close, bill, and call; (2) medical cardiac, cardiology, complications, injury, and coronary; (3) facility center, clinic, access, action, and assist; and (4) care complaint, concern, attach, and care.

Classifier Selection

The final step was to assess the best classifier to employ for our problem. Due to the special nature of the problem, selecting a classifier prospectively was difficult. We implemented a supervised learning framework to capture the relation between patient text and the resultant physician action. The models then could detect whether the complaint was related to a physician's practice. Our framework supported six well-known classifiers. We used RTextTools [24] as the library to implement the classifiers shown in Table 1. After experimenting with these classifiers on the same dataset, we selected the best overall performing classifier.

Evaluation

We divided the dataset into a training and a testing dataset. We used one of the six classifiers to train a model over the mapped dataset. We then used the testing dataset to validate the accuracy of our classifiers. Accuracy is defined by Equation 3:

(3) Accuracy=(true positives+true negatives)/(true positives+false positives+true negatives+false negatives)

Sensitivity captures how many patients with a condition are detected (ie, the avoidance of false negatives) as in Equation 4:

(4) Sensitivity=true positives/(true positives+false negatives)

Specificity captures how many patients without a condition are not detected (ie, the avoidance of false positives) as shown in Equation 5:

(5) Specificity=true negatives/(true negatives+false positives)

The F-score captures how accurate the test was. It is computed using both the precision and the recall as shown in Equation 6:

(6) F-score=2×(precision×recall) /(precision+recall)

Institutional Review Board Approval

This research was reviewed and approved by the Vanderbilt Medical Center Institutional Review Board and the North Carolina State University Institutional Review Board.

Results

We first report our full 10-splits results for each classifier's predictions. Figure 1 shows the results obtained using TF-extracted features. We experimented with changing the sparsity from 0.90 to 0.99 to reduce the number of selected features. The prediction accuracy either slightly improved or remained steady with the reduced number of features except in the random forests case, where the accuracy peaked and dropped slightly at the end of the range.

The case is a bit different with results obtained using TF-IDF-extracted features, as shown in Figure 2. The prediction of all classifiers improved notably (from 2.5% in the case of random forests to 12.1% in the case of SLDA) because we reduced the number of selected features. The gap between the best-performing classifier using TF-IDF and the rest of the classifiers was more pronounced as well. Because results were generally better at higher sparsity, we reported the detailed results at sparsity of 0.99 with accuracy, sensitivity, and specificity for both TF and TF-IDF in Table 2 as well as the harmonic mean (F-score) over each of the six classifiers we implemented.

Table 1. Implemented classifiers.

Classifier	Description
Scaled linear discriminant analysis (SLDA)	Expresses one dependent variable as a linear combination of other variables. SLDA is similar to ANOVA, but with the difference that SLDA assumes continuous independent variables and categorical dependent labels. SLDA is widely used in image and pattern recognition [25].
Support vector machines (SVM)	Divides the dataset via a set of hyperplanes during the learning phase and maps new data to fall into one of the hyperplanes. SVM has been used for text classification [26].
Glmnet	An implementation of the Lasso and elastic-net regularized generalized linear models, Glmnet is popular for domains with large databases [27].
Max entropy	A probabilistic classifier that selects the model with maximum entropy from among a set of models and uses it to classify data [28].
Boosting	Aggregates a set of weak learners (classifiers that perform slightly better than random) to create a strong learner by weighting them appropriately [29].
Random forests	An ensemble learning method, similar to boosting, that learns and combines many decision trees and subsequently selects the best performing from among multiple learning algorithms to improve predictions.

Table 2. Classifiers term frequency (TF) versus term frequency-inverse document frequency (TF-IDF) accuracy, sensitivity, specificity, and F-score using 10-splits Monte Carlo cross-validation at 0.99 sparsity.

Classifier	TF				TF-IDF			
	Accuracy	Sensitivity	Specificity	F-score	Accuracy	Sensitivity	Specificity	F-score
SLDA	0.76	0.72	0.80	0.76	0.74	0.66	0.83	0.74
SVM	0.79	0.71	0.86	0.78	0.75	0.67	0.82	0.74
Glmnet	0.76	0.71	0.81	0.75	0.76	0.64	0.86	0.73
Max entropy	0.77	0.71	0.83	0.76	0.77	0.69	0.84	0.76
Boosting	0.70	0.85	0.55	0.67	0.73	0.82	0.64	0.72
Random forests	0.80	0.74	0.87	0.80	0.82	0.76	0.87	0.81

Figure 1. Term frequency-generated features using 10-splits Monte Carlo cross-validation accuracy.
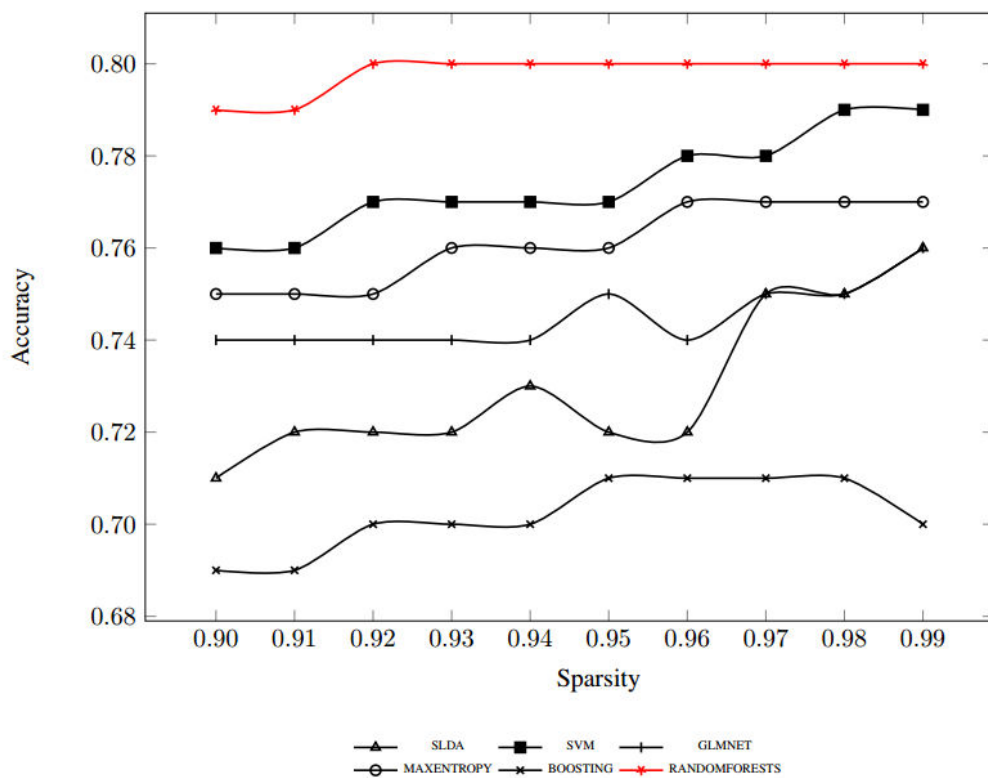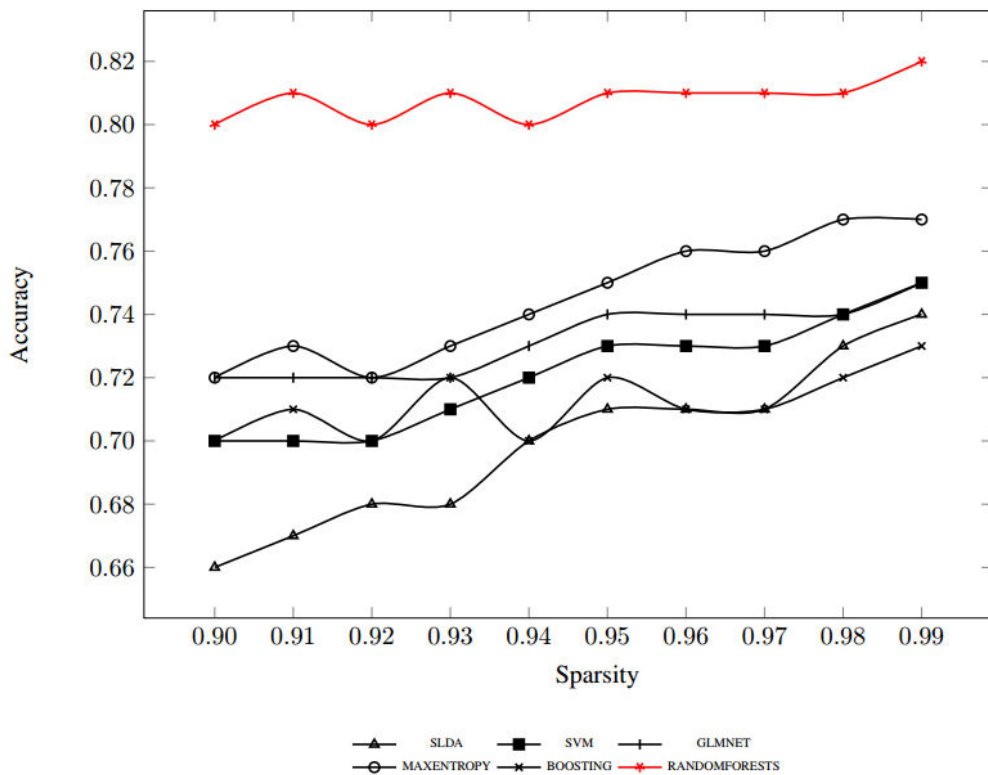

Figure 2. Term frequency-inverse document frequency-generated features using 10-splits Monte Carlo cross-validation accuracy.

Discussion

The results of this study indicate that a machine learning approach can be effective in identifying patient complaints that involve physicians. It is interesting that using term sparsity to reduce the feature set provides robust improvement until we arrive at a point where the terms are too few to provide any meaningful discrimination between the labels and, thus, the prediction accuracy falls. Adding IDF adjusts the weights assigned by TF in TF-IDF, which helps remove features that do not contribute significant information. Although common terms would be more prone to appear in the TF less sparse terms, the TF-IDF would have removed those terms before we get to this point. Our results are consistent with prior research (eg, Liu et al [30] and Cho and Lee [31]), showing improved results with a reduced (and hence a more representative) set of features. The insight here is that although reducing the number of features

leads to better prediction performance, knowing which features to keep plays a significant role as well.

Our specific findings are that the best-performing classifier was random forests with 82% accuracy and 81% F-score using TF-IDF for feature generation, followed by the SVM classifier, which achieved 79% accuracy using the simpler TF for feature generation. Adopting our automated approach would lead to the identification of patient complaints that should be shared with a physician much faster than any manual approach and thereby encourage thoughtful review and potential improvements.

Error Analysis

Error analysis is a critical step to understanding the failure mode of the classifiers [32]. We attempted to understand the general trends underpinning the classifier error. In Table 3, we show the percentage of total false prediction, positives and negatives, versus the number of classifiers that shared the error prediction.

Table 3. Classifier error analysis (n=3010).

Number of classifiers sharing an error prediction	% of errors
6	17.97
5	43.99
4	1.99
3	1.00
2	1.00
1	33.99

We note that in 61.96% (1865/3010) of cases, at least five of the six used classifiers shared the erroneous prediction. In 98.80% (2974/3010) of those cases, the classifiers predicted that the complaint required physician action, although it did not.

We wanted to understand why the classifiers were confused in this specific manner. In analyzing the complaints, a pattern emerged. The complaints mainly shared a few topics: patient falling, medical records, or billing issue. The terms used in those complaints contain a mix of both cases because a physician may be involved or mentioned in those cases and the complaint topic does not require physician action. The insight we draw from our error analysis is that although TF-IDF provides a good approach for weighting the features, it is not sensitive enough to distinguish mixed cases. Potential methods for alleviating the errors that appear in patient falling, medical records, and billing issues would potentially include using dependency-based features [33,34] to capture contextual information or a health care-specific lexicon.

Limitations

Modeling the content of patient complaints is a challenging problem. We limited feature extraction to TF and TF-IDF, which although generating robust results, still leaves unanswered the question of whether more useful data could yet be extracted. Using TF-IDF does not always work well. For example, the

term "doctor" was very frequent and is an important feature, although it was not determined to be important using TF-IDF due to the prevalence of the term in the medical domain. TF-IDF can easily confuse such terms with more noisy terms as illustrated with the term "her."

Exploring more advanced NLP methods to dive into the underlying language structure and reduce the noise would represent a potential future line of inquiry. Although 82% accuracy and 81% F-score is a promising start in regards to our specific problem, extracting better features may help improve the accuracy. Another limitation of our work is our focus on the binary classification we have used. Patient complaints involving physicians' practices are not all the same; rather, some may be treatment concerns, environmental issues, physician behavioral issues, or competency questions. It would be interesting to expand our scope to address those issues.

Future Directions

A future direction is to extend methods outlined by Tausczik and Pennebaker [35] and Zhang and Singh [36] to build a lexicon specific to health care complaints, which could yield superior metrics such as accuracy. Another interesting direction is to evaluate the influence of geography; specifically, do patients from different locations express themselves differently and do their differences in phrasing affect the underlying meaning?

XSL•FO

RenderX

Acknowledgments

Conflicts of Interest

None declared.

References

1. Hayden AC, Pichert JW, Fawcett J, Moore IN, Hickson GB. Best practices for basic and advanced skills in health care service recovery: a case study of a re-admitted patient. Jt Comm J Qual Patient Saf 2010 Jul;36(7):310-318. [Medline: 21226384]

2. Hickson GB, Pichert JW, Webb LE, Gabbe SG. A complementary approach to promoting professionalism: identifying, measuring, and addressing unprofessional behaviors. Acad Med 2007 Nov;82(11):1040-1048. [doi: 10.1097/ACM.0b013e31815761ee] [Medline: 17971689]

3. The Joint Commission. 2016. Sentinel event data summary URL: https://www.jointcommission.org/sentinel_event_statistics_quarterly/ [accessed 2017-06-20] [WebCite Cache ID 6rNP5tkc3]

4. Catron TF, Guillamondegui OD, Karrass J, Cooper WO, Martin BJ, Dmochowski RR, et al. Patient complaints and adverse surgical outcomes. Am J Med Qual 2016 Sep;31(5):415-422. [doi: 10.1177/1062860615584158] [Medline: 25916627]

5. Bendall-Lyon D, Powers TL. The role of complaint management in the service recovery process. Jt Comm J Qual Improv 2001 May;27(5):278-286. [Medline: 11367775]

6. de Luca LM, Atuahene-Gima K. Market knowledge dimensions and cross-functional collaboration: examining the different routes to product innovation performance. J Marketing 2007 Jan;71(1):95-112. [doi: 10.1509/jmkg.71.1.95]

7. Pichert JW, Moore IN, Karrass J, Jay JS, Westlake MW, Catron TF, et al. An intervention model that promotes accountability: peer messengers and patient/family complaints. Jt Comm J Qual Patient Saf 2013 Oct;39(10):435-446. [Medline: 24195197]

8. Baud RH, Rassinoux AM, Scherrer JR. Natural language processing and semantical representation of medical texts. Methods Inf Med 1992 Jun;31(2):117-125. [Medline: 1635463]

9. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA 2011 Aug 24;306(8):848-855. [doi: 10.1001/jama.2011.1204] [Medline: 21862746]

10. Bejan CA, Denny JC. Learning to identify treatment relations in clinical text. In: AMIA Annu Symp Proc. 2014 Presented at: 17th Annual Symposium of American Medical Informatics Association (AMIA); Nov 15-19, 2014; Washington, DC p. 282-288.

11. Licong C, Shiqiang T, Guo-Qiang Z. A semantic-based approach for exploring consumer health questions using UMLS. In: AMIA Annu Symp Proc. 2014 Presented at: 17th Annual Symposium of American Medical Informatics Association (AMIA); Nov 15-19, 2014; Washington, DC p. 432-441.

12. Sakai AM, Rossaneis MA, Haddad M, Sardinha D. Feelings of nurses in the reception and risk classification evaluation in the emergency room. Northeast Netw Nurs J 2016 May 2;17(2):233-241. [doi: 10.15253/2175-6783.2016000200011]

13. Zhou L, Melton GB, Parsons S, Hripcsak G. A temporal constraint structure for extracting temporal information from clinical narrative. J Biomed Inform 2006 Aug;39(4):424-439 [FREE Full text] [doi: 10.1016/j.jbi.2005.07.002] [Medline: 16169282]

14. Garla V, Re VL, Dorey-Stein Z, Kidwai F, Scotch M, Womack J, et al. The Yale cTAKES extensions for document classification: architecture and application. J Am Med Inform Assoc 2011;18(5):614-620 [FREE Full text] [doi: 10.1136/amiajnl-2011-000093] [Medline: 21622934]

15. Vuokko R, Mäkelä-Bengs P, Hyppönen H, Doupi P. Secondary use of structured patient data: interim results of a systematic review. Stud Health Technol Inform 2015;210:291-295. [Medline: 25991152]

16. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol 1993;78(1):98-104. [doi: 10.1037/0021-9010.78.1.98]

17. Gwet K. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters. Gaithersburg, MD: STATAXIS Publishing Company; 2001.

18. Aggarwal CC, Yu PS. Finding generalized projected clusters in high dimensional spaces. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. 2000 May Presented at: The 19th ACM International Conference on Management of Data (SIGMOD); May 15-18, 2000; Dallas, TX p. 70-81. [doi: 10.1145/342009.335383]

19. Xu Q, Liang Y. Monte Carlo cross validation. Chemometr Intell Lab 2001 Apr;56(1):1-11. [doi: 10.1016/S0169-7439(00)00122-2]

20. Wilcox AB, Hripcsak G. The role of domain knowledge in automating medical text report classification. J Am Med Inform Assoc 2003;10(4):330-338 [FREE Full text] [doi: 10.1197/jamia.M1157] [Medline: 12668687]

21. Rajaraman A, Ullman JD. Mining of Massive Datasets. Cambridge: Cambridge University Press; 2012.

XSL•FO
RenderX

22. Dumais S, Platt J, Heckerman D, Sahami M. Inductive learning algorithms and representations for text categorization. In: Proceedings of the 7th International Conference on Information and Knowledge Management. 1998 Nov Presented at: The 7th International Conference on Information and Knowledge Management (CIKM); Nov 2-7, 1998; Bethesda, MD p. 148-155.

23. Saif H, Fernández M, He Y, Alani H. On stopwords, filtering and data sparsity for sentiment analysis of Twitter. In: Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC). 2014 Presented at: The 9th International Conference on Language Resources and Evaluation (LREC); May 26-31, 2014; Reykjavik, Iceland p. 810-817.

24. Jurka TP, Collingwood L, Boydstun AE, Grossman E, Van Atteveldt W. RTextTools: automatic text classification via supervised learning. The R Journal 2013 Jun;5(1):6-12.

25. Martínez AM, Kak AC. PCA versus LDA. IEEE T Pattern Anal 2001;23(2):228-233. [doi: 10.1109/34.908974]

26. Peng T, Zuo W, He F. SVM based adaptive learning method for text classification from positive and unlabeled documents. Knowl Inf Sys 2007 Sep 6;16(3):281-301. [doi: 10.1007/s10115-007-0107-1]

27. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]

28. Osborne M. Using maximum entropy for sentence extraction. In: Proceedings of the ACL-02 Workshop on Automatic Summarization. 2002 Jul Presented at: The Association for Computational Linguistics (ACL) Workshop on Automatic Summarization; Jul 11-12, 2002; Philadelphia, PA p. 1-8.

29. Schapire RE. The strength of weak learnability. Mach Learn 1990 Jun;5(2):197-227. [doi: 10.1007/BF00116037]

30. Liu T, Liu S, Chen Z, Ma WY. An evaluation on feature selection for text clustering. In: Proceedings of the Twentieth International Conference on International Conference on Machine Learning. 2003 Aug Presented at: The 20th International Conference on Machine Learning (ICML); Aug 21-24, 2003; Washington, DC p. 488-495.

31. Cho H, Lee JS. Data-driven feature word selection for clustering online news comments. In: Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp). 2016 Jan Presented at: The 3rd International Conference on Big Data and Smart Computing (BigComp); Jan 18-20, 2016; Hong Kong p. 494-497.

32. Vo DT, Zhang Y. Target-dependent Twitter sentiment classification with rich automatic features. In: Proceedings of the 24th International Conference on Artificial Intelligence. 2015 Jul Presented at: The 24th International Conference on Artificial Intelligence; Jul 25-31, 2015; Buenos Aires p. 1347-1353.

33. Talavera L. Dependency-based feature selection for clustering symbolic data. J Intell Data Anal 2000 Jan;4(1):19-28.

34. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005 Aug;27(8):1226-1238. [doi: 10.1109/TPAMI.2005.159] [Medline: 16119262]

35. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2010 Mar;29(1):24-54. [doi: 10.1177/0261927X09351676]

36. Zhang Z, Singh MP. ReNew: a semi-supervised framework for generating domain-specific lexicons and sentiment analysis. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL). 2014 Jun Presented at: The 52nd Annual Meeting of the Association for Computational Linguistics (ACL); Jun 23-25, 2014; Baltimore, MD p. 542-551.

Abbreviations

NLP: natural language processing
PARS: Patient Advocacy Reporting System
SLDA: scaled linear discriminant analysis
SVM: support vector machines
TF: term frequency
TF-IDF: term frequency-inverse document frequency

XSL•FO
RenderX

Original Paper

What Patients Can Tell Us: Topic Analysis for Social Media on Breast Cancer

Mike Donald Tapi Nzali[1,2], PhD; Sandra Bringay[2,3], PhD; Christian Lavergne[1,3], PhD; Caroline Mollevi[4], PhD; Thomas Opitz[5], PhD

[1]Institut Montpelliérain Alexander Grothendieck (IMAG), Department of Mathematics, Montpellier University, Montpellier, France

[2]Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Department of Computer Science, Montpellier University, Montpellier, France

[3]Paul Valery University, Montpellier, France

[4]Biometrics Unit, Institut du Cancer Montpellier (ICM), Montpellier, France

[5]BioSP Unit, Institut National de la Recherche Agronomique (INRA), Avignon, France

Corresponding Author:
Mike Donald Tapi Nzali, PhD
Institut Montpelliérain Alexander Grothendieck (IMAG)
Department of Mathematics
Montpellier University
Place Eugène Bataillon
Case Courrier 051
Montpellier, 34095
France
Phone: 33 4 67 41 86 72
Fax: 33 4 67 41 85 00
Email: tapinzali@lirmm.fr

Abstract

Background: Social media dedicated to health are increasingly used by patients and health professionals. They are rich textual resources with content generated through free exchange between patients. We are proposing a method to tackle the problem of retrieving clinically relevant information from such social media in order to analyze the quality of life of patients with breast cancer.

Objective: Our aim was to detect the different topics discussed by patients on social media and to relate them to functional and symptomatic dimensions assessed in the internationally standardized self-administered questionnaires used in cancer clinical trials (European Organization for Research and Treatment of Cancer [EORTC] Quality of Life Questionnaire Core 30 [QLQ-C30] and breast cancer module [QLQ-BR23]).

Methods: First, we applied a classic text mining technique, latent Dirichlet allocation (LDA), to detect the different topics discussed on social media dealing with breast cancer. We applied the LDA model to 2 datasets composed of messages extracted from public Facebook groups and from a public health forum (cancerdusein.org, a French breast cancer forum) with relevant preprocessing. Second, we applied a customized Jaccard coefficient to automatically compute similarity distance between the topics detected with LDA and the questions in the self-administered questionnaires used to study quality of life.

Results: Among the 23 topics present in the self-administered questionnaires, 22 matched with the topics discussed by patients on social media. Interestingly, these topics corresponded to 95% (22/23) of the forum and 86% (20/23) of the Facebook group topics. These figures underline that topics related to quality of life are an important concern for patients. However, 5 social media topics had no corresponding topic in the questionnaires, which do not cover all of the patients' concerns. Of these 5 topics, 2 could potentially be used in the questionnaires, and these 2 topics corresponded to a total of 3.10% (523/16,868) of topics in the cancerdusein.org corpus and 4.30% (3014/70,092) of the Facebook corpus.

Conclusions: We found a good correspondence between detected topics on social media and topics covered by the self-administered questionnaires, which substantiates the sound construction of such questionnaires. We detected new emerging topics from social media that can be used to complete current self-administered questionnaires. Moreover, we confirmed that social media mining is an important source of information for complementary analysis of quality of life.

XSL•FO
RenderX

Introduction

Social media such as Facebook, Twitter, or Internet forums dedicated to health-related topics have evolved into easily accessible participatory tools for the exchange of knowledge, experience, and opinions through structured collections of text documents [1]. Online health forums are used by patients to exchange information [2]. Patients maintain their anonymity while discussing freely with other patients. Whereas communication with doctors and the medical staff in hospitals mainly revolve around technical issues of the disease and treatment, social media give patients access to more general exchanges of information, experiences, and mutual support among former and current patients [3]. Such forums can therefore be considered as a valuable resource for the study of health-related quality of life (QoL). As shown by some studies (eg, [4]), the anonymous environment of social media facilitates the unbiased expression of opinions and of feelings such as doubt or fear. Internet users have been shown to be primarily interested in specific information on health problems or diseases [5-7] and in adopting a healthier lifestyle and looking for alternative points of view [5]. Here we propose an approach to structure and evaluate clinically relevant information in narratives extracted from online health social media, with a focus on the QoL of patients with breast cancer.

While constant progress in medical science leads to new treatments and improved chances to prolong lives, such treatments can be difficult to undergo. QoL can be considered as an alternative clinical end point in this context, moving the focus away from quantity to quality [8-11]. QoL falls within the scope of patient-reported outcomes; that is, measures of perceived health [12,13]. These measures must therefore be reported by patients themselves. For instance, alternative treatments such as palliative treatment of terminal cancer may be less efficient from a traditional clinical stance but may still be preferable with respect to the patients' QoL [14,15]. Moreover, health economists must take into account the expense of treatments with respect to their effective benefits, for instance measured by the improvement in QoL (see Hirth et al [16] and Cutler and McClellan [17] for a general discussion, and Hillner and Smith [18] for a cost-effectiveness study of chemotherapy in certain cases of breast cancer).

Since QoL is a multidimensional, subjective, and culture-dependent concept, its quantification is not as straightforward, as shown in the literature review of Garratt et al [19]. This concept includes at least physical, psychological, and social well-being, as well as symptoms related to illness and treatment. Today, QoL is assessed in cancer clinical trials by self-administered questionnaires developed by the European Organization for Research and Treatment of Cancer (EORTC). The EORTC Quality of Life Questionnaire Core 30 (QLQ-C30) [20] is a generic self-administered questionnaire often associated with disease-specific modules, such as the EORTC breast cancer module (QLQ-BR23). The EORTC QLQ-C30 contains 30 items and evaluates 15 dimensions of QoL: 5 functional scales, 1 QoL and global health status scale, and 8 symptomatic scales, as well as 1 scale measuring the financial difficulties associated with the disease. The EORTC QLQ-BR23 contains 23 questions. It is usually administered with the EORTC QLQ-C30 and is designed to measure QoL for breast cancer patients at various stages and with different treatment modalities. The evaluation consists of 4 functional scales and 4 symptomatic scales. Usually, self-administered questionnaires evaluate functional and symptomatic dimensions and are filled in at a predefined time of the study protocol, such as at baseline, during treatment, and at follow-up. In this context, an advantage of social media is that they allow patients to leave a written trace of their sentiment at any time, therefore avoiding potential self-reporting bias owing to a change of perception due to time lag.

Opitz et al [21] developed an automated approach for the supervised detection of topics defined in QLQ-BR23 questionnaire items for cancerdusein.org, a French forum specialized in breast cancer. In this new work, we used an unsupervised method to discover topics covered by health social media. Unsupervised methods have been successfully applied to biomedical data. For example, Arnold and Speier [22] presented a topic model tailored to the clinical reporting environment that allows for individual patient timelines. Lu et al [23] used text clustering algorithms on social media data to discover health-related topics. Zhang et al [24] applied a convolutional neural network classifier to an online breast cancer community and carried out a longitudinal analysis to show topic distributions and topic changes throughout the members' participation. In our study, the main medical application was to help improve questionnaires by including new topics of interest for patients (topics frequently discussed by patients and the impact on QoL) as new items in the questionnaires.

Researchers have developed several topic models, including latent semantic analysis [25], probabilistic latent semantic analysis [26], latent Dirichlet allocation (LDA) [27], and latent semantic indexing [28]. In this study, we defined a general process based on LDA [27] and applied this model to social media. LDA, an unsupervised generative probabilistic method for modeling a corpus, is the most commonly used topic modeling method. The main disadvantage of LDA is that there are no objective metrics that justify the choice of the hyperparameters. However, the main advantage of LDA is that it is a probabilistic model with interpretable topics. Nowadays, a growing number of probabilistic models are based on LDA and dedicated to particular tasks. For example, Zhan et al [29] used LDA to identify topics among posts generated by e-cigarette users in social media. Wang et al [30] and Paul and Dredze [31] constructed a specialized and advanced LDA model using biomedical terms to provide a more effective way of exploring the biomedical literature. LDA has also been successfully used for patient-generated data [32-36] and in

particular for online breast cancer discussions [3,24]. Hao and Zhang [37] used LDA to examine what Chinese patients said about their physicians in 4 major specialty areas. Hao et al [38] used LDA to identify topics in positive and negative textual reviews of obstetricians and gynecologists from the 2 most popular online doctor rating websites in the United States and China. Yesha and Gangopadhyay [39] described methods to identify topics and patterns within patient-generated data related to suicide and depression. LDA has also been used as a feature to build machine learning models to automatically identify the extent to which messages contain emotional and informational support on online health forums dealing with breast cancer [40] or on Chinese social media [41].

Conducting automated research as we have done here is of considerable interest for processing a large amount of text obtained from social media. The LDA approach for extracting topics allows for better targeting for information exploration, reducing search time, and treating topics as a flat set of probability distribution; it can also be used to recover a set of topics from a corpus. In this work, we only used the LDA model and tuned parameters to align the topics found with QoL questionnaires. The originality of our approach is to automatically relate the topics obtained with the LDA method

to the questionnaire items with an adaptation of the Jaccard coefficient.

In this study, the purpose of our approach was diverse: (1) to provide a nonconventional analysis of QoL from social media and put the topics identified with this nonconventional analysis into perspective with those of classical QoL questionnaires collected in clinical trials (in particular in breast cancer: EORTC QLQ-C30 and QLQ-BR23); (2) to apply the LDA model to patient data with relevant pretreatments; (3) to index the narratives with respect to topics extracted through an unsupervised statistical analysis of forum content and to predefined topics from questionnaires used in cancer clinical trials; and (4) to discover new topics directly from patients' concerns that are not included in the current questionnaires used to evaluated QoL, with the possibility that these topics could be included in these questionnaires if sufficiently relevant.

Methods

Data

Data Description

In this work, we used datasets from 2 different social media sources: cancerdusein.org and Facebook groups. Table 1 summarizes statistics from these 2 datasets.

Table 1. Number of users, threads, and posts on a social network and a health forum analyzed in this study.

Characteristics	Health forum (cancerdusein.org)	Social network (Facebook groups)
Date	October 2010-October 2014	October 2010-October 2014
No. of users	675	1394
No. of discussion threads	1050	11,013
No. of messages	16,868	70,092

The first dataset contained the forum posts from cancerdusein.org, a French health forum with more than 16,000 posts. These posts cover a large number of topics related to health issues. This forum is recommended to patients in a brochure of the *Institut National du Cancer* (INCA), which is the French reference organization in oncology. The forum is recommended for patients to exchange information and find comfort and potential solutions to their problems. It serves as an online cancer support community, where cancer patients, cancer survivors, and their families share information about cancer and their conditions. The second dataset contains posts from groups on Facebook, one of the most well-known social networks. We extracted 70,092 posts from 4 different public groups or communities on Facebook: *Cancer du sein*, *Octobre rose 2014*, *Cancer du sein - breast cancer*, and *brustkrebs*. We collected data from groups focusing on the adult population (the targeted users) and in which users were very active.

On both social media platforms, patients freely exchange information without the need for moderators to supervise discussions. New messages can either be added to an existing thread or be posted to open a new thread. In cancerdusein.org, a thread appears in exactly 1 of the 13 predefined subforums, for example, *Discussion générale* [general discussion], *Vivre mon cancer au quotidien* [daily life with my cancer], *Les bonnes*

nouvelles [good news], or *Récidives et combats au long cours* [relapses and long-term battles]. In Facebook groups, there are no predefined topics to index the threads. Structuring topics according to the subforum structure is possible in cancerdusein.org, but this structure underlines the relatively uninformative and widely spread topics, covering a strongly unbalanced number of messages. Such indexing is not possible in Facebook groups. Interestingly, we propose to accomplish a finer analysis of topics in the next section, which further enables the presence of several topics within 1 message.

Data Preprocessing

Texts on social media are often strongly heterogeneous and noisy, with many deviations from standards of spelling, syntax, and abbreviations, which impede efficient natural language processing. The French language has a rich spelling and grammar, characterized by special characters such as *ç*, various kinds of accented vowels (eg, *é*, *è*, *ê*, *ë*, *â*, and *à*), and many flexional variants. Additional rules exist for linking subsequent terms in certain situations (eg, the contraction *du* formed from *de+le* and the contraction *des* formed from *de+les*). As a consequence, automatic correction of text not obeying those rules is relatively difficult in practice. Furthermore, semantic analysis of texts is complicated by a large number of homonymy

relationships: for example, *pas* can either mean *step* (noun) or can be the negation adverb *not*. As Balahur [42] and Farzindar and Inkpen [43] have pointed out, these linguistic peculiarities may affect classification performance. For this reason, we developed the following preprocessing steps.

- *Removal of user tags.* All user tags that have been identified in our corpus are removed, for example, @name, @surname.
- *Replacement of hyperlinks and email addresses.* All the hypertext links are replaced by the term "link" and all the email addresses are replaced by the term "mail." Hyperlinks (Internet, email, etc) are deleted. Emoticons are coded as :smile:, :sad:, etc.
- *Replacement of slang.* Some expressions frequently used on social media, such as lol, mdr[lol], and xD, are removed.
- *Lemmatization.* All words are lemmatized (using TreeTagger [44]).
- *Lowercasing.* Capitals letters are lowercased.
- *Removal of stopwords.*
- *Replacement of specific patient terms.* The texts for the 2 corpora are usually highly focused on a specific domain (breast cancer, in our case). Most often, as patients are laypersons in the medical field, they use slang, abbreviations, and their own vocabulary during their exchanges. To automatically analyze text from social networks, we need a specific vocabulary. In this work, we use the vocabulary created by Tapi Nzali et al [45] to replace the patients' terms with biomedical terms used by health professionals and presented in shared medical resources. For example, *crabe* [crab] is replaced by *cancer*, *onco* is replaced by *oncologue* [oncologist].
- *Correction of spelling.* Spelling correction is important to remove redundant dimensions of data and to improve part-of-speech tagging, which is the basis for many statistical and rule-based methods in natural language processing. We apply spelling correction based on specialized dictionaries constructed ad hoc and the open source tool GNU Aspell version 0.60.6.1, whose algorithm proposes a list of possible corrections for unknown terms from the corpus. We use the following ad hoc dictionaries: lists of breast cancer drugs and of secondary effects, and proper names extracted from forum metadata (usernames, user residence) and from narratives (terms with capital first letter not at the beginning of a sentence; usernames identified from salutations at the beginning of forum posts).
- *Extraction and deletion of forum pseudonyms.* All the pseudonyms, previously extracted from each website, are used. The pseudonyms are extracted and deleted if they exist in the post.

Unsupervised Topic Detection and Assigning

Modeling Topics With Latent Dirichlet Allocation

Today, detection of latent semantic structures and topics has become a very active field of research in the text mining community. We focused on the LDA model [27], which has become a standard model for unsupervised topic detection from a text corpus. It is a probabilistic model with a hierarchical definition of its components. With the LDA model, we generated new documents from a given model. Based on the relatively simple and robust bag-of-words representation of text documents, it leaves the order of occurrence of terms and sentence structure out for consideration. For a given corpus of D documents, we first defined the relevant vocabulary V, a preprocessed collection of terms occurring in the corpus. Typical preprocessing steps include spelling correction, lemmatization, and the removal of noisy or irrelevant terms. To define a topic t, we associated a nonnegative weight ω_{ti} with each of the vocabulary's terms, w_i, so that weights summed up to 1 ($\sum_{i=1}^{V} \omega_{ti}=1$). In practice, each topic typically consisted of a relatively small number of terms with nonnegligible weight. An LDA model uses a fixed number $K>1$ of topics. For each document d, weights $\omega_{dt} \geq 0$ indicate the occurrence probability of terms from topic t, where the sum of ω_{dt} over all topics t yields 1 ($\sum_{t=1}^{K} \omega_{dt}=1$). If document d contains l_d terms (or "positions"), we associated a topic t_{dj} with each of the positions $j=1,..., l_d$, where the probability of associating topic t is α_{dt}. Finally, each position was filled with a term, w_{dj}, from the vocabulary, where the probability of using term w_i is ω_{tdj}.

The corpus-generation model is proposed by the algorithm shown in Figure 1.

The principal information that we can learn from using such a model on a corpus of text data is the structure of represented topics and the distribution of topics over the documents contained in the corpus. The high number of unknown parameters in this model makes inference challenging, yet Bayesian techniques such as Gibbs sampling [46] have proven reliable. Based on prior assumptions about the distribution of the weights of terms in topics and of topics in documents on a range from very uniform to very spiked, these inference techniques are applied to the data to estimate the posterior distributions of the model. Most importantly, the most likely topic structure and the occurrence probabilities for topics in each document are proposed. In this work, we considered a message as a document.

Figure 1. Algorithm proposing the corpus-generation model.

- For each topic $t \, \epsilon \, \{1, \ldots, K\}$, choose a distribution over the vocabulary according to a

  Dirichlet prior $\beta_t = (\beta_{t1}, \ldots, \beta_{tV}) \sim Dirichlet(\lambda_\beta, \ldots, \lambda_\beta)$

- For each document $d \, \epsilon \, \{1, \ldots, D\}$

  - Choose a distribution over topics according to a Dirichlet prior

    $\alpha_d = (\alpha_{d1}, \ldots, \alpha_{dK}) \sim Dirichlet(\lambda_\alpha, \ldots, \lambda_\alpha)$

    Each α_{dt} indicates the proportion of occurrences of document d which are

    associated with the topic t.

  - For each position of term i in d, $i \, \epsilon \, \{1, \ldots, l_d\}$,

    - Choose a topic from the document's distribution over topics

      $T_{di} \sim Multinomial(\alpha_d)$

    - Choose a word from that topic $W_{di} \sim Multinomial(\beta_{T_{di}})$

Crucial Model Parameters

Besides K, 2 parameters often denoted as α and β strongly influence the distribution of topic probabilities for each of the messages. They are concentration parameters for the prior distributions of topics over a message (α) and of words over a topic (β). When α or β is smaller than 1 and decreases, prior mass concentrates closer and closer to the border of the simplex with spikes at each of its vertices. Then, 1 or fewer components (topics for α, words for β) carry strong probability in the mixture distribution. In the limit 0, a single component is selected with a probability of 1. On the contrary, when α or β is larger than 1 and increases, mass concentrates more and more in the barycenter of the simplex, leading to a mixture of the distribution, which is more and more balanced over all components. In the limit ∞, each component is selected with a probability of 1 over the number of components.

Now we will explain our choice of α based on the influence of α on the distribution of topic probabilities for messages and of term distributions for topics. When $\alpha=1$, the prior distribution for the vector of topic probabilities corresponds to a uniform distribution on the simplex with K vertices. As α increases, the distribution concentrates more and more strongly toward the center of the simplex, such that most of the probabilities are closer to $1/K$. As α decreases, it concentrates more and more strongly toward the vertices, leading to some probabilities being further away from $1/K$. For fixed α, probabilities concentrate more and more around $1/K$ as K increases. In Griffiths and Steyvers [47], values $\alpha=\alpha_0/K$ with the constant $\alpha_0=50$ are encouraged, where dividing through K constantly keeps a certain complexity measure of the model. Exploratory analysis showed

that $\alpha_0=50$ led to very flat probability vectors in our case, which made it difficult to attribute a small number of topics for indexation to each message. On the other hand, smaller values of α_0 led to topics becoming more difficult to interpret due to flatter distribution of term probabilities within topics and similar dominating terms in multiple topics. After careful analysis of topics and posterior distributions for a range of values of α_0, we decided to fix $\alpha_0=10$. Whereas higher values of α_0 yielded a better fit of the model in terms of its likelihood, it led to very flat posterior probabilities for the topic distribution of messages. As in Griffiths and Steyvers [47], we decided to fix the value of parameter β to 0.1 for our experiments.

There is evidence [48] that automatic choice of parameters through a model selection criterion may result in an unsatisfactory topic collection, whose interpretation is more challenging than topics associated with suboptimal values of the criterion. Often, the calculation of held-out likelihood is used, allowing for approaches such as likelihood cross-validation. However, the likelihood calculation is not trivial, and some standard methods produce inaccurate results (see [49]).

Vocabulary Definition

To avoid noisy topics that are difficult to interpret, it is useful to focus on terms with potential medical relevance. Here, we defined terms as sequences of words, and often there was only a single word. To begin, we used terms indexed in the French version of the Medical Subject Headings (MeSH) [50]. Then we added terms figuring in a list of breast cancer drugs (extracted from the online resource) or appearing in a list of

XSL•FO
RenderX

nonconventional treatments (extracted from the French Wikipedia entry). We denoted this term set as *MED*. We retained 481,111 occurrences of 18,672 terms in 16,868 messages on cancerdusein.org, and 626,043 occurrences of 18,741 terms in 70,092 messages on Facebook. The resulting topics, often strongly dominated by a single term, appeared to be rather difficult to interpret by clinical experts, possibly due to the relatively small dimension of the term-document space. We categorized terms figuring in the representative terms according to their grammatical role: nouns/proper names (*NN*), verbs (*V*), and adjectives (*A*). Then, we extracted topics by applying LDA to the original *MED* term set, extended by terms according to scenarios *MED+NN+V+A*. Based on the exploratory inspection of topics extracted by LDA in the approaches presented in the following, we further removed a small number of strongly

represented terms leading to strong noise (*femme* [woman], *temps* [time or weather]), and medically meaningless topics.

Align Topics and Questionnaires

With the topics returned by the LDA model, we automatically identified correspondences between the topics and the questionnaires, as shown in Figure 2. To align topics and questionnaires, we computed a distance between each question q_j and all topics t_i in *T*. We kept the topic with the higher distance. To compute the distance between an LDA topic and an item of the questionnaire, we customized the Jaccard coefficient [51] by taking into account the probability of the words obtained with the LDA model, as shown in Figure 3 (equation 1).

Figure 2. Automatic identification of correspondences between topics and questionnaires. LDA: latent Dirichlet allocation; *MED + NN + V + A*: set of medically relevant terms (*MED*) extended by terms categorized by their grammatical role (*NN*: nouns and proper names; *V*: verbs; *A*: adjectives).

Let $Q = \{q_1, q_2, \ldots, q_{53}\}$ be the set of items in the questionnaires and $T = \{t_1, t_2, \ldots, t_K\}$ be the set of topics returned by the LDA model.

For a given topic t_i, let $W_i = \{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$ and $P_i = \{p_{i_1}, p_{i_2}, \ldots, p_{i_n}\}$ be the set of words and the set of corresponding probabilities of the words.

For example, w_{2_2} is the second word of topic t_2 and p_{2_2} is the probability of word w_{2_2} for the topic t_2.

For a given question q_j, let $W'_j = \{w'_{j_1}, w'_{j_2}, \ldots, w'_{j_m}\}$ be the set of m words obtained after lemmatization of the words of the question and the selection according to the scenario *MED+NN+V+A*.

Figure 3. Equation to calculate the distance between a latent Dirichlet allocation topic and an item of the questionnaire.

Let L be the number of common words between W_i and W'_j, such as $W_i \cap W'_j = \{w_{l_1}, w_{l_2}, \ldots, w_{l_L}\}$ and $P_L = \{p_{l_1}, p_{l_2}, \ldots, p_{l_L}\}$ be the set of corresponding probabilities of the words.

$$Distance(t_i, q_j) = \frac{\sum_{k=1}^{L} p_{l_k}}{|W_i \cup W'_j|} \quad (1)$$

Results

Topic Modeling Result

To run experiments, we used the R package LDA [52] and the R environment version 3.2.5 (R Foundation) for the implementation. We tested different scenarios, and an expert validated and labeled the topics and verified the association between topics and questionnaires items. The expert is a biostatistician and QoL researcher in the cancer field [53,54].

In scenario *MED + NN*, most of the topics were of a factual nature, whereas scenario *MED + NN + V* led to a more complete description of topics, where verbs often add information about actions undertaken by users and other stakeholders (wait,

consult, seek, support, etc) and about user sentiment (feel, cry, tire, fear, accept, etc). In scenario *MED + NN + V + A*, several topics consisting mainly of emotional words were difficult to interpret from a medical point of view. We reported the stability of the majority of topics that were identified through the scenarios *MED + NN*, *MED + NN + V*, and *MED + NN + V + A* due to the similarity of dominating terms. After careful analysis, we narrowed down the choice of *K* to a value between 20 and 30. With more than 20 topics, we found duplication of topics (2 topics may deal with the same subject). In addition, some are unable to be interpreted (the medical expert found no meaning). Consequently, we decided to retain scenario *MED + NN + V + A* with 20 topics. Finally, we fixed *K*=20 for the duration of this study. For each topic, we showed only 20

XSL•FO
RenderX

keywords having higher probabilities under that topic. These keywords were presented to the expert. Table 2 and Table 3 list the topic modeling results of the 2 corpora. We show the top 10 keywords for each topic. Table 4 shows the results of the 20 topics interpreted by the medical expert on the 2 corpora.

Relationships Between Questionnaire Topics

In this work, we used 2 QoL questionnaires (EORTC QLQ-C30 and EORTC QLQ-BR23) to look for relationships between the studied dimensions in these previous questionnaires and topics that we interpreted. The EORTC QLQ-C30 is a 30-item, self-administered, cancer-specific questionnaire designed to measure QoL in the cancer population. The assessment comprises 5 functional scales (physical, role, cognitive, emotional, and social), 8 symptomatic scales (fatigue, nausea and vomiting, pain, dyspnea, insomnia, loss of appetite, constipation, and diarrhea), and 1 scale measuring financial difficulties and 1 measuring global health status and QoL by a score ranging from 0 to 100 through the 30 items [20]. The EORTC QLQ-BR23 is a 23-item, self-administered, breast cancer-specific questionnaire, usually administered with the EORTC QLQ-C30, designed to measure QoL in the breast cancer population at various stages and with patients with differing treatment modalities. The assessment comprises 4 functional scales (body image, sexual functioning, sexual enjoyment, and future perspective) and 4 symptomatic scales (systemic therapy side effects, breast symptoms, arm symptoms, and hair loss) [55]. The EORTC health-related QoL questionnaires are built on a Likert scale with polytomous items.

To find the theme corresponding to a question, we used equation 1 (Figure 3) proposed above. We obtained the following relationships:

- Topic *sexuality* is related to items 44 (To what extent were you interested in sex?) and 45 (To what extent were you sexually active?).
- Topic *hair loss* is related to item 34 (Have you lost any hair?).
- Topic *body care and body image during cancer* is related to items 39 (Have you felt physically less attractive as a result of your disease or treatment?) and 40 (Have you been feeling less feminine as a result of your disease or treatment?).

These relationships were validated by a medical expert. Following validation of the results, we calculated the precision. On cancerdusein.org data, for the 53 items, 39 relationships with topics were validated by the medical expert and 14 were invalidated, for a precision of 74%. On Facebook data, for the 53 items, 36 relationships were validated by the medical expert and 17 were invalidated, for a precision of 68%. The medical expert also manually examined the invalidated relationships. This step reduced the time spent by the expert to find relationships between the questions and the topics. The obtained precision rates can be explained by the fact that the items of the questionnaires are composed of very short sentences. On average, these sentences contain fewer than 5 words.

Table 2. Top 10 frequently occurring words for the first 10 topics (among the 20 found) on cancerdusein.org forum data.

Topic no.	Top 10 words with their translation		Topic label[a]
	French	English translation	
1	*cheveu, perdre, perruque, tomber, tête, commencer, repousser, chimiothérapie, perte, foulard*	hair, lose, wig, fall, head, begin, regrowth, chemotherapy, loss, scarf	Hair loss
2	*prendre, temps, travail, demander, soin, reprendre, charge, travailler, aide, payer*	take, time, job, ask, care, restart, charge, work, help, pay	Work life during cancer and financial aspects
3	*effet, chimiothérapie, secondaire, cure, douleur, passer, mammographie, nausée, docétaxel, fatigue*	effect, chemotherapy, secondary, treatment, pain, pass, mammography, nausea, docetaxel, fatigue	Chemotherapy and its secondary effects
4	*prendre, effet, douleur, traitement, problème, tamoxifène, prise, penser, secondaire, arrêter*	take, effect, pain, treatment, problem, tamoxifen, catch, think, secondary, stop	Hormone therapy and its secondary effects
5	*sein, bras, chirurgie, reconstruction, opération, douleur, prothèse, opérer, enlever, cicatrice*	breast, arm, surgery, reconstruction, operation, pain, prosthesis, operate, remove, scar	Breast reconstruction
6	*baiser, petit, beau, super, attendre, soutien, nouveau, guerrier, grand, vérité*	kiss, little, beautiful, great, wait, support, new, warrior, big, truth	Support from patient's family and friends
7	*ongle, peau, radiothérapie, main, séance, pied, rayon, brûlure, crème, conseil*	nail, skin, radiotherapy, hand, session, foot, radius, burn, cream, council	Radiotherapy and its secondary effects
8	*prendre, manger, boire, essayer, miel, aider, produit, demander, santé, complément*	take, eat, drink, try, honey, help, product, ask, health, complement	Complementary and alternative medicine
9	*lire, forum, message, venir, nouveau, donner, trouver, site, réponse, écrire*	read, forum, message, come, new, give, find, site, response, write	Media and forum information exchange
10	*homonymie, enfant, fille, maman, vie,* cancer, *vérité, vivre, malade, famille*	homonymy, child, girl, mom, life, cancer, truth, live, sick, family	Family background and breast cancer

[a]Topic label was assigned by a medical expert.

XSL•FO
RenderX

Table 3. Top 10 frequently occurring words for the first 10 topics (among the 20 topics found) on Facebook data.

Topic no.	Top 10 words		Topic label[a]
	French	English translation	
1	*voir, attendre, résultat, médecin, oncologie, examen, biopsie, mammographie, contrôle, scanner*	see, wait, result, doctor, oncology, examination, biopsy, mammography, test, scanner	Diagnosis
2	*douleur, effet, chimiothérapie, secondaire, jour, prendre, mal, fatigue, nausée, chaleur*	pain, effect, chemotherapy, secondary, day, take, bad, fatigue, nausea, heat	Chemotherapy and its secondary effects
3	*justice, moral, garder, aller, fort, dureté, battre, étape, force, combat*	justice, morale, keep, go, strong, hardness, beat, step, strength, fight	Breast cancer as a daily battle
4	*cheveu, perdre, tomber, repousser, perruque, couper, raser, tête, joli, foulard*	hair, lose, fall, growth, wig, cut, shave, head, beautiful, scarf	Hair loss
5	*prendre, suivre, dire, soin, arrêter, traitement, tamoxifène, poids, perdre, homonymie*	take, follow, tell, care, stop, treatment, tamoxifen, weight, lose, homonymy	Secondary effect of treatment
6	*aller, justice, passer, sexologie, allergologie, baiser, penser, meilleur, voir, reposer*	go, justice, pass, sexology, allergology, kiss, think, best, see, rest	Body care and body image during cancer
7	*homonymie, dire, vérité, suivre, peur, sexologie, comprendre, croire, dureté, enfant*	homonymy, tell, truth, follow, fear, sexology, understand, believe, hardness, child	Family background and breast cancer
8	*demander, suivre, droit, travail, aide, médecin, payer, charge, travailler, donner*	ask, follow, law, job, help, doctor, pay, charge, work, give	Work life during cancer and financial aspects
9	*sein, opération, reconstruction, enlever, bras, opérer, mastectomie, cicatrice, retirer, prothèse*	breast, operation, reconstruction, remove, arm, operate, mastectomy, scar, withdraw, prosthesis	Breast reconstruction
10	*suivre, aller, fille, sol, voir, rire, regarder, marier, croire, lire*	follow, go, girl, ground, see, laugh, look, marry, believe, read	Support from patient's family and friends

[a]Topic label was assigned by a medical expert.

Table 4. List of identified topic titles with $K=20$ in collaboration with an expert.

Topic no.	cancerdusein.org	Facebook
1	Hair loss	Diagnosis
2	Work life during cancer and financial aspects	Chemotherapy and its secondary effects
3	Chemotherapy and its secondary effects	Breast cancer as a daily battle
4	Hormone therapy and its secondary effects	Hair loss
5	Breast reconstruction	Secondary effects of treatments
6	Support from patient's family and friends	Body care and body image during cancer
7	Radiotherapy and its secondary effects	Family background and breast cancer
8	Complementary and alternative medicine	Work life during cancer and financial aspects
9	Media and forum information exchange	Breast reconstruction
10	Family members with breast cancer	Support from patient's family and friends
11	Treatment period	Interaction with nurses and doctors
12	Everyday life during cancer	Anxiety and fatigue
13	Healing	Healing of family member
14	Search for medical information	Relapse
15	Mourning	Sexuality
16	Diagnosis	Body care and body image during cancer
17	Breast cancer as a daily battle	Family members with breast cancer
18	Body care and body image during cancer and sexuality	Healing
19	Surgery	Support from patient's family and friends
20	Waiting for results of analysis, concerns	Treatment period

XSL·FO

RenderX

Table 5 shows the relationships between topics from questionnaires and those we found in the 2 corpora. The first column lists the topics of the 2 questionnaires, with the corresponding questionnaires items shown in column 2. Columns 3 and 4 give the corresponding topics obtained with LDA in the 2 corpora. Table 6 shows the percentage of documents belonging to each topic in cancerdusein.org and Facebook. We noticed that the numbers of messages belonging to each topic are almost equal; this shows the importance of all the topics that we found and that were discussed by patients.

Data From cancerdusein.org

We succeeded in interpreting the 20 topics obtained from the output of our model on the cancerdusein.org corpus. Table 2 presents the 10 first topics and the top 10 words obtained by our model that were interpreted by an expert. Some relationships were established. In the QLQ-C30, we found matches for all of the topics except for global health status and QoL. In the QLQ-BR23 form, we matched all of the topics.

Data From Facebook

We succeeded in interpreting the 20 topics obtained from the output of our model on the Facebook corpus. Table 3 presents the 10 first topics and the top 10 words obtained by our model that were interpreted by an expert. Some relationships were established. In the QLQ-C30, we found matches for all of the topics except for role functioning, cognitive functioning, and global health status and QoL. In the QLQ-BR23 form, we matched all of the topics.

Discussion

We have presented what we believe to be the first study of health social media data in French, as a potential source of analysis of the QoL for breast cancer patients. We used accurate machine learning models to identify topics discussed in online breast cancer support groups. Then we examined the relationships between the discovered topics and studied dimensions from QoL self-administered questionnaires. Exploratory and in-depth analysis of these data is a potential source of candid information as an alternative to analysis of QoL based on self-administered questionnaires.

Limitations

Patient-Authored Text

The first limitation of this study is the type of users, which produced the patient-authored text exploited in our process. Indeed, unless a group has formal gatekeeping of members, it is difficult to know for sure whether people posting to a forum or in a Facebook group are patients, survivors, health care professionals, care providers, family, or friends of patients. Consequently, topics extracted with our method may have been generated by users who do not have breast cancer. In particular, it has been known for decades that health information is sought principally by friends or family members, and then after that by patients [56]. In this work, we assumed that the relatives' topics of interest were similar to patients' topics of interest.

However, in a previous work [57], we proposed a method to automatically deduce the role of the forum user. This method can be used at the beginning of our chain to exclude the posts of individuals who are not actual patients.

Generalization of the Method

The second limitation is that we harvested data from only 1 forum and different Facebook groups. However, this forum is frequently recommended by French physicians to patients. It is also recommended by INCA, which is the French reference organization in oncology. We deliberately selected this forum and these Facebook groups to examine similarities and differences within and between these 2 particular communities. Of course, there are certainly many other online communities related to breast cancer, and the users in these 2 online communities were not necessarily representative of users of all breast cancer social media.

It is also important to note that our method can be easily applied to other diseases. For example, we can (1) use brain cancer forum data to align topics discussed by patients with items of the EORTC QLQ-C30 and the brain cancer module (QLQ-BN20) [58] questionnaires, and (2) use lung cancer forum data to align topics discussed by patients with items of the QLQ-C30 and the lung cancer module (QLQ-LC13) [59] We have already also applied a similar approach to study other social media data such as Twitter [60]. The main adaptation is relative to the acquisition of the patient terms, which are specific to the disease and the social media as mentioned in the Data Preprocessing section above.

Latent Dirichlet Allocation Model

A third limitation was the choice of LDA. LDA requires much manual tuning of its parameters, which vary from task to task. We spent a lot of time finding the best parameters so that the results could be interpreted meaningfully. Such analysis makes itself a sort of "overfitting" to the task at hand, making it very hard to generalize the method to other datasets and other tasks. However, we efficiently defined parameters of 2 types of text (forum and Facebook posts), which can be reused for other studies on comparable corpora.

Topics covered on social media focused on a specific domain, breast cancer. It was difficult to adjust the number of topics because topics were closed: all of the users were discussing breast cancer. When we adjusted the model and sought the optimal K with methods such as those used in other studies (eg, [47,61,62]), we obtained more than 50 topics. An interesting perspective was using the heuristic approach defined by Zhao et al [63] to determine an appropriate number of topics. This method is based on the rate of perplexity change [62,64]. This measure is commonly used in information theory to evaluate how well a statistical model describes a dataset, with lower perplexity denoting a better probabilistic model [63]. Finally, as in Arnold et al [65], we observed that an expert is not able to interpret so many topics. In this study, we manually fixed K=20. We interpreted all the topics with minimal redundancies.

XSL·FO

RenderX

Table 5. Distribution of documents on each topic on cancerdusein.org and Facebook.

Questionnaires and their scales	Questionnaire items	cancerdusein.org	Facebook
EORTC QLQ-C30[a]			
Functional scales			
Physical functioning	1-5	Everyday life during cancer	Treatment period
		Treatment period	
Role functioning	6, 7	Everyday life during cancer	
Emotional functioning	21-24	Diagnosis	Diagnosis
		Breast cancer as a daily battle	Breast cancer as a daily battle
		Waiting for results of analysis, concerns	Anxiety and fatigue
		Support from patient's family and friends	Support from patient's family and friends
Cognitive functioning	20, 25	Search for medical information	
		Media and forum information exchange	
Social functioning	26, 27	Support from patient's family and friends	Support from patient's family and friends
		Work life during cancer and financial aspects	Work life during cancer and financial aspects
Symptom scales			
Fatigue	10, 12, 18	Chemotherapy and its secondary effects	Anxiety and fatigue
			Secondary effects of treatments
Nausea and vomiting	14, 15	Chemotherapy and its secondary effects	Secondary effects of treatments
Pain	9, 19	Chemotherapy and its secondary effects	Secondary effects of treatments
		Surgery	
Dyspnea	8	Chemotherapy and its secondary effects	Secondary effects of treatments
Insomnia	11	Chemotherapy and its secondary effects	Secondary effects of treatments
Appetite loss	13	Chemotherapy and its secondary effects	Secondary effects of treatments
Constipation	16	Chemotherapy and its secondary effects	Secondary effects of treatments
Diarrhea	17	Chemotherapy and its secondary effects	Secondary effects of treatments
Financial difficulties	28	Work life during cancer and financial aspects	Work life during cancer and financial aspects
Global health status			
Global health status and quality of life	29, 30		
EORTC QLQ-BR23[b]			
Functional scales			
Body image	39-42	Breast reconstruction	Breast reconstruction
		Body care and body image during cancer, and sexuality	Body care and body image during cancer
		Surgery	
Sexual functioning	44, 45	Body care and body image during cancer, and sexuality	Sexuality
Sexual enjoyment	46	Body care and body image during cancer, and sexuality	Sexuality
Future perspectives	43	Healing	Healing
			Relapse

Questionnaires and their scales		Questionnaire items	cancerdusein.org	Facebook
Symptom scales				
	Systemic therapy	31-34	Chemotherapy and its secondary effects	Secondary effects of treatments
	Side effects	36-38	Hormone therapy and its secondary effects	Chemotherapy and its secondary effects
	Breast symptoms	50-53	Breast reconstruction	Breast reconstruction
			Radiotherapy and its secondary effects	
			Surgery	
	Arm symptoms	47-49	Breast reconstruction	Breast reconstruction
			Surgery	
	Hair loss	35	Hair loss	Hair loss
Topics without a relationship				
			Complementary and alternative medicine	
			Mourning	
				Family background and breast cancer
			Family members with breast cancer	Family members with breast cancer
				Healing of family member

[a]EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.

[b]QLQ-BR23: breast cancer module.

Table 6. Distribution of documents in each topic on cancerdusein.org and Facebook.

Topic no.	cancerdusein.org (n=16,868) n (%)	Facebook (n=70,092) n (%)
1	978 (5.80)	3294 (4.70)
2	590 (3.50)	3925 (5.60)
3	1147 (6.80)	3785 (5.40)
4	860 (5.10)	4065 (5.80)
5	1315 (7.80)	2804 (4.00)
6	759 (4.50)	3715 (5.30)
7	810 (4.80)	3014 (4.30)
8	523 (3.10)	3084 (4.40)
9	877 (5.20)	3645 (5.20)
10	692 (4.10)	3505 (5.00)
11	675 (4.00)	2804 (4.00)
12	523 (3.10)	2734 (3.90)
13	1113 (6.60)	5047 (7.20)
14	692 (4.10)	3014 (4.30)
15	843 (5.00)	2804 (4.00)
16	1063 (6.30)	2734 (3.90)
17	1248 (7.40)	3575 (5.10)
18	540 (3.20)	5607 (8.00)
19	1198 (7.10)	3432 (4.90)
20	422 (2.50)	3505 (5.00)

XSL•FO

RenderX

Relationships Between Self-Administered Questionnaires and Social Media

We were able to match most of the topics from QoL self-administered questionnaires in social media. These topics correspond to a total of 95% (22/23) of topics in the cancerdusein.org corpus and 86% (20/23) of topics in the Facebook corpus. These figures underline the importance of studying QoL, because they correspond to patients' real concerns. The topics that corresponded with those of the EORTC QLQ-C30 and the EORTC QLQ-BR23 questionnaires were hair loss, work life during cancer and financial aspects, chemotherapy and its secondary effects, breast reconstruction, support from the patient's family and friends, treatment period, healing, diagnosis, breast cancer as a daily battle, body care and body image during cancer and sexuality, hormone therapy and its secondary effects, radiotherapy and its secondary effects, media and forum information exchange, everyday life during cancer, search for medical information, surgery, waiting for results of analysis, concerns, secondary effects of treatments, interaction with nurses and doctors, anxiety and fatigue, and relapse.

Emerging Topics in Social Media

We also found 5 topics that are not present in QoL questionnaires. These topics correspond to a total of 15% (3/20) of the cancerdusein.org corpus and 15% (3/20) of the Facebook corpus. Of the 5 topics that do not appear in the questionnaires, 2 focus on patients. The emerging topics are complementary and alternative medicine, mourning, family background and breast cancer, family members with breast cancer, and healing of a family member. Among these 5 topics, we believe that 2 of them (complementary and alternative medicine, and family background and breast cancer) could be added to the QoL questionnaires. The topic complementary and alternative medicine focuses on nonconventional treatments and corresponded to a total of 3.10% (523/16,868) of the cancerdusein.org corpus. The topic family background and breast cancer focuses on the relationships of patients with their family, especially healing and grieving for a family member. This topic corresponded to a total of 4.30% (3014/70,092) of the Facebook corpus. The 3 others topics are not related to QoL. These topics deal with mourning, having family members with breast cancer, and healing of a family member. They were discussed by relatives of patients and not by patients.

Different Uses of Forums and Social Networks

One of the reasons that led us to use 2 data resources (social networks and a health forum) was to discover the topics discussed in each platform. Table 7 presents the relationships between topics found in both social media and the percentage distribution of messages in each topic. Of 20 topics detected by our model in the corpus forum and Facebook, we found 11 common topics in the 2 corpora. Some of them have a similar frequency of discussion (Table 6). These topics are hair loss, work life during cancer, support from patient's family and friends, treatment period, diagnosis, and family members with breast cancer. We observed that topics such as chemotherapy and its secondary effects, breast reconstruction, and breast cancer as daily battle were discussed more on the forum than on Facebook, maybe because the subject is more technical. As Table 7 shows, we noted that the topics support from a patient's family and friends, body care and body image during cancer, and sexuality were discussed more on Facebook than on the forum because of visibility to friends. In the end, the topics discovered were quite similar. However, we observed a difference of length in the posts. Most of the time, posts from the health forum were longer than posts from Facebook. Even if the topics found in both social media were similar, messages from the forum provided more information and were better interpreted than messages from Facebook.

Table 7. Relationships between topics found on both social media (cancerdusein.org and Facebook) with K=20 in collaboration with an expert.

Topic names	cancerdusein.org (n=16,868)		Facebook (n=70,092)		Matched to questionnaire item
	Topic no.	n (%)	Topic no.	n (%)	
Topics on both social media					
Hair loss	1	978 (5.80)	4	4065 (5.80)	Yes
Work life during cancer and financial aspects	2	590 (3.50)	8	3084 (4.40)	Yes
Chemotherapy and its secondary effects	3	1147 (6.80)	2	3925 (5.60)	Yes
Breast reconstruction	5	1315 (7.80)	9	3645 (5.20)	Yes
Support from patient's family and friends	6	759 (4.50)	10	3505 (5.00)	Yes
			19	3432 (4.90)	Yes
Family members with breast cancer	10	692 (4.10)	17	3575 (5.10)	No
Treatment period	11	675 (4.00)	20	3505 (5.00)	Yes
Healing	13	1113 (6.60)	18	5607 (8.00)	Yes
Diagnosis	16	1063 (6.30)	1	3294 (4.70)	Yes
Breast cancer as a daily battle	17	1248 (7.40)	3	3785 (5.40)	Yes
Body care and body image during cancer, and sexuality	18	540 (3.20)	6	3715 (5.30)	Yes
			15	2804 (4.00)	Yes
			16	2734 (3.90)	Yes
Topics on only 1 social media					
Hormone therapy and its secondary effects	4	860 (5.10)	N/A[a]	N/A	Yes
Radiotherapy and its secondary effects	7	810 (4.80)	N/A	N/A	Yes
Complementary and alternative medicine	8	523 (3.10)	N/A	N/A	No
Media and forum information exchange	9	877 (5.20)	N/A	N/A	Yes
Everyday life during cancer	12	523 (3.10)	N/A	N/A	Yes
Search for medical information	14	692 (4.10)	N/A	N/A	Yes
Mourning	15	843 (5.00)	N/A	N/A	No
Surgery	19	1198 (7.10)	N/A	N/A	Yes
Waiting for results of analysis, concerns	20	422 (2.50)	N/A	N/A	Yes
Secondary effects of treatments	N/A	N/A	5	2804 (4.00)	Yes
Family background and breast cancer	N/A	N/A	7	3014 (4.30)	No
Interaction with nurses and doctors	N/A	N/A	11	2804 (4.00)	Yes
Anxiety and fatigue	N/A	N/A	12	2734 (3.90)	Yes
Healing of member family	N/A	N/A	13	5047 (7.20)	No
Relapse	N/A	N/A	14	3014 (4.30)	Yes

[a]N/A: not applicable.

Conclusions

In this work, we used an unsupervised learning model known as LDA to detect the different topics on a health forum and social network discussed by patients. We demonstrated how we used the LDA model on patient data with relevant preprocessing applied to 2 datasets obtained from a forum and Facebook messages. We used MeSH as the principal resource for medical terms and for patients' and doctors' vocabulary [45]. We automatically detected relationships between topics and questions. We found good relationships between detected topics and the dimensions of internationally standardized questionnaires used for breast cancer patients, which substantiate the sound construction of such questionnaires. We detected new emerging topics from social media that could be used to complete actual QoL questionnaires. Moreover, we confirmed that social media can be an important source of information for the study of QoL in the field of cancer.

In our ongoing work [21], we are targeting the classification of whole messages or text snippets with respect to the role of the

narrator (patient, confidant of a patient, expert, health professional) and to the location within the trajectory of care (before or after an operation, first cancer or relapse). One potential limitation of this work was the number of topics (K=20) selected for our LDA model. This limitation may be overcome by using the number of topics for which the model is better adjusted [47,61,62], then, first, to merge topics that are close, and second, to find topics that could not be interpreted by humans and eliminate them. Moreover, the actual comparison of the 2 corpora (Facebook and forum) was done manually by the expert. A possibility is to adapt equation 1 (Figure 3) used to align LDA topics and questionnaire items in order to automatically compare topics extracted from the 2 corpora.

Of course, the lack of informed consent given by social media users for data usage leads to ethical questions. In particular, confidentiality with respect to the publication of research results is an issue (see others' discussion and guidelines [66-68]). We adhered to those guidelines. We have presented results with a degree of detail that does not permit conclusions on individual users to be drawn. In the long term, we will study emotions described by patients in their messages for each topic and make some statistical analyses. Finally, we will use the emotion classification system built by Abdaoui et al [69] to detect polarity (positive, negative, or neutral), subjectivity (objective, subjective), and feelings (joy, surprise, anger, fear, etc) of users' messages, and we will relate this information to the detected topics in order to determine patients' perception of their disease. What are the topics that frighten patients the most and that need prevention?

Conflicts of Interest

None declared.

References

1. Robinson KM. Unsolicited narratives from the Internet: a rich source of qualitative data. Qual Health Res 2001 Sep;11(5):706-714. [doi: 10.1177/104973201129119398] [Medline: 11554197]
2. Seale C, Charteris-Black J, MacFarlane A, McPherson A. Interviews and internet forums: a comparison of two sources of qualitative data. Qual Health Res 2010 May;20(5):595-606. [doi: 10.1177/1049732309354094] [Medline: 20008955]
3. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
4. Hancock JT, Toma C, Ellison N. The truth about lying in online dating profiles. 2007 Presented at: The SIGCHI conference on Human factors in computing systems of the ACM; Apr 28–May 3, 2007; San Jose, CA, USA p. 449-452.
5. Lemire M, Paré G, Sicotte C, Harvey C. Determinants of Internet use as a preferred source of information on personal health. Int J Med Inform 2008 Nov;77(11):723-734. [doi: 10.1016/j.ijmedinf.2008.03.002] [Medline: 18434246]
6. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
7. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
8. Ganz PA, Rowland JH, Desmond K, Meyerowitz BE, Wyatt GE. Life after breast cancer: understanding women's health-related quality of life and sexual functioning. J Clin Oncol 1998 Feb;16(2):501-514. [doi: 10.1200/JCO.1998.16.2.501] [Medline: 9469334]
9. King MT, Kenny P, Shiell A, Hall J, Boyages J. Quality of life three months and one year after first treatment for early stage breast cancer: influence of treatment and patient characteristics. Qual Life Res 2000;9(7):789-800. [Medline: 11297021]
10. Lidgren M, Wilking N, Jönsson B, Rehnberg C. Health related quality of life in different states of breast cancer. Qual Life Res 2007 Aug;16(6):1073-1081. [doi: 10.1007/s11136-007-9202-8] [Medline: 17468943]
11. Montazeri A. Health-related quality of life in breast cancer patients: a bibliographic review of the literature from 1974 to 2007. J Exp Clin Cancer Res 2008;27(1):32. [doi: 10.1186/1756-9966-27-32] [Medline: 18759983]
12. Doward L, McKenna S. Defining patient-reported outcomes. Value Health 2004;7 Suppl 1:S4-S8 [FREE Full text] [doi: 10.1111/j.1524-4733.2004.7s102.x] [Medline: 15367236]
13. Fayers P, Machin D. Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes. 2nd edition. New York, NY: John Wiley & Sons; 2013.

XSL•FO

RenderX

14. Bausewein C, Hartenstein R. Oncology and palliative care. Oncol Res Treat 2001 Jan 15;23(6):534-537. [doi: 10.1159/000055002] [Medline: 11441257]

15. Bausewein C, Higginson IJ. Appropriate methods to assess the effectiveness and efficacy of treatments or interventions to control cancer pain. J Palliat Med 2004 Jun;7(3):423-430. [doi: 10.1089/1096621041349572] [Medline: 15265352]

16. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: in search of a standard. Med Decis Making 2000;20(3):332-342. [doi: 10.1177/0272989X0002000310] [Medline: 10929856]

17. Cutler D, McClellan M. Is technological change in medicine worth it? Health Aff (Millwood) 2001;20(5):11-29 [FREE Full text] [Medline: 11558696]

18. Hillner B, Smith TJ. Efficacy and cost effectiveness of adjuvant chemotherapy in women with node-negative breast cancer. A decision-analysis model. N Engl J Med 1991 Jan 17;324(3):160-168. [doi: 10.1056/NEJM199101173240305] [Medline: 1898533]

19. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 2002 Jun 15;324(7351):1417-1421 [FREE Full text] [Medline: 12065262]

20. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 3;85(5):365-376. [Medline: 8433390]

21. Opitz T, Bringay S, Azé J, Joutard C, Lavergne C, Mollevi C. Breast cancer and quality of life: medical information extraction from health forums. 2014 Presented at: Medical Informatics Europe; Aug 31–Sept 3, 2014; Istanbul, Turkey p. 1070-1074.

22. Arnold C, Speier W. A topic model of clinical reports. 2012 Presented at: The 35th international ACM SIGIR conference on Research development in information retrieval; Aug 12-16, 2012; Portland, OR, USA p. 12-16.

23. Lu Y, Zhang P, Liu J, Li J, Deng S. Health-related hot topic detection in online communities using text clustering. PLoS One 2013 Jan;8(2):e56221 [FREE Full text] [doi: 10.1371/journal.pone.0056221] [Medline: 23457530]

24. Zhang S, Grave E, Sklar E, Elhadad N. Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks. 2016 Mar. URL: http://adsabs.harvard.edu/abs/2016arXiv160308458Z [accessed 2017-07-18] [WebCite Cache ID 6s3SA3w5B]

25. Landauer TK, Dumais ST. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol Rev 1997;104(2):211-240. [doi: 10.1037/0033-295X.104.2.211]

26. Hofmann T. Unsupervised learning by probabilistic latent semantic analysis. Mach Learn 2001;42(1):177-196. [doi: 10.1023/A:1007617005950]

27. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022. [doi: 10.1162/jmlr.2003.3.4-5.993]

28. Deerwester S, Dumais S, Furnas G, Landauer T, Harshman R. Indexing by latent semantic analysis. J Am Soc Inf Sci 1990 Sep;41(6):391-407. [doi: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9]

29. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. J Med Internet Res 2017 Jan 20;19(1):e24 [FREE Full text] [doi: 10.2196/jmir.5780] [Medline: 28108428]

30. Wang H, Ding Y, Tang J, Dong X, He B, Qiu J, et al. Finding complex biological relationships in recent PubMed articles using Bio-LDA. PLoS One 2011 Mar 23;6(3):e17243 [FREE Full text] [doi: 10.1371/journal.pone.0017243] [Medline: 21448266]

31. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]

32. Portier K, Greer GE, Rokach L, Ofek N, Wang Y, Biyani P, et al. Understanding topics and sentiment in an online cancer survivor community. J Natl Cancer Inst Monogr 2013 Dec;2013(47):195-198. [doi: 10.1093/jncimonographs/lgt025] [Medline: 24395991]

33. Attard A, Coulson NS. A thematic analysis of patient communication in Parkinson's disease online support group discussion forums. Comput Hum Behav 2012 Mar;28(2):500-506. [doi: 10.1016/j.chb.2011.10.022]

34. Selby P, van MT, Voci SC, Parent D, Cunningham JA. Online social and professional support for smokers trying to quit: an exploration of first time posts from 2562 members. J Med Internet Res 2010;12(3):e34 [FREE Full text] [doi: 10.2196/jmir.1340] [Medline: 20719739]

35. Himmel W, Reincke U, Michelmann HW. Text mining and natural language processing approaches for automatic categorization of lay requests to web-based expert forums. J Med Internet Res 2009 Jul 22;11(3):e25 [FREE Full text] [doi: 10.2196/jmir.1123] [Medline: 19632978]

36. Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. J Biomed Inform 2013 Dec;46(6):998-1005 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.011] [Medline: 24025513]

37. Hao H, Zhang K. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]

38.   Hao H, Zhang K, Wang W, Gao G. A tale of two countries: international comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Mar;99:37-44. [doi: 10.1016/j.ijmedinf.2016.12.007] [Medline: 28118920]

39.   Yesha R, Gangopadhyay A. A method for analyzing health behavior in online forums. 2015 Presented at: The 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics; Sept 9-12, 2015; Atlanta, GA, USA p. 615-621.

40.   Wang YC, Kraut R, Levine JM. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. 2012 Presented at: Conference on Computer Supported Cooperative Work; Feb 11-15, 2012; Seattle, WA, USA p. 833-842.

41.   Wang S, Paul MJ, Dredze M. Exploring health topics in Chinese social media: an analysis of Sina Weibo. 2014 Presented at: The AAAI Workshop on the World Wide Web and Public Health Intelligence; July 27-28 , 2014; Quebec, QC, Canada p. 20-23.

42.   Balahur A. Sentiment analysis in social media texts. 2013 Presented at: The 4th workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; June 14, 2013; Atlanta, GA, USA p. 120-128.

43.   Farzindar A, Inkpen D. Natural language processing for social media. Synthesis Lectures on Human Language Technologies 2015 Aug 28;8(2):1-166. [doi: 10.2200/S00659ED1V01Y201508HLT030]

44.   Schmid H. Probabilistic part-of-speech tagging using decision trees. 1994 Presented at: International Conference on New Methods in Language Processing; Sept 14-16, 1994; Manchester, UK p. 14-16.

45.   Tapi Nzali MD, Bringay S, Lavergne C, Opitz T, Azé J, Mollevi C. Construction d'un vocabulaire patient/médecin dédié au cancer du sein à partir des médias sociaux. 2015 Presented at: The 26th Conference on Ingénierie des Connaissances; July 1-3, 2015; Rennes, France p. 9-20.

46.   Asuncion A, Welling M, Smyth P, Teh YW. On smoothing and inference for topic models. 2009 Presented at: The 25th Conference on Uncertainty in Artificial Intelligence; June 18-21, 2009; Montreal, QC, Canada p. 27-34.

47.   Griffiths TL, Steyvers M. Finding scientific topics. Proc Natl Acad Sci U S A 2004 Apr 06;101 Suppl 1:5228-5235 [FREE Full text] [doi: 10.1073/pnas.0307752101] [Medline: 14872004]

48.   Chang J, Gerrish S, Wang C, Boyd-Graber JL, Blei DM. Reading tea leaves: how humans interpret topic models. 2009 Presented at: The 22nd International Conference on Neural Information Processing Systems; Dec 7-10, 2009; Vancouver, BC, Canada p. 288-296.

49.   Wallach HM, Murray I, Salakhutdinov R, Mimno D. Evaluation methods for topic models. 2009 Presented at: Annual International Conference on Machine Learning ACM; June 14-18, 2009; Montreal, QC, Canada p. 1105-1112.

50.   Thirion B, Pereira S, Névéol A, Dahamna B, Darmoni S. French MeSH Browser: a cross-language tool to access MEDLINE/PubMed. 2006 Presented at: Annual Symposium Proceedings/AMIA Symposium; Nov 11-15, 2006; Washington, DC, USA p. 1132.

51.   Jaccard P. Etude Comparative de la Distribution Florale dans une Portion des Alpes et du Jura. Zurich, Switzerland: Imprimerie Corbaz; 1901.

52.   Grün B, Hornik K. topicmodels: an R package for fitting topic models. J Stat Softw 2011;40(13). [doi: 10.18637/jss.v040.i13]

53.   Anota A, Barbieri A, Savina M, Pam A, Gourgou-Bourgade S, Bonnetain F, et al. Comparison of three longitudinal analysis models for the health-related quality of life in oncology: a simulation study. Health Qual Life Outcomes 2014 Dec 31;12:192 [FREE Full text] [doi: 10.1186/s12955-014-0192-2] [Medline: 25551580]

54.   Barbieri A, Anota A, Conroy T, Gourgou-Bourgade S, Juzyna B, Bonnetain F, et al. Applying the longitudinal model from item response theory to assess health-related quality of life in the PRODIGE 4/ACCORD 11 randomized trial. Med Decis Making 2016 Jul;36(5):615-628. [doi: 10.1177/0272989X15621883] [Medline: 26683246]

55.   Sprangers MA, Groenvold M, Arraras JI, Franklin J, te Velde A, Muller M, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. J Clin Oncol 1996 Oct;14(10):2756-2768. [doi: 10.1200/JCO.1996.14.10.2756] [Medline: 8874337]

56.   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]

57.   Abdaoui A, Azé J, Bringay S, Grabar N, Poncelet P. Analysis of forum posts written by patients and health professionals. 2014 Presented at: Medical Informatics Europe; Aug 31–Sept 3, 2014; Istanbul, Turkey p. 1185.

58.   Taphoorn MJB, Claassens L, Aaronson NK, Coens C, Mauer M, Osoba D, EORTC Quality of Life Group, Brain Cancer, NCIC and Radiotherapy Groups. An international validation study of the EORTC brain cancer module (EORTC QLQ-BN20) for assessing health-related quality of life and symptoms in brain cancer patients. Eur J Cancer 2010 Apr;46(6):1033-1040. [doi: 10.1016/j.ejca.2010.01.012] [Medline: 20181476]

59.   Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. Eur J Cancer 1994;30A(5):635-642. [Medline: 8080679]

60.   Abboute A, Boudjeriou Y, Entringer G, Azé J, Bringay S, Poncelet P. Mining Twitter for suicide prevention. 2014 Presented at: International Conference on Applications of Natural Language to Data Bases/Information Systems; June 18–20, 2014; Montpellier, France p. 250-253.

XSL•FO

RenderX

61. Arun R, Suresh V, Madhavan C, Murthy M. On finding the natural number of topics with latent Dirichlet allocation: some observations. 2010 Presented at: Pacific-Asia Conference on Knowledge Discovery and Data Mining; June 21–24, 2010; Hyderabad, India p. 391-402.

62. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. Neurocomputing 2009 Mar;72(7):1775-1781. [doi: 10.1016/j.neucom.2008.06.011]

63. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinformatics 2015;16 Suppl 13:S8 [FREE Full text] [doi: 10.1186/1471-2105-16-S13-S8] [Medline: 26424364]

64. Blei D, Lafferty J. Correlated topic models. 2005 Presented at: The 18th International Conference on Neural Information Processing Systems; Dec 5-8, 2005; Vancouver, BC, Canada p. 147-154.

65. Arnold CW, Oh A, Chen S, Speier W. Evaluating topic model interpretability from a primary care physician perspective. Comput Methods Programs Biomed 2016 Feb;124:67-75 [FREE Full text] [doi: 10.1016/j.cmpb.2015.10.014] [Medline: 26614020]

66. KING SA. Researching Internet communities: proposed ethical guidelines for the reporting of results. Inf Soc 1996 Jun;12(2):119-128. [doi: 10.1080/713856145]

67. Frankel M, Siang S. Ethical and legal aspects of human subjects research on the Internet. 1999. URL: https://nationalethicscenter.org/resources/187/download/ethical_legal.pdf [accessed 2017-07-19] [WebCite Cache ID 6s4oEE0mK]

68. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. Am Psychol 2004;59(2):105-117. [doi: 10.1037/0003-066X.59.2.105] [Medline: 14992637]

69. Abdaoui A, Tapi Nzali MD, Azé J, Bringay S, Lavergne C, Mollevi C, et al. Analyse du sentiment, de l'opinion et de l'émotion sur des Tweets Français. 2015 Presented at: The 11th Workshop on Défi Fouille de Texte, Traitement Automatique des Langues Naturelles; June 22, 2015; Caen, France p. 78-87.

Abbreviations

EORTC: European Organization for Research and Treatment of Cancer
INCA: Institut National du Cancer
LDA: latent Dirichlet allocation
MeSH: Medical Subject Headings
QLQ-C30: Quality of Life Questionnaire Core 30
QoL: quality of life

Viewpoint

Hierarchical Medical System Based on Big Data and Mobile Internet: A New Strategic Choice in Health Care

Yaogang Wang[1*], MD, PhD; Li Sun[1*], MD; Jie Hou[1*], MD

Tianjin Medical University, Tianjin, China

[*]all authors contributed equally

Corresponding Author:
Yaogang Wang, MD, PhD
Tianjin Medical University
22 Qixiangtai Road
Tianjin,
China
Phone: 86 02283336636
Fax: 86 02283336636
Email: wyg@tmu.edu.cn

Abstract

China is setting up a hierarchical medical system to solve the problems of biased resource allocation and high patient flows to large hospitals. The development of big data and mobile Internet technology provides a new perspective for the establishment of hierarchical medical system. This viewpoint discusses the challenges with the hierarchical medical system in China and how big data and mobile Internet can be used to mitigate these challenges.

Introduction

The hierarchical medical system has become an essential system in many developed countries. It plays an important role of forming the basis for guaranteeing health care [1-4]. The basic working of a hierarchical medical system involves initial diagnoses at primary medical institutions and two-way referrals among hospitals. In China, which has a population of over 1.37 billion [5], there are many problems in the medical system, such as biased resource allocation and extremely high patient flows to large hospitals [6]. Since 2009, China has vigorously promoted the implementation of the hierarchical medical system to realize rational allocation of medical resources, promote the equalization of primary medical services, and reduce the cost of medical services. On the basis of this background, this paper aims to analyze the difficulties of hierarchical medical system in China and opens a dialogue on the challenges associated with the innovation model of the hierarchical medical system by using perspectives based on big data and mobile Internet.

Difficulties of Hierarchical Medical System in China

So far, China has not established an effective model of the hierarchical medical system. In China, hospitals at different levels formed a regional medical consortium (RMC). The government is urging hospitals at different levels in an RMC to strengthen their cooperation and recognize each others' patient medical results, while encouraging two-way referrals among them. However, an RMC cannot achieve the results that can be expected from a hierarchical medical system. From 2005 to 2014, the number of hospitals in China increased by an average of 716 per year, whereas the primary medical institutions increased by an average of 6785 per year. By contrast, the average annual growth rate of outpatients in hospitals and primary medical institutions were 11.43% and 6.82%, respectively (Figure 1) [7]. The growth rate of outpatients in primary medical institutions has not matched the growth rate of outpatients in institutions. Additionally, the number of beds and the rate of bed utilization increased more in hospitals than in primary medical institutions. Large hospitals are still overcrowded, while primary medical institutions are sparsely populated. High-quality medical resources are concentrated in large hospitals, but primary medical institutions are seriously lacking in medical resources. In addition, the health-information-sharing platforms and associated mechanisms have not been established. Patients' information cannot be shared among hospitals at different levels so patients cannot enjoy the continuity of medical services between different

hospitals. These problems increase the difficulty and cost of medical services in China.

Therefore, the biggest challenge for China is to find a successful way to solve these problems.

Challenge and Opportunities

In the 21st century, traditional health care has been rapidly changing owing to Internet-based big data and cloud computing [8,9]. Big data is being generated by all digital operations at all times during routine use. Every digital process and social media exchange produces data through systems, sensors, and mobile devices that transmit this data. Therefore, big data is generated by multiple sources with an alarming velocity, volume, and variety. Four continuous stages including the generation, acquisition, storage, and analysis constitute the big data value-chain.

Pervasive Internet access has enabled patients around the world to seek information on the best care available. Additionally, the Internet facilitates efficient communication of medical information globally [10]. Personalized health technologies are also developing rapidly. Sensors, smart-watches, and mobile health apps are strapped to wrists and placed in pockets to monitor and help in modifying health behaviors [11]. Some high-profile devices include Fitbit, Jawbone, Microsoft Band, and Apple smart watch. These technologies help with self-quantification of physical activity and health self-management [12].

In China, "Mobile Internet + Medical Care" oriented by the big data value chain is gradually changing medical practices and processes. The main functions of mobile medical care include making appointments, consulting service, acquiring health information, and providing guidance. By the end of 2016, the number of mobile phone users in China reached over 1.32 billion, while 4G Mobile Internet users accounted for 58.2%(58.2/100) of the total [13]. Meanwhile, the construction of hospital information is developing quickly as the usage of electronic medical records, digital hospitals, telemedicine, and collaborative medical care is on the rise. The development of big data and mobile Internet technology provides a new perspective for the establishment of a hierarchical medical system.

Figure 1. Changes in the numbers of medical institutions, outpatients, beds and bed utilization between hospitals and primary medical institutions from 2005 to 2014 in China.
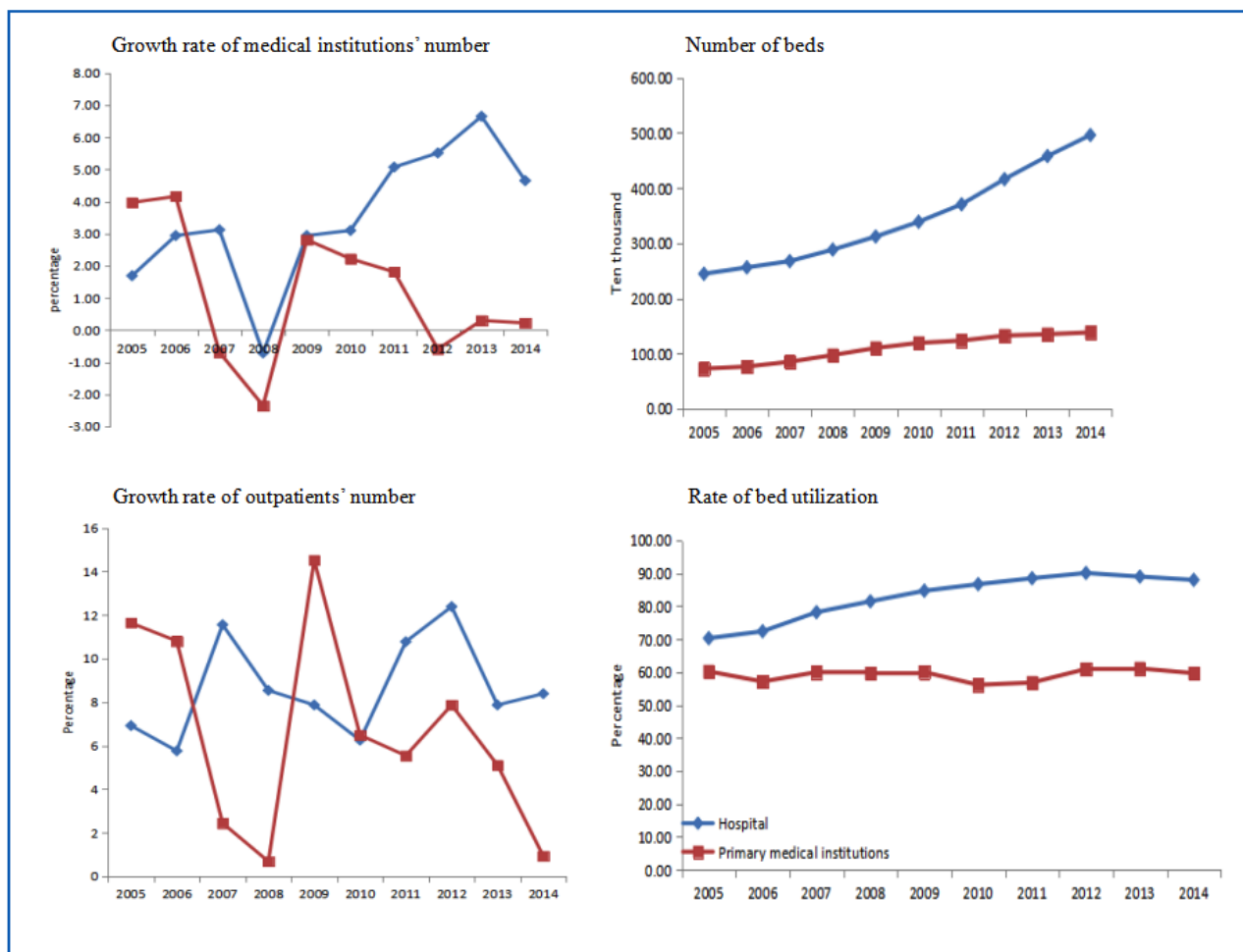
Hierarchical Medical System Based on Big Data and Mobile Internet

This paper proposes an innovative model of the hierarchical medical system based on big data and mobile Internet that may become a new strategic choice to resolve the imbalance in the availability of medical resources in China.

This innovative model aims to materialize 5 successive medical services. These services would involve linkages at different levels such as institutional, interdisciplinary, interpersonal, patient satisfaction, and management. These linkages have no interruptions by using the Data Sharing and Processing System (DSPS) (Figure 2).

In this model, medical data can be recorded and shared among hospitals and health facilities at different levels across the DSPS. The important basic patient information including the physical examination file, medical records, laboratory results, imaging results, medication records, self-monitoring, and other relevant information can be transmitted to medical service providers in different institutions, which would result in patients being able to receive continuous medical services when they go to different institutions to receive care. In addition, advanced medical resources in large hospitals can also accessed by the primary medical institutions without the constraints of time and space. Doctors at more advanced hospitals can participate in the medical activities of lower level medical institutions online and help medical professionals in primary medical institutions to improve the quality of their services. Owing to this connectivity and cooperation, service providers at different facilities can become familiar with one another and establish long-term relationships. In addition, patients would have the benefit of receiving continuous medical service at different institutions seamlessly (Figure 3).

Figure 2. The theoretical model of hierarchical medical system based on big data and mobile Internet.
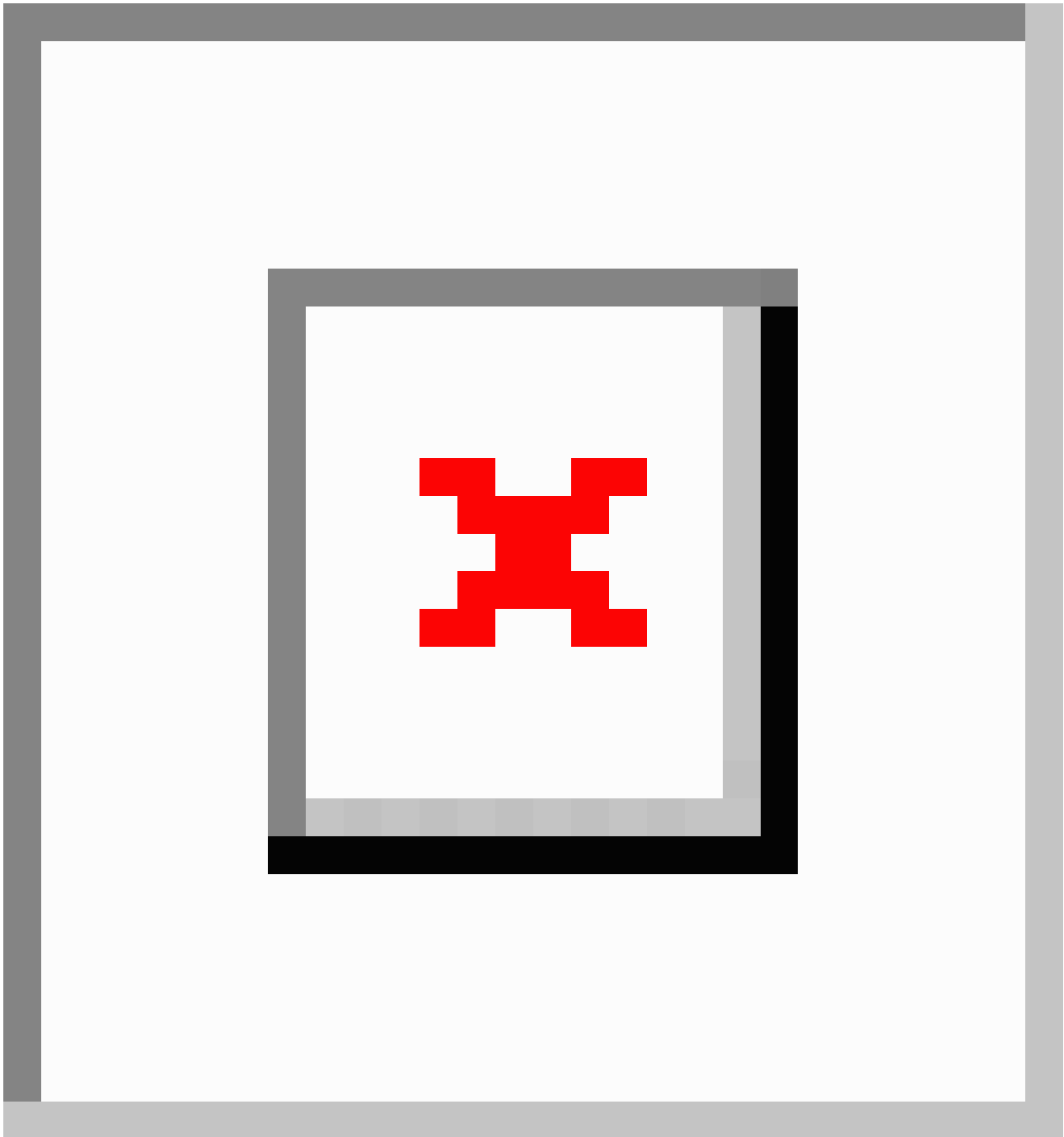
Figure 3. Functions of hierarchical medical system based on big data and mobile Internet.
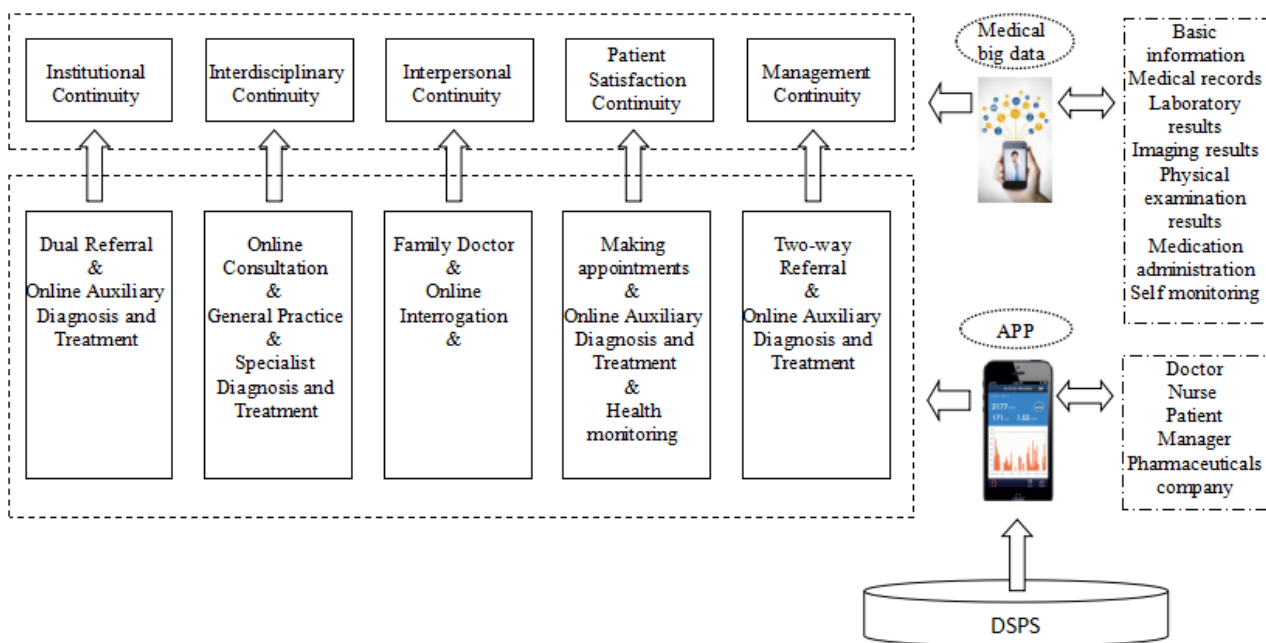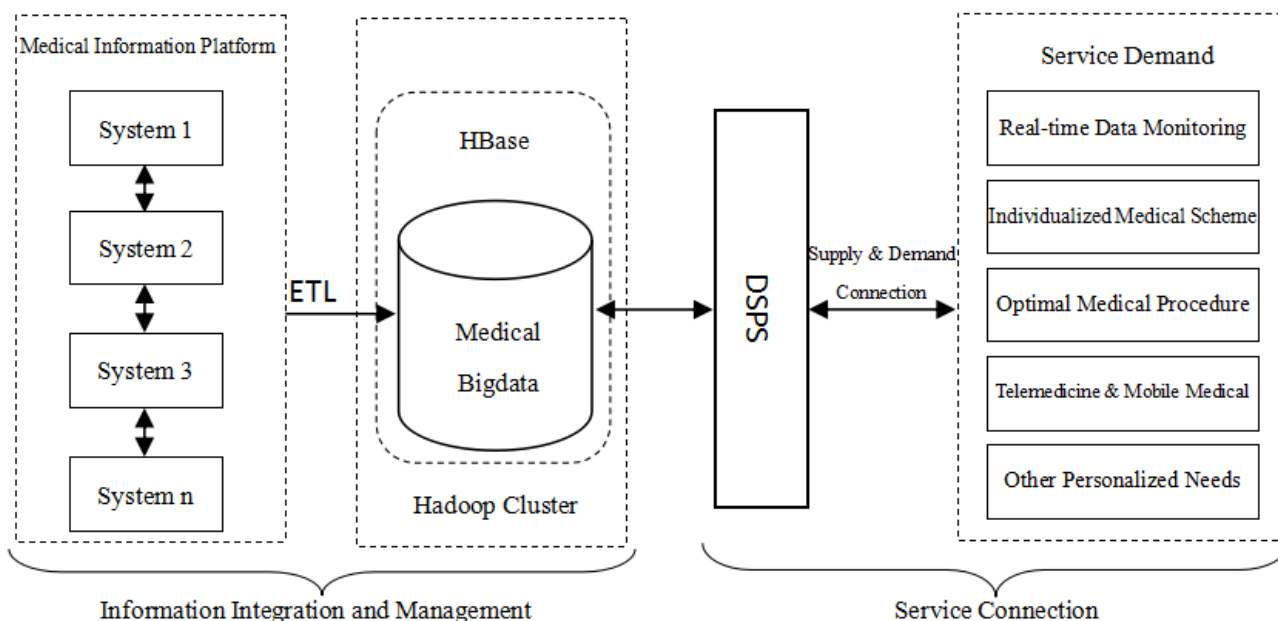

Figure 4. The data sharing and processing system architecture.

Integrated and shared medical data provide the technical support to realize these services. As previously stated, this innovative model could be achieved by using the DSPS (Figure 4). This system is based on mobile Internet technology that is driven by the big data value chain, and it connects the patients with different medical service providers, including doctors at the hospitals, family doctors, pharmaceutical companies, and other relevant participants in the medical system. The advantage of this connection is making information sharing possible by breaking the constraints of time and space.

First, based on Hadoop software, it would be practical to construct the hierarchical medical data-sharing platform that can realize medical big data integration and sharing. The integration scheme based on Hadoop platform consists of three parts. The first part consists of the medical institutions and the clinical information systems (EMR, PACS, LIS, etc), and the message engine would be based on Apache Camel. The second part utilizes the enterprise information portal (EIP) to design the routing rules and process flow. The third part is the Hadoop platform that is composed of the Hadoop cluster, Zookeeper, and HBase cluster.

Second, it is possible to explore the evolution of structures, agent behavior, and interactive form of subsystems in this model by using the system dynamics methods.

Third, it is feasible to develop the mobile application system, which includes the modules of authentication, appointment registration, electronic medical records, prescriptions, and test results query.

These 3 steps provide the necessary theoretical basis, technical support, and ways to develop this innovative model.

This huge data platform will take a single sign-on mode. Users need to log in, and based on the confidentiality level of the dataset, the users will have different levels of authority for access. Single sign-on unified identity authentication and authority control technology and strict control of user access are some of the ways to effectively ensure the safety of such large data applications. The system is based on the Hadoop platform. The cluster consists of a master node and a slave node. The master node needs to install and configure Hadoop, Hive, Sqoop, and MySQL in order to manage the cluster. Slave node only needs to install Hadoop; it can also be configured similar to the master node.

To ensure data privacy, the protection model mainly includes two aspects. One is the user-querying privacy protection; the query content cannot be leaked out. The second is data privacy. This protection model is based on the key. The data is protected by the form of anonymous processing based on the third party and query split. In this way, the data platform cannot connect the privacy content query with the users, thus protecting the user-querying privacy.

The application IoT technology in medicine covers almost all aspects, including medical practice, remote monitoring and home care, medical information, hospital first aid, medical equipment and medical waste monitoring, blood bank management, and infection control. An important application of IoT in medical information technology is mobile medical services, which are based on wireless LAN (local-area network) technology and RFID (radio-frequency identification).

In the DSPS, the Internet of Things (IoT) technology can be used to materialize the intelligent and real-time management of patients and the related systems such as digital collection, processing, storage, transmission, and sharing of information. The intelligent character recognition (ICR) technology is used at the hospitals to build the main index of patients and drugs. Using bar code scanning and RFID technology, hospitals have

an accurate information confirmation and recognition system. This technology allows doctors to receive abundant data using the mobile sensor device and the medical instrument, which are both suitable for household use. IoT technology can support the collection of all kinds of vital signs data whenever and wherever possible. It can then automatically transfer the data to hospitals. Data mining and machine learning can be used to discover the hidden knowledge from these data. Interconnection technology can integrate patients and processes to allow for the standardized management of processes.

The main data collected by using IoT may include basic patient information (eg, full name, gender, date of birth, phone number, social security card number or ID number, and photo), medical information (eg, blood type, disease status such as diabetes, epilepsy, and hypertension), patient registration and treatment number, and medical treatment information (ie, medical records). It can also include the doctor's electronic clinical orders (combined with mobile operation), patient medication treatment records (combined with medicare electronic settlement and payment), and patient tracking and positioning. Medical staff can access all these medical records both via portable hand held devices and desktop computers.

Telemedicine services can alleviate the queuing problem and reduce the cost of transportation for patients. It allows health care professionals to efficiently monitor patients' indicators and give suggestions at any time. It is a low cost, rapid, and stable health monitoring method that can also extend medical services to subhealthy people, elderly patients, and patients with chronic diseases.

Perspective

Surely, we need more research and dialogue on these issues and the impact associated with this innovative model of hierarchical medical system based on big data and mobile Internet. This model is likely to provide a new perspective and strategic choice of health care service, not only for China, but also for other countries.

Conflicts of Interest

None declared.

References

1. Scott C. Public and private roles in health care systemsxperiences from seven OECD countries. Milton Keynes: Open University Press; 2001.
2. Goldfield N, Gnani S, Majeed A. Primary care in the United States: profiling performance in primary care in the United States. BMJ 2003 Apr 5;326(7392):744-747 [FREE Full text] [doi: 10.1136/bmj.326.7392.744] [Medline: 12676844]
3. Palmer G, Reid B. Evaluation of the performance of diagnosis-related groups and similar casemix systems: methodological issues. Health Serv Manage Res 2001 May;14(2):71-81. [Medline: 11373998]

4. Pettoello-Mantovani M, Namazova-Baranova L, Ehrich J. Integrating and rationalizing public healthcare services as a source of cost containment in times of economic crises. Ital J Pediatr 2016 Feb 24;42:18 [FREE Full text] [doi: 10.1186/s13052-016-0231-1] [Medline: 26911573]

5. National Bureau of Statistics of the People's Republic of China. Gov.cn. The 6th National Census Data Bulletin (2010) URL: http://www.gov.cn/test/2012-04/20/content_2118413.htm [accessed 2017-03-01] [WebCite Cache ID 6odoQpNtm]

6. Yang G, Wang Y, Zeng Y, Gao GF, Liang X, Zhou M, et al. Rapid health transition in China, 1990-2010: findings from the global burden of disease Study 2010. Lancet 2013 Jun 8;381(9882):1987-2015. [doi: 10.1016/S0140-6736(13)61097-1] [Medline: 23746901]

7. National HealthFamily Planning Commission of the People's Republic of China. China health statistical yearbook (2006-2015). Beijing: Pecking Union Medical College Press; 2006.

8. Thornton J. What you need to know to make the most of big data in biology. Lancet 2015 Feb 26;385(Suppl 1):S5-S6. [doi: 10.1016/S0140-6736(15)60321-X] [Medline: 25726450]

9. Fung IC, Tse ZT, Fu K. Converting big data into public health. Science 2015 Feb 6;347(6222):620. [doi: 10.1126/science.347.6222.620-b] [Medline: 25657237]

10. Crisp N. Turning the World Upside Down: The Search for Global Health in the 21st Century. New York: CRC Press; 2010.

11. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]

12. Allen LN, Christie GP. The emergence of personalized health technology. J Med Internet Res 2016;18(5):e99. [doi: 10.2196/jmir.5357] [Medline: 27165944]

13. Ministry of Industry and Information Technology of the People's Republic of China. MIIT. Communications and transport statistics bulletin 2016 URL: http://www.miit.gov.cn/n1146290/n1146402/n1146455/c5471508/content.html [accessed 2017-07-10] [WebCite Cache ID 6rqE1MSWY]

Abbreviations

DSPS: data sharing and processing system
EIP: enterprise information portal
EMR: electronic medical record
ICR: intelligent character recognition
IoT: Internet of Things
LAN: local-area network
LIS: laboratory information system
PACS: picture archiving and communication system
RFID: radio-frequency identification
RMS: regional medical consortium

XSL•FO
RenderX

Original Paper

Issues Associated With the Use of Semantic Web Technology in Knowledge Acquisition for Clinical Decision Support Systems: Systematic Review of the Literature

Seyedjamal Zolhavarieh[1], BSc, MKM; David Parry[1*], PhD; Quan Bai[1*], PhD

Department of Computer Science, Auckland University of Technology, Auckland, New Zealand

[*]these authors contributed equally

Corresponding Author:
Seyedjamal Zolhavarieh, BSc, MKM
Department of Computer Science
Auckland University of Technology
Private Bag 92006
Auckland,
New Zealand
Phone: 64 99219999
Fax: 64 99219944
Email: szolhava@aut.ac.nz

Abstract

Background: Knowledge-based clinical decision support system (KB-CDSS) can be used to help practitioners make diagnostic decisions. KB-CDSS may use clinical knowledge obtained from a wide variety of sources to make decisions. However, knowledge acquisition is one of the well-known bottlenecks in KB-CDSSs, partly because of the enormous growth in health-related knowledge available and the difficulty in assessing the quality of this knowledge as well as identifying the "best" knowledge to use. This bottleneck not only means that lower-quality knowledge is being used, but also that KB-CDSSs are difficult to develop for areas where expert knowledge may be limited or unavailable. Recent methods have been developed by utilizing Semantic Web (SW) technologies in order to automatically discover relevant knowledge from knowledge sources.

Objective: The two main objectives of this study were to (1) identify and categorize knowledge acquisition issues that have been addressed through using SW technologies and (2) highlight the role of SW for acquiring knowledge used in the KB-CDSS.

Methods: We conducted a systematic review of the recent work related to knowledge acquisition MeM for clinical decision support systems published in scientific journals. In this regard, we used the keyword search technique to extract relevant papers.

Results: The retrieved papers were categorized based on two main issues: (1) format and data heterogeneity and (2) lack of semantic analysis. Most existing approaches will be discussed under these categories. A total of 27 papers were reviewed in this study.

Conclusions: The potential for using SW technology in KB-CDSS has only been considered to a minor extent so far despite its promise. This review identifies some questions and issues regarding use of SW technology for extracting relevant knowledge for a KB-CDSS.

Introduction

Decision-making is an essential activity for clinicians. In this paper, we are primarily concerned with knowledge-based diagnostic decisions. Other decisions include image interpretation, drug discovery, and others. Such decisions are clearly critical. Clinical decision-making is a daily process for all practitioners making decisions about patient care. The quality of decisions depends on how much experience experts have and how much accurate knowledge is available. Clinical diagnostic decision-making is a complex activity and requires clinicians to have access to relevant, up-to-date, and accurate knowledge sources to support appropriate patient care. A knowledge-based clinical decision support system (KB-CDSS) uses

XSL•FO
RenderX

machine-stored knowledge to assist clinicians. Other CDSSs may learn from large amounts of data via machine learning techniques or act as a case-based reasoning system [1].

Recently, Informatics researchers have proposed several computerized methods to find relevant and accurate knowledge to assist in diagnosis. KB-CDSS requires knowledge to be available, rather than generating its own knowledge through machine learning. Knowledge-based approaches may be more effective in cases where little data is available, or there is a need for explanatory capacity. Early decision support systems such as MYCIN [2] used knowledge-based approaches, albeit from knowledge collected by experts. However, there are limitations in their use in terms of the need to fit together with the use of clinical experience. KB-CDSSs may be most useful where the clinician does not have recent experience of a particular problem or may not feel that their knowledge is up to date.

The core of each KB-CDSS contains three components including a central knowledge base (KB), an inference or reasoning engine, and a user or communication interface [3]. The KB-CDSS receives patient data and inputs and provides a diagnosis as an output. In this regard, the KB plays a vital role in this scenario for collecting, classifying, and sharing the knowledge [4].

The KB-CDSS works by extracting knowledge from various knowledge sources. However, knowledge acquisition is one of the well-known bottlenecks for any kind of KB-CDSS. Providing an intelligent mechanism for communicating between KB-CDSSs and knowledge resources is a major concern of today's researchers since inappropriate or low-quality knowledge may not give appropriate outputs. More precisely, the KB-CDSS cannot be effective if it uses limited or outdated knowledge in response to a given query about a particular disease or set of symptoms [5].

Semantic Web (SW) technology is an effort to make knowledge on the Web both human-understandable and machine-readable [6]. In the context of KB-CDSSs, there is well-known biomedical research that has used SW technologies [7-9] and semantic mechanisms [10,11] to improve the process of knowledge acquisition in KB-CDSSs [12-15]. However, it is still unclear how SW technologies can be efficiently used to support KB-CDSS knowledge acquisition. For example, the quality of extracted knowledge has not been evaluated yet, even in the SW-based KB-CDSSs.

Recently, there has been exponential growth in the amount of published medical knowledge. For example, PubMed has grown by around 4% a year and contains more than 20 million articles [16]. Available knowledge resources are very diverse in terms of formats, structure, and vocabulary. Since 2005, researchers have been developing SW-based KB-CDSSs to effectively extract knowledge from such heterogeneous environments [15,17-20].

This paper is a systematic review that aims to identify and describe the knowledge acquisition issues related to SW technologies for KB-CDSSs. It attempts to classify the issues discovered and offer suggestions for open questions in this field.

Methods

Search Criteria and Selection

In this study, a systematic review framework was applied to search, extract, and assess articles. We used a keyword search strategy to find relevant articles that contain "Semantic Web Technology" and "Clinical Decision Support System" (see Table 1). SW technologies started to be used to support KB-CDSSs [21] after 2005. In order to extract related articles, we queried PubMed, Web of Science, Journal of Biomedical Informatics, Knowledge and Information Systems, Journal of Medical Systems, Artificial Intelligence in Medicine, Current Bioinformatics, Journal of Convergence Information Technology (JCIT), eHealth Networking Application and Services, and Health Science.

Table 1. Clinical variables among responders by type of diabetes.

Search lines	Search terms	Filtered by
Line 1	"Semantic technology" OR "Semantic Web technology" OR "Semantic Web" OR "Semantic Web techniques" OR "Semantic-based" OR "Semantic-Web-based."	Title or abstract
2. AND	"Clinical Decision Support" OR "Clinical Decision making" OR "Medical Decision Support" OR "Medical Decision making" OR "Clinical Decision Support System" OR "Medical Decision Support System" OR "CDS" OR "CDSS."	Title or abstract
3. AND	"Architecture" OR "Framework" OR "System" OR "Model."	Title or abstract
4. AND	"Health" OR "disease" OR "case study" OR "public health."	Title or abstract
5. AND	"Diagnosis" OR "treatment" OR "prediction" OR "reasoning."	Title or abstract

Results

Recovered Documents

Details about the inclusion and exclusion of articles are provided in Figure 1. These queries returned 2240 articles. This is a result of querying "Semantic Web Technology" and "Clinical Decision Support System" together. We then checked the titles and abstracts to refine the set. By checking the title or abstract, we reduced the number of articles to 283. In the general selection phase, we considered those papers that pointed to the concepts of SW technologies and CDSSs together. In this regard, there is a large number of articles that discuss CDSSs. However, few of these articles review the importance and benefit of SW technologies in the area of KB-CDSSs. We only considered papers that strongly focused on improving knowledge acquisition issues in the context of KB-CDSSs through applying

SW technologies. We also excluded articles if they were not in the English language. Of the remaining 283 articles, 27 met the inclusion criteria.

categorized articles based on two main issues: (1) format and data heterogeneity and (2) lack of semantic analysis. In Textboxes 1 and 2, we describe the issues along with the number of papers that focus on the issues (n/N).

Overview

To achieve a better understanding of knowledge acquisition issues that have been addressed using SW technologies, we

Figure 1. Flow diagram showing the exclusion of articles at various stages of the research.
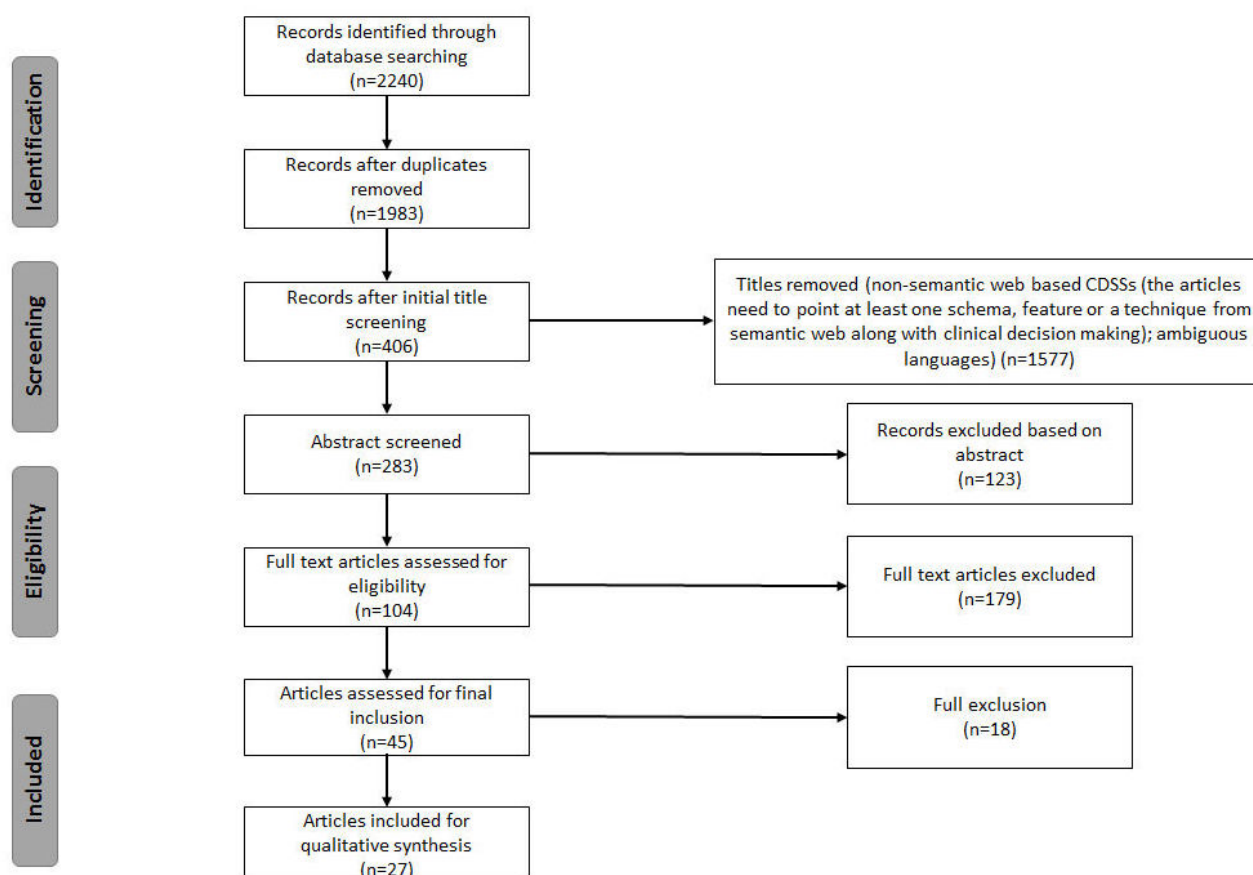
Textbox 1. Format and data heterogeneity issues (18/27).

Format and data heterogeneity were divided into the following subcategories:

- Format heterogeneity: This issue comes from different ways of representing and storing the same data. Due to the inconsistency among data models, connecting different biomedical knowledge sources is not an easy task.

- Data heterogeneity: This issue refers to the redundant results for the same entry such as having multiple entries for the same data.

- Lack of data integration: This issue is related to the lack of having a unified model for combining data residing in different sources. In this regard, clinical health care systems need a unified model to share and reuse knowledge among each other.

Textbox 2. Lack of semantic analysis issues (9/27).

Lack of Semantic Analysis was divided into two subcategories:

- Weak semantic infrastructure: Lack of a semantic infrastructure, effectively shared understanding of meaning, reduces the value of results of queries from health care knowledge sources. Resource Description Framework (RDF) is a unified data model proposed by the Semantic Web community that is useful for remedying such an issue.

- Lack of semantic definitions: Without sufficient semantic definitions, knowledge-based clinical decision support systems are not able to interpret the meaning of extracted knowledge. Such knowledge is usually encoded in the ontology (ie, schema-level) which is the backbone of Semantic Web.

XSL•FO

RenderX

This review discovered that the most important issue is related to format and data heterogeneity. Figure 2 depicts how using SW technologies helps to remedy the issues. SW technology is an efficient way to improve knowledge acquisition for several reasons such as providing an intelligent query processing mechanism rather than keyword-based answering process, providing an easy inference process, organizing the knowledge in conceptual domains, supporting consistency, facilitating knowledge extraction, supporting data integration as well as semantic interoperability, providing knowledge retrieval, and knowledge representation. In the following, we review the recent related work, which deal with SW technologies for remedying the knowledge acquisition of KB-CDSSs for the purpose of diagnosis.

Figure 2. Knowledge acquisition issues of clinical decision support systems (CDSSs) improved by Semantic Web (SW) technologies.
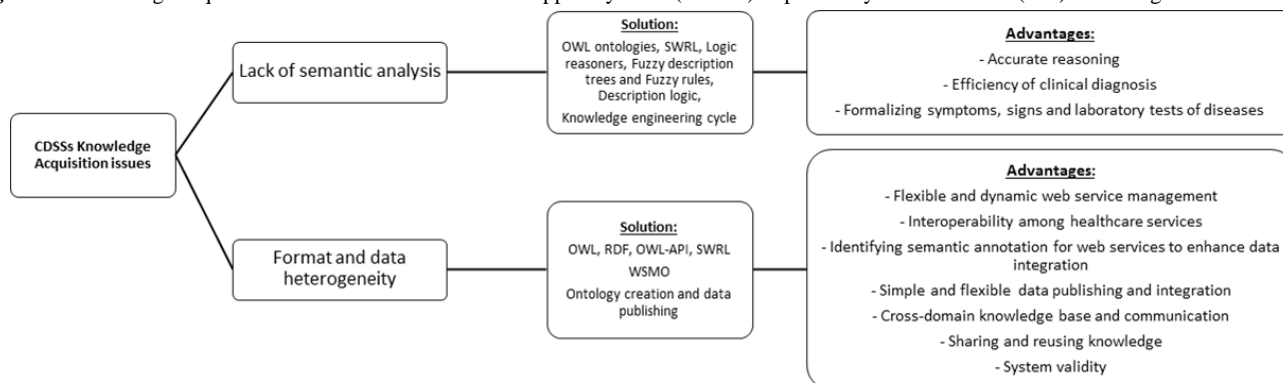
Format and Data Heterogeneity

It is important to mention that the reviewed papers proposed two different types of frameworks to overcome the issues. These frameworks, which have been developed by utilizing SW technologies, are ontological-based structures and SW services. In this regard, SW technologies have been used to boost the process of knowledge acquisition in KB-CDSSs.

Ontological-Based Structures

By using emerging SW technologies such as Resource Description Framework (RDF), Web Ontology Language (OWL), and Semantic Web Rule Language (SWRL), researchers have started to utilize SW technologies to empower and facilitate the process of knowledge sharing among KB-CDSSs. An ontology is potentially very useful in SW as it identifies the relationships between concepts in a domain. One of the most popular approaches for reducing the problem of data and format heterogeneity of KB-CDSSs is therefore to use an ontological-based structure.

[22] addressed antimicrobial health problems and inappropriate antibiotic prescribing in the health care domain. In this study, an application-independent KB-CDSS model has been developed by using formal ontological methods. The method used some SW standardizations such as OWL and SWRL to evaluate the results through intrinsic and extrinsic evaluation studies. However, this study suffers from the lack of having an accurate evaluation mechanism. The results of the study were mostly gathered in a laboratory setting rather than a clinical setting.

[23] proposed an ontological-mediated decision support system for breast cancer by utilizing SW technologies to use data for the decision-making process. The benefit of using SW technologies in such systems is to integrate heterogeneous formats of knowledge sources together. It also helps to handle complex and large datasets in order to share and reuse knowledge. Although the system is not scalable enough in a large clinical setting, it provides a flexible architecture.

Bio-DASH [24] is a SW-based prototype of a drug development dashboard. In this type of KB-CDSS, users use an RDF model to diagnose diseases, compounds, drug progression stages, molecular biology, and pathway knowledge. This system addressed the problem of sharing heterogeneous knowledge in the KB-CDSSs. To tackle this issue, the authors proposed a SW-based framework using RDF or OWL languages to describe objects and the relations between them. The framework supports data integration and user authorization. The proposed method suffers from the lack of having an appropriate platform for sharing and aggregating knowledge. High memory usage is another drawback of the proposed model.

A number of papers described a proposed clinical practitioner guideline (CPG) KB-CDSSs [1,17,25,26]. The main idea behind these series of papers is to integrate different types of ontologies such as the domain ontology, CPG ontology, and patient ontology by developing a knowledge-centric system. This system, which has been developed for the community of Breast Cancer Follow-up (BCF), contains three main components including (1) paper-based BCF CPG computerization, (2) ontology development, and (3) executing BCF CPG in a logic-based engine. Technically, this structure helps to reduce the workload of the specialist cancer center. The simple and flexible usage of data publishing and integration along with user interaction are the advantages of using SW technology in these frameworks. However, the proposed system is quite generic and needs to be validated in different situations.

[27] offered an ontology-based approach for predicting the risk of hypertension and diabetes in KB-CDSSs. To this aim, the authors used ontologies for representing patient medical profile and improving an inference mechanism for clinical decision making.

[28] proposed a SW-based KB for clinical pharmacogenetics in order to manage data. The KB has been developed by utilizing SW standardizations such as RDF and OWL. The OWL ontology contains the details of drug product labels of pharmacogenomics information. The advantages of using SW technologies have been highlighted in this study. The ontological-based structure can increase the likelihood of successful long-term maintenance and growth of KB. They are also beneficial for handling an enormous amount of datasets and share and reuse ontological concepts.

The Cleveland clinic supported a project called Semantic-DB [29] that proposed a framework to collect, store, and reuse knowledge to support sufficiency, flexibility, and extensibility of different clinical data. The reliability of research results and the accuracy of quality metrics are the addressed issues in this paper. The proposed model contains three main components: (1) content repository, (2) query interface, and (3) data production. The results obtained by the method show that the system can guarantee the quality of care measurements. It has also reduced the duplicate efforts as well as imposing transparency to deduct errors in the reported data. For the future, this research needs to be improved from different aspects such as ontology alignment, maintaining semantic alignment, and improving performance.

Finally, [30] focused on answering the question of "how the SW tools such as ontologies and rules can be applied to connect the medical and oral health (M-OH) domains by developing a KB." The KB can be reused by the medical information systems for semantic interoperability and reasoning process. The system has been developed by utilizing OWL and SWRL rules. According to the results, effectiveness in reasoning, comprehensive cross-domain KB, and cross-domain communication are the strengths of the proposed system.

Semantic Web (SW) Services

COCOON glue [21] is an SW-based service to integrate complex eHealth services. It uses Web Service Modeling Ontology (WSMO) with an open source f-logic inference engine called Flora2 to run over an open source deductive database system. This system aims at reducing medical errors and developing an efficient Web service management system to publish, discover, and compose services. This system has two main advantages: (1) providing a clear separation between the ontologies and (2) preparing good performance. The major weakness of this study is related to the use of f-logic technique for defining similarity metrics. The f-logic is a set of predefined rules for making deductions. Basically, methods developed by f-logic technique are not scalable enough and cannot be applied on the large volume of data.

ARTEMIS [31,32] is a project supported by the European Commission based on the SW services using the OWL. The structure of this system is similar to COCOON. It aims to describe the semantics of Web service functionality. It also supports the semantic meaning of messages or documents exchanged through Web services. As previously mentioned, using SW technologies not only enables health care services to easily interact with each other but also helps to integrate data across the clinical Web service by using semantic annotations.

However, this system does not provide a secure platform for protecting data.

[15] addressed the interoperability problem in both the domain of data integration and heterogeneous systems. They proposed a SW-based service framework to tackle the problem and empower the semantic interoperability among health care systems.

Despite improving health care quality, sharing and extracting knowledge in a heterogeneous environment is the most popular limitation among KB-CDSSs. Therefore, [33] proposed a sharable KB-CDSS that meets this challenge. This system has a SW service framework to identify, access, and leverage independent and reusable knowledge modules located in the central KB. The knowledge modules are defined by the ontological model, terminologies, and representation formalisms to support sharable KB-CDSS. Their contributions consist of representing unified knowledge and patient data in heterogeneous domains, knowledge integration and data interoperation, and semantic development of sharable knowledge for automated knowledge acquisition. This system has been evaluated by two applications including model-level and application-level evaluation. The coherent knowledge representation is confirmed by model-level evaluation. The high accuracy and completeness is validated by application-level evaluation. These evaluations show this system is feasible and useful in providing sharable and reusable knowledge for the purpose of diagnosis in decision making. It is also offers time-saving benefits and cost effectiveness in comparison with the other KB-CDSSs. The system improves the maintainability and scalability of systems to contribute with the other KB-CDSS.

[34] suggested a SW-based framework to support reasoning to remedy diagnostic errors. The authors believe that diagnostic errors are derived from flawed reasoning, incomplete knowledge, faulty information discovery, and inappropriate decision making. This approach contains a case-based fuzzy cognitive mapping to support diagnosis. The framework also evaluates the clinical knowledge for decision making by using Bayesian belief networks. The reasoning methods for this framework used statistical approach to solve the diagnosis issues and enhance the efficiency of the system. The reasoning methods used in this approach are implemented by using the SW tools such as Notation 3 or RDF and Euler Sharp inference engine. The strength of this system is in handling approximate reasoning, incomplete information, control rules for clinical conditions, and patient profiling. This approach is in the first stages of development for implementation. It needs to be tested with larger datasets and allowing updating of the system by integrating new knowledge.

Another study proposed by [35] developed a multiagent framework called MAPP4MD to provide a privacy preserving mechanism for clinical data in a heterogeneous environment. In this study, each agent utilizes ontologies and SW technologies to apply reasoning for a privacy-preserving algorithm. This approach supports data integration and sharing among agents in the various environments for knowledge discovery. The evaluations of this system show that the distributed multiagent

framework is flexible. One of the benefits of this approach is to improve data sharing for medical research, population-level analysis, and evaluation of population-level in health care activities. Although this framework works fine in the limited datasets, it needs to be checked in the larger datasets to show its scalability.

[36] addressed the problem of standalone KB-CDSS and having a universal KB-CDSS. The authors developed a semiautomated approach to discover, select, and compose KB-CDSSs available as Web services. The proposed system is at the elementary level and needs to be implemented and validated. The lack of identifying formalized semantics attached to the services is an obvious challenge for this research.

Lack of Semantic Analysis

The reviewed papers in this section have proposed two SW frameworks to improve the lack of semantic analysis. They consist of knowledge engineering technique and logic reasoning structure. The main goal of these papers is to improve knowledge acquisition process in the KB-CDSSs by utilizing SW technologies.

Knowledge Engineering Technique

Many of the non–SW-based KB-CDSSs suffer from the lack of automatic analysis systems. This issue can be addressed by using SW technologies. A knowledge engineering approach was taken in [37,38] for detecting Alzheimer's disease to help physicians to detect it in the early stages by using multidisciplinary knowledge and reasoning over the underlying KBs. In this paper, researchers used ontologies (eg, MIND ontology, Semantic Web applications in neuromedicine [SWAN] ontology, and systematized nomenclature of medicine-clinical terms [SNOMED-CT]). Although this project needs to be tested on the larger ontological domains, the authors improved the accuracy of results for further decision-making processes. In 2012, the system improved for discovering new knowledge and generating new rules for clinical decision making [39,40]. It is important to mention that physicians take advantage of this system to help patients discover relevant knowledge for Alzheimer's disease diagnosis. This KB-CDSS not only works in the Alzheimer's disease domain but also supports the other domains such as cancer.

The authors in 2013 [40] proposed a more generic software architecture called S-KB-CDSS to solve some of the challenges of KB-CDSSs. They improved the system by adding new tasks to the system such as diagnosis, prognosis, treatment, evolution, and prevention. It helps the system to integrate and reutilize clinical workflow of KB-CDSSs. They mentioned that discovering new knowledge methods in a previous study was implicit and that they want to solve other challenges of KB-CDSSs. They mentioned that because of the nature of a system that is based on knowledge model provided by a team of domain experts, classical validation is not possible in this stage. Therefore, they assumed that the system is correct. They validated their system by comparing system decisions with end-user decisions.

In another paper, [41] developed a model for semantic enhancement of KB-CDSS by using knowledge engineering technique to express the domain of knowledge and the patient data in a unified model. The architectures included four different phases: (1) knowledge acquisition, (2) knowledge representation, (3) knowledge application, and (4) knowledge evaluation. The main motivations for developing such architectures were to handle multidisciplinary and heterogeneous platforms. The authors claimed that their system was useful because it could reduce the reduplication of data in the KB. However, it needed to support experience based reasoning, as well as bridge the gap between semantic health care KB and existing knowledge representation model.

Another knowledge engineering approach that aims to improve the performance of KB-CDSS was proposed by [42]. This approach answers queries by integrating deterministic and plausible knowledge from heterogeneous environments. Researchers in this study used SW technologies to leverage reasoning and extend the coverage of a medical KB. Extending the coverage of medical KB, by considering potential correlations between decisional attributes is useful, especially when KB-CDSSs need to have complete knowledge for decision making.

There is some rationale for using SW technologies in this approach, such as data management, description logic (DL)–based inferring methods, and the opportunity to support plausible reasoning. Moreover, using ontology inference and conceptual similarity check, improves the accuracy of reasoning in the system. The result of the system evaluation shows that this multi strategy approach improves knowledge coverage of clinical KB and helps to have better diagnostic process for complex diseases. In addition, inferred knowledge can be used in future decision making.

Logic Reasoning Structure

[43] proposed a knowledge-based preoperative decision support system to assist health professionals in secondary care in the preoperative assessment of patient before elective surgery. In this system, the authors applied SW technologies such as OWL and logic reasoning to develop an automatic analysis system. The system attaches patient information to the medical context. However, the collected information from patients is still a kind of "coarse-grained" information and needs to be transformed into a "fine-grained" model.

[44] proposed a personalized treatment flow without user intervention. The method was developed by using fuzzy decision tree, fuzzy rules, and SWRL. The advantages of such systems are to provide a user-friendly environment to improve memory performance and to reduce the time on patient care. The scalability of the proposed model is still under investigation.

SeDelo [45] suggested a computer-aided diagnostic system to help experts and nonexperts to recommend the clinical diagnosis. In this study, the authors developed a KB-CDSS by utilizing SW technologies and description logics to diagnose diseases by using symptoms, signs, and laboratory tests. This system is more efficient and accurate in decision-making processes compared with previous systems proposed by the same authors. Although this method achieves a better result in terms of the

accuracy of the system, it is still not scalable enough and needs to be developed for the rule description process.

Discussion

Principal Findings

SW technology and its applications are useful since in principle they can deal with data from multiple sources and facilitate machine-machine communication. The SW is an effort to make knowledge on the Web both human-understandable and machine-readable. There is no need to provide a database schema for sharing data since it has its own universal data structure and can be used among knowledge sources. SW technologies and their features such as semantic interoperability, knowledge integration, and knowledge reusing to upgrade and transform old applications into modern and intelligent models [46].

In order to make well-informed decisions, health care applications need to be able to discover knowledge among many heterogeneous KBs. Having diverse data models and formats lead researchers to use SW technologies to facilitate data integration processes. SW technologies allow researchers to analyze incompatible biological descriptions in one unified format. For example, using SW technologies helps to mesh datasets about protein-protein interaction to reveal obscure correlations that could help identify promising medications [7].

In the context of KB-CDSSs, different issues have been improved by utilizing SW technologies such knowledge acquisition and data collection, and data integration of clinical systems. In this paper, we have reviewed and highlighted the issues of knowledge acquisition improved by SW technologies. The review shows some of the potential approaches of SW technology in supporting KB-CDSSs.

A Proposed Model for Semantic Web (SW) Use in Knowledge-Based Clinical Decision Support System (KB-CDSS)

To discuss how the knowledge can be discovered and updated with the SW technologies, we have proposed a knowledge broker framework to apply in a KB-CDSS [47]. In this framework, we focus on assessing clinical knowledge and delivering high-quality knowledge for a KB-CDSS. The overall framework is shown in Figure 3. The proposed model contains five major parts.

Figure 3. A proposed model for using Semantic Web (SW) in knowledge acquisition for knowledge-based clinical decision support system (KB-CDSS).
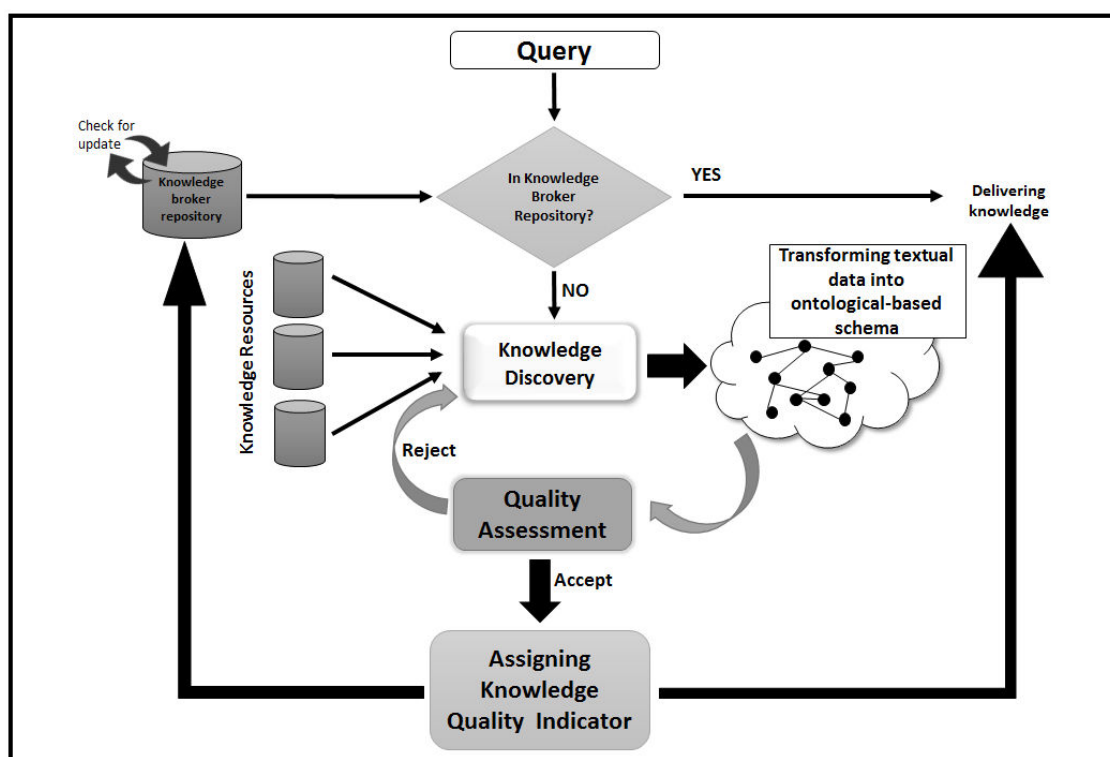
Knowledge Discovery

After receiving a query, the knowledge broker will check the existing knowledge in its repository to find a related result. If the knowledge exists in the repository, the system will deliver the knowledge immediately, if it does not exist, the new knowledge will be extracted from electronic knowledge resources (especially PubMed) based on query characteristics.

Note that in the system, there is a knowledge repository, which records all of the extracted knowledge with a knowledge quality indicator. The knowledge quality indicator can support relevancy, currency, and accuracy of knowledge to use in decision making. The knowledge repository will check the quality of knowledge regularly to provide high quality knowledge every time. If the knowledge needs updating, the

knowledge broker improves the knowledge and sends it back to the repository.

We assume that all of the knowledge that we have used is of the OWL or Semantic format.

Constructing Knowledge

In this step, the extracted knowledge will be converted to the ontology format and annotated by other information to enrich the knowledge. To achieve this, Protégé Ontology editor has been utilized. In the knowledge discovery step, the knowledge broker may extract several items of knowledge that are useful. Therefore, we may use their information to annotate the extracted knowledge. The output will be an enriched knowledge for the system.

Quality Assessment

This step is related to checking the quality of the knowledge to ensure it is useful for decision making. We may use different metrics to check the quality of knowledge.

Assigning Knowledge Quality Indicator (KQI)

In this step, a knowledge quality indicator will be assigned to the knowledge item to show how much knowledge is qualified. The knowledge quality indicator (KQI) can support the approval for knowledge quality. It will be more useful to have a knowledge quality indicator when reusing the knowledge in the future.

Updating Knowledge Repository and Delivering Knowledge

Finally, the high quality knowledge will be sent to the knowledge repository to be used again. It will deliver to the KB-CDSS for decision making respectively.

Limitations and Future Directions

Although SW technologies improve the problem of knowledge acquisition in KB-CDSSs, there are still some issues that have not been considered yet. For example, in the context of KB-CDSSs, most existing methods do not properly evaluate the quality of extracted knowledge. Here the question is how to make sure that the knowledge used by KB-CDSSs is reliable.

Conventional search engines cannot evaluate whether the knowledge is accurate, reliable, and relevant in the case of comorbidities. Inappropriate knowledge can have negative effects on the decision-making process. Therefore, there is a need to propose new methods to check the quality of extracted knowledge using SW technologies for KB-CDSSs.

There are also some limitations in applying SW technologies to systems such as KB-CDSSs. Apart from using and managing personal data and knowledge, the privacy issues around using SW could be a significant problem in such systems, primarily because everything that is published online will be shared using SW technologies. Another issue that can be problematic for applying SW technologies can be resouce requirements to support complete features of SW. The SW technology may need some specific resources to work; however, some of them may not exist in the current environment. There is still a very long way to go before the SW dream becomes true and changes the information society and the information economy. SW technologies aim to convert syntactical structure to semantical structure. They also aim to facilitate the process of retrieving information to delegating tasks. In this regard, health informatic experts need to make efforts to utilize SW technologies in the body of CDSSs.

In addition, we have identified a number of still-open questions: (1) Are KB-CDSS approaches more effective than machine learning systems, or should they be combined with them? (2) How can the quality of knowledge discovered by a SW approach be evaluated? (3) Is it possible or even desirable to use an SW approach to automatically update KB-CDSSs? and (4) Should knowledge sources conform to a particular standard in order to support SW-based knowledge acquisition, and would this justify the overhead associated with such work?

Conclusions

The rise of precision medicine is also a key driver in the need to identify both knowledge and data from heterogeneous sources [48,49]. The aim of this systematic review paper is to highlight the importance of using SW technologies for improving the knowledge acquisition issues in the context of KB-CDSSs. In this paper, the potential for using SW technology has been described. We have categorized the recent knowledge acquisition issues of KB-CDSSs improved by SW technologies into 2 main groups including format and data heterogeneity and lack of semantic analysis. In this regard, we have reviewed the recent related work in this context to highlight the necessity of using SW technologies in the body of current KB-CDSSs.

As discussed previously, the existing health care search engines (ie, PubMed and Clinical Trials) do not comprehensively extract and identify high quality knowledge for serving in the KB-CDSSs. The ever-growing amount of clinical knowledge makes the process of extracting high quality knowledge increasingly difficult. None of the reviewed papers have addressed the issue of quality assessment for KB-CDSSs. For future work, we aim to develop an automatic system to measure, extract, and rate the high quality knowledge for KB-CDSSs [47]. Such a system should be able to support knowledge brokers to extract and rate knowledge from multiple heterogeneous sources (ie, PubMed and other sources) to keep KB-CDSSs current and provide optimal decision making. There is also the possibility of integrating such systems with a precision medicine–based approach [49] to allow a KB-CDSS to discover appropriate cases and outcomes that may need to be included in rule revision.

Conflicts of Interest

None declared.

References

1. Hussain S, Abidi SSR. Integrating healthcare knowledge artifacts for clinical decision support: towards semantic web based healthcare knowledge morphing. In: Artificial Intelligence in Medicine. Berlin: Springer-Verlag; 2009:171-175.

2. Shortliffe E, editor. Computer-based medical consultations: MYCIN. New York: Elsevier; 1976.

3. O'Kane T, O'Donoghue J, Gallagher J, Aftab A, Casey A, Angove P, et al. MEWS to e-MEWS: From a paper-based to an electronic clinical decision support system. In: Proceedings of the 4th European Conference on Information Management and Evaluation. 2010 Presented at: ECIME; 2010; Portugal p. 301.

4. Leuf B. The semantic web: crafting infrastructure for agency. Chichester, West Sussex, England: John Wiley & Sons Ltd; 2006.

5. von Michalik K, Kwiatkowska M, Kielan K. Application of knowledge-engineering methods in medical knowledge management. In: Fuzziness and Medicine: Philosophical Reflections and Application Systems in Health Care. Berlin, Heidelberg: Springer-Verlag; 2013:205-214.

6. Brüggemann S, d'Amato C, editors. Collaboration and the Semantic Web: Social Networks, Knowledge Networks and Knowledge Resources. Hershey, PA: Information Science Reference; 2012.

7. Feigenbaum L, Herman I, Hongsermeier T, Neumann E, Stephens S. The semantic web in action. Sci Am 2007 Dec;297(6):64-71. [Medline: 18237102]

8. Wroe C. Is semantic web technology ready for healthcare? 2006 Presented at: 3rd Annual European Semantic Web Conference (ESWC'06); 2006; Budva, Montenegro.

9. Tao C, Jiang G, Oniki TA, Freimuth RR, Zhu Q, Sharma D, et al. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. J Am Med Inform Assoc 2013 May 01;20(3):554-562 [FREE Full text] [doi: 10.1136/amiajnl-2012-001326] [Medline: 23268487]

10. Al-Mubaid H, Nguyen H. Measuring semantic similarity between biomedical concepts within multiple ontologies. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2009 Jul;39(4):389-398.

11. Al-Mubaid H, Nguyen HA. A cluster-based approach for semantic similarity in the biomedical domain. Conf Proc IEEE Eng Med Biol Soc 2006;1:2713-2717. [doi: 10.1109/IEMBS.2006.259235] [Medline: 17946134]

12. Schulz S, Martínez-Costa C. How Ontologies Can Improve Semantic Interoperability in Health Care. In: Process Support and Knowledge Representation in Health Care. Lecture Notes in Computer Science. Cham: Springer; 2013:1-10.

13. Blomqvist E. The use of Semantic Web technologies for decision support- a survey. Semant Web 2014;5(3):177-201.

14. Pathak J, Kiefer RC, Chute CG. Using semantic web technologies for cohort identification from electronic health records for clinical research. AMIA Jt Summits Transl Sci Proc 2012;2012:10-19 [FREE Full text] [Medline: 22779040]

15. Sonsilphong S, Archint N. Semantic Interoperability for data integration framework using semantic web services and rule-based inference: A case study in healthcare domain. Journal of Convergence Information Technology (JCIT) 2013 Feb;8(3):150-159.

16. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) 2011 Jan 18;2011:baq036 [FREE Full text] [doi: 10.1093/database/baq036] [Medline: 21245076]

17. Hussain S, Abidi S, Abidi S. Semantic web framework for knowledge-centric clinical decision support systems. In: AIME '07 Proceedings of the 11th conference on Artificial Intelligence in Medicine.: Springer-Verlag; 2007 Presented at: Artificial Intelligence in Medicine; July 2007; Amsterdam, The Netherlands p. 451-455. [doi: 10.1007/978-3-540-73599-1_60]

18. Wright A, Sittig DF. SANDS: a service-oriented architecture for clinical decision support in a National Health Information Network. J Biomed Inform 2008 Dec;41(6):962-981. [Medline: 18434256]

19. Dixon BE, Simonaitis L, Goldberg HS, Paterno MD, Schaeffer M, Hongsermeier T, et al. A pilot study of distributed knowledge management and clinical decision support in the cloud. Artif Intell Med 2013 Sep;59(1):45-53. [doi: 10.1016/j.artmed.2013.03.004] [Medline: 23545327]

20. Minutolo A, Esposito M, De Pietro G. A pattern-based knowledge editing system for building clinical Decision Support Systems. Knowledge-Based Systems 2012 Nov;35:120-131.

21. Valle ED, Cerizza D. COCOON glue: a prototype of WSMO discovery engine for the healthcare field. In: Proceedings of the WIW 2005 Workshop on WSMO Implementations. 2005 Presented at: 2nd WSMO Implementation Workshop; July 2005; Innsbruck, Austria.

22. Bright TJ, Yoko Furuya E, Kuperman GJ, Cimino JJ, Bakken S. Development and evaluation of an ontology for guiding appropriate antibiotic prescribing. J Biomed Inform 2012 Feb;45(1):120-128 [FREE Full text] [doi: 10.1016/j.jbi.2011.10.001] [Medline: 22019377]

23. Dasmahapatra S, Dupplaw D, Hu B, Lewis P, Shadbolt N. Ontology-mediated distributed decision support for breast cancer. Berlin: Springer; 2005 Presented at: 10th Conference on Artificial Intelligence in Medicine; 23-27 July; Aberdeen, UK p. 221-225.

24. Neumann EK, Quan D. BioDash: a Semantic Web dashboard for drug development. Pac Symp Biocomput 2006:176-187. [Medline: 17094238]

25. Abidi SR, Abidi SS, Hussain S, Shepherd M. Ontology-based modeling of clinical practice guidelines: a clinical decision support system for breast cancer follow-up interventions at primary care settings. Stud Health Technol Inform 2007;129(Pt 2):845-849. [Medline: 17911835]

XSL•FO

RenderX

26.   Jafarpour B, Abidi SR, Abidi SSR. Exploiting OWL reasoning services to execute ontologically-modeled clinical practice guidelines. In: Proceedings of the 13th conference on Artificial intelligence in medicine.: Springer; 2011 Presented at: Conference on Artificial Intelligence in Medicine; 2-6 July 2011; Bled, Slovenia p. 307-311.

27.   Sherimon PC, Vinu PV, Krishnan R, Takroni Y. Ontology Based System Architecture to Predict the Risk of Hypertension in Related Diseases. International Journal of Information Processing and Management 2013 Jun;4(4):44-50. [doi: 10.4156/ijipm]

28.   Samwald M, Freimuth R, Luciano JS, Lin S, Powers RL, Marshall MS, et al. An RDF/OWL knowledge base for query answering and decision support in clinical pharmacogenetics. Stud Health Technol Inform 2013;192:539-542 [FREE Full text] [Medline: 23920613]

29.   Pierce CD, Booth D, Ogbuji C, Deaton C, Blackstone E, Lenat D. SemanticDB: a semantic web infrastructure for clinical research and quality reporting. Curr. Bioinform 2012;7(3):267-277.

30.   Shah T, Rabhi F, Ray P, Taylor K. CoRR. 2013. Enhancing automated decision support across medical and oral health domains with semantic web technologies URL: http://mo.bf.rmit.edu.au/acis2013/382.pdf [accessed 2017-06-22] [WebCite Cache ID 6rPLXJ9sG]

31.   Bicer V, Laleci GB, Dogac A, Kabak Y. Artemis message exchange frameworkmantic interoperability of exchanged messages in the healthcare domain. ACM Sigmod Record 2005 Sep;34(3):71-76.

32.   Dogac A, Laleci GB, Kirbas S, Kabak Y, Sinir SS, Yildiz A, et al. Artemis: deploying semantically enriched web services in the healthcare domain. Information Systems 2006;31(4-5):321-339.

33.   Zhang YF, Gou L, Tian Y, Li TC, Zhang M, Li JS. Design and Development of a Sharable Clinical Decision Support System Based on a Semantic Web Service Framework. J Med Syst 2016 May;40(5):118. [Medline: 27002818]

34.   Douali N, Csaba H, De Roo J, Papageorgiou EI, Jaulent MC. Diagnosis support system based on clinical guidelines: comparison between case-based fuzzy cognitive maps and Bayesian networks. Comput Methods Programs Biomed 2014;113(1):133-143. [doi: 10.1016/j.cmpb.2013.09.012] [Medline: 24599907]

35.   Wimmer H, Yoon V, Sugumaran V. A multi-agent system to support evidence based medicine and clinical decision making via data sharing and data privacy. Decis Support Syst 2016 Aug;88:51-66.

36.   Hederman L, Khan H. ResearchGate. 2015. A Universal Clinical Decision Support System using semantic web services URL: https://www.researchgate.net/profile/Lucy_Hederman/publication/ 268293066_A_Universal_Clinical_Decision_Support_System_using_semantic_web_services/links/ 54ff3bcc0cf2eaf210b82c95.pdf [accessed 2017-06-11] [WebCite Cache ID 6r9VbmHXv]

37.   Sanchez E, Toro C, Carrasco E, Bonachela P, Parra C, Bueno G, et al. A Knowledge-based Clinical Decision Support System for the diagnosis of Alzheimer Disease. In: e-Health Networking Applications and Services (Healthcom). 2011 Presented at: 13th IEEE International conference on e-Health Networking Applications and Services (Healthcom); 2011 June; Columbia, MO, USA p. 351-357.

38.   Sanchez E, Toro C, Carrasco E, Bueno G, Parra C, Bonachela P, et al. An architecture for the semantic enhancement of clinical decision support systems. In: Knowlege-Based and Intelligent Information and Engineering Systems. Berlin, Heidelberg: Springer; 2011:611-620.

39.   Toro C, Sanchez E, Carrasco E, Mancilla-Amaya L, Sanín C, Szczerbicki E, et al. Using set of experience knowledge structure to extend a rule set of clinical decision support system for Alzheimer's disease diagnosis. Cybernet. Syst 2012;43(2):81-95.

40.   Sanchez E, Toro C, Artetxe A, Graña M, Sanin C, Szczerbicki E, et al. Bridging challenges of clinical decision support systems with a semantic approach:A case study on breast cancer. Pattern Recogn. Lett 2013 Oct 15;34(14):1758-1768.

41.   Zhang YF, Tian Y, Zhou TS, Araki K, Li JS. Integrating HL7 RIM and ontology for unified knowledge and data representation in clinical decision support systems. Comput Methods Programs Biomed 2016 Jan;123:94-108. [doi: 10.1016/j.cmpb.2015.09.020] [Medline: 26474836]

42.   Mohammadhassanzadeh H, Van Woensel W, Abidi SR, Abidi SSR. A semantic web-based approach to plausible reasoning for improving clinical knowledge engineering. In: Biomedical and Health Informatics (BHI). 2016 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); 24-27 Feb. 2016; Las Vegas, NV, USA. [doi: 10.1109/BHI.2016.7455950]

43.   Bouamrane MM, Rector A, Hurrell M. Using OWL ontologies for adaptive patient information modelling and preoperative clinical decision support. Knowl Inf Syst 2011 Nov;29(2):405-418.

44.   Muthuraman T, Sankaran G. A framework for personalized decision support system for the healthcare application. Health Science Journal 2014;8(2):262-273.

45.   Rodríguez-González A, Labra-Gayo JE, Colomo-Palacios R, Mayer MA, Gómez-Berbís JM, García-Crespo A. SeDeLo: using semantics and description logics to support aided clinical diagnosis. J Med Syst 2012 Aug;36(4):2471-2481. [doi: 10.1007/s10916-011-9714-1] [Medline: 21537850]

46.   Lozano-Rubí R, Pastor X, Lozano E. OWLing clinical data repositories with the ontology web language. JMIR Med Inform 2014;2(2):e14. [Medline: 25599697]

47.   Zolhavarieh S, Parry D. A model of knowledge quality assessment in clinical decision support system. 2015 Presented at: 14th Annual HiNZ conference; 2015; Christchurch, New Zealand.

48.   Isaac A, Haslhofer B. Europeana linked open data - data.europeana.eu. Semantic Web 2013 Jul;4(3):291-297.
49.   Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795 [FREE Full text] [doi: 10.1056/NEJMp1500523] [Medline: 25635347]

Abbreviations

AD: Alzheimer's disease
BCF: Breast Cancer Follow-up
CDSS: clinical decision support system
CPG: clinical practitioner guideline
KB: knowledge base
KB-CDSS: knowledge-based clinical decision support system
KQI: knowledge quality indicator
M-OH: medical and oral health
OWL: Web Ontology Language
SNOMED-CT: systematized nomenclature of medicine-clinical terms
SW: Semantic Web
SWRL: Semantic Web Rule Language

XSL•FO
RenderX

<u>Original Paper</u>

Estimating One-Year Risk of Incident Chronic Kidney Disease: Retrospective Development and Validation Study Using Electronic Medical Record Data From the State of Maine

Shiying Hao[1,2,3*], PhD; Tianyun Fu[4*], BS; Qian Wu[5,6*], BS; Bo Jin[4], MS; Chunqing Zhu[4], MS; Zhongkai Hu[2,3], PhD; Yanting Guo[5,7], BS; Yan Zhang[5,8], MD; Yunxian Yu[1], MD, PhD; Terry Fouts[9], MD; Phillip Ng[10], MD; Devore S Culver[11], MM; Shaun T Alfreds[11], MBA; Frank Stearns[4], MHA; Karl G Sylvester[5], MD; Eric Widen[4], MHA; Doff B McElhinney[2,3*], MD; Xuefeng B Ling[1,3,5*], PhD

[1]Department of Epidemiology and Health Statistics, School of Public Health, School of Medicine, Zhejiang University, Hangzhou, China

[2]Department of Cardiothoracic Surgery, Stanford University, Stanford, CA, United States

[3]Clinical and Translational Research Program, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Palo Alto, CA, United States

[4]HBI Solutions Inc, Palo Alto, CA, United States

[5]Department of Surgery, Stanford University, Stanford, CA, United States

[6]China Electric Power Research Institute, Beijing, China

[7]School of Management, Zhejiang University, Hangzhou, China

[8]Department of Oncology, The First Hospital of Shijiazhuang, Shijiazhuang, China

[9]Empactful Capital, San Francisco, CA, United States

[10]Sequoia Hospital, Redwood City, CA, United States

[11]HealthInfoNet, Portland, ME, United States

[*]these authors contributed equally

Corresponding Author:
Xuefeng B Ling, PhD
Department of Surgery
Stanford University
S370 Grant Bldg
Stanford, CA,
United States
Phone: 1 650 427 9198
Fax: 1 650 723 1154
Email: bxling@stanford.edu

Abstract

Background: Chronic kidney disease (CKD) is a major public health concern in the United States with high prevalence, growing incidence, and serious adverse outcomes.

Objective: We aimed to develop and validate a model to identify patients at risk of receiving a new diagnosis of CKD (incident CKD) during the next 1 year in a general population.

Methods: The study population consisted of patients who had visited any care facility in the Maine Health Information Exchange network any time between January 1, 2013, and December 31, 2015, and had no history of CKD diagnosis. Two retrospective cohorts of electronic medical records (EMRs) were constructed for model derivation (N=1,310,363) and validation (N=1,430,772). The model was derived using a gradient tree-based boost algorithm to assign a score to each individual that measured the probability of receiving a new diagnosis of CKD from January 1, 2014, to December 31, 2014, based on the preceding 1-year clinical profile. A feature selection process was conducted to reduce the dimension of the data from 14,680 EMR features to 146 as predictors in the final model. Relative risk was calculated by the model to gauge the risk ratio of the individual to population mean of receiving a CKD diagnosis in next 1 year. The model was tested on the validation cohort to predict risk of CKD diagnosis in the period from January 1, 2015, to December 31, 2015, using the preceding 1-year clinical profile.

Results: The final model had a *c*-statistic of 0.871 in the validation cohort. It stratified patients into low-risk (score 0-0.005), intermediate-risk (score 0.005-0.05), and high-risk (score ≥ 0.05) levels. The incidence of CKD in the high-risk patient group was 7.94%, 13.7 times higher than the incidence in the overall cohort (0.58%). Survival analysis showed that patients in the 3 risk categories had significantly different CKD outcomes as a function of time ($P<.001$), indicating an effective classification of patients by the model.

Conclusions: We developed and validated a model that is able to identify patients at high risk of having CKD in the next 1 year by statistically learning from the EMR-based clinical history in the preceding 1 year. Identification of these patients indicates care opportunities such as monitoring and adopting intervention plans that may benefit the quality of care and outcomes in the long term.

Introduction

Chronic kidney disease (CKD) is a major public health concern in the United States. The National Health and Nutrition Examination Survey (NHANES) reported a prevalence of 15.2% in the general population [1], and it is growing annually, from less than 2% in 2000 to nearly 4.5% in 2008 [2]. The end-stage renal disease prevalence was 2067 per million in the United States in 2014, ranging from 965 to 1754 per million in different health service areas in Maine [3]. CKD is highly associated with other chronic conditions such as diabetes, hypertension, and cardiovascular defects and is associated with poor outcomes and high resource burden [4,5]. Timely recognition and treatment of patients with CKD has been shown to reduce the risk of mortality and complications and slow down disease progression [6-9]. Taken together, these factors highlight a critical need for early detection and intervention to mitigate the impact of CKD.

A barrier to timely recognition and management of CKD is the long clinically silent phase of the disease. Patients with CKD tend to be asymptomatic in the early stage, resulting in generally low awareness of the disease. NHANES reported a self-awareness rate of less than 10% among patients with CKD at stages 1 to 3 and less than 50% at stage 4 [10]. Low awareness of CKD was also found at the provider level, mainly due to poor documentation of the disease and lack of knowledge and education about disease recognition [11-13]. The low awareness at both patient and clinician levels is an impediment to improving the quality of patient care. To increase awareness and thus improve the early recognition from both sides, annual screening with CKD diagnostic testing including serum creatinine and urine albumin testing was recommended for patients at increased risk of CKD, including those with diabetes, hypertension, or family history of kidney disease [14,15]. Yet the existing screening guidelines focus on selected patients rather than the general population, which inevitably tends to ignore a number of CKD patients, especially for those without a history of diabetes or hypertension.

Recent attempts to improve the timely recognition of CKD include identifying risk factors predictive of CKD and combining them to develop a risk score [16-23]. Risk scores stratify individuals based on their probability of having incident CKD or further progression, which can help clinicians to make decisions about intervention. Limitations of those efforts include lack of generalizability across the population, insufficient predictive accuracy, loss to follow-up, and dependence on specific laboratory test results. So far, there is no widely accepted risk assessment model implemented for clinical use in a large, general population.

The widespread use of electronic medical records (EMRs) affords a unique opportunity to understand health care status and improve care management at the population level. The successful use of EMR data to develop risk scores for population stratification has facilitated better patient care for other conditions [24-28]. Enabled by information technology, analysis using EMR data provides a unique perspective on population health tendencies, with large numbers of patients and high dimensional clinical data elements. In this study, we aimed to develop an EMR-based risk model to estimate the probability of receiving an incident diagnosis of CKD within the next 1 year. The model was derived through statistical learning from patients' prior 1-year clinical history, combined with domain knowledge of risk factors of CKD. The data sources were EMRs collected from 35 hospitals, 34 federally qualified health centers, and more than 400 ambulatory practices in the state of Maine covering more than 1 million patients [27,29]. We aimed to predict patients with newly recognized CKD within the next 1 year. The term "recognized CKD" included patients having diagnosis codes from the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) associated with CKD diagnosis. We hypothesized that the proposed risk model would be able to identify high-risk patients prior to the assignment of a CKD diagnosis code. To our knowledge it is the first study to predict the 1-year risk of being diagnosed with CKD by using EMR data in an all-age, all-disease, and all-payer group general population.

Methods

Reporting Method

The study was reported according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines for a derivation and validation predictive model [30] (Multimedia Appendix 1).

Ethics Statement

Protected personal health information was removed for the purpose of this research. Because it analyzed de-identified data, this study was exempted from ethics review by the Stanford University Institutional Review Board (October 16, 2014).

Studied Population and Source of Data

Patient information for this study was extracted from the Health Information Exchange (HIE) dataset administered by HealthInfoNet, an independent nonprofit organization. The dataset contains records of nearly 95% of the population in the state of Maine. Data elements include demographic information, socioeconomic status, laboratory and radiographic tests coded according to Logical Observation Identifier Names and Codes, outpatient medication prescriptions coded according to the National Drug Code, and primary and secondary diagnoses and procedures which are coded using ICD-9-CM. Missing data handling is described in Multimedia Appendix 2. The study included patients who visited any care facility in the Maine HIE

network any time from January 1, 2013, through December 31, 2015. Patients who died or had a history of CKD diagnosis at any time between 2009 (the first time deployment of any EMR system in the state of Maine) and the time point of prediction or a history of treatment or diagnosis for end-stage renal disease were excluded from the study.

Outcome Definition

In this study, a CKD case was defined as having an ICD-9-CM diagnosis code of CKD assigned during any visit during the next 1 year, which refers to the period from January 1, 2014, to December 31, 2014, in the derivation cohort and from January 1, 2015, to December 31, 2015, in the validation cohort. A full list of ICD-9-CM codes of CKD was shown in Table m.1 in the 2015 Annual Data Report of the United State Renal Data System [3]. All cases of CKD, including those specified as stages 1 to 5 as well as those with unspecified stages, were included as study cases. The validity of ICD codes of CKD was reported in previous reports [3,31].

Figure 1. Flow chart of study. Study population was split into two parts based on time frames of electronic medical records (2013-2014 for derivation and 2014-2015 for validation).
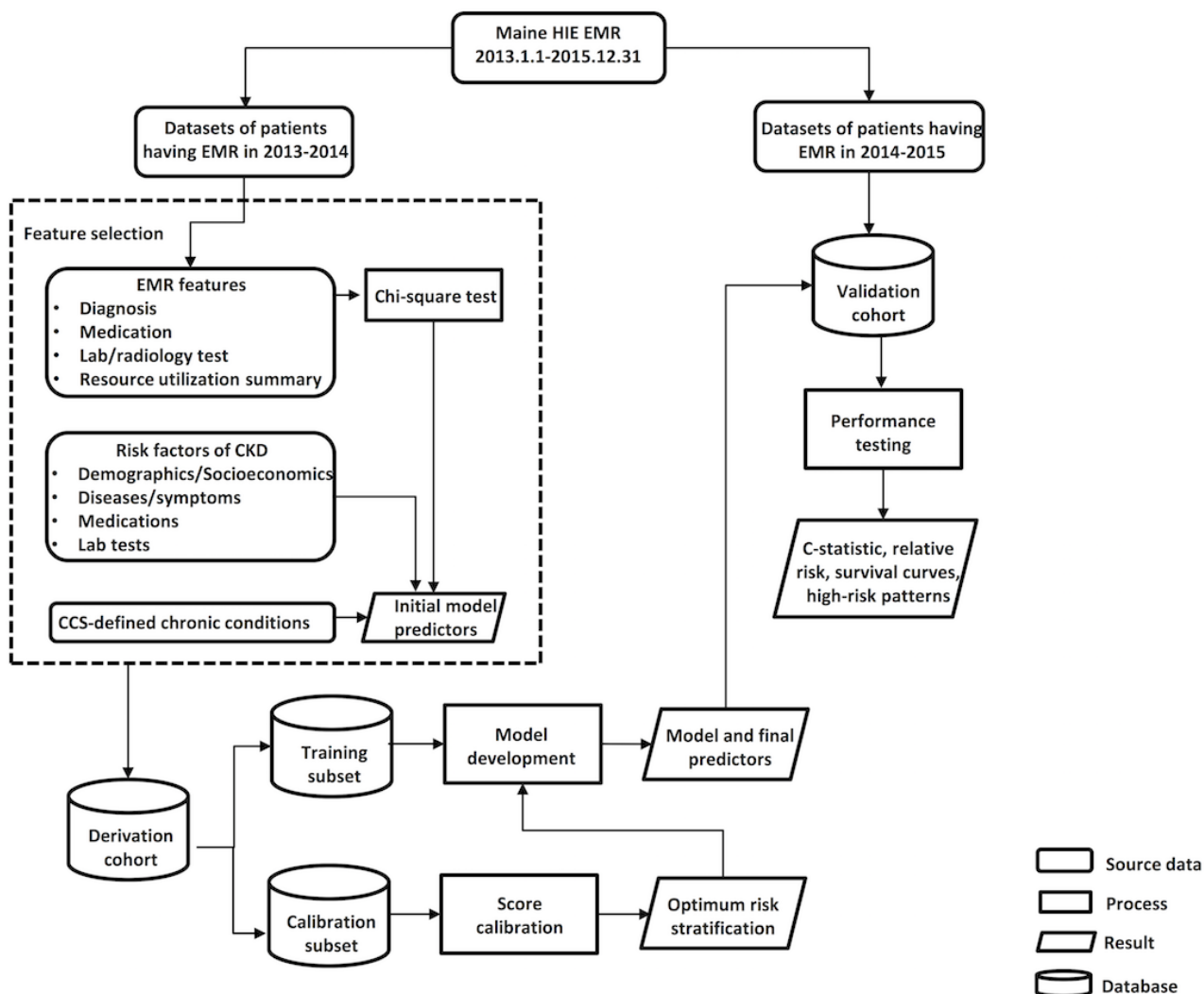
Figure 2. Formula of a tree ensemble model developed with the training subset.

$$y_i' = \sum_{k=1}^{K} f_k(x_i) \qquad i = 1, \ldots, n$$

Figure 3. Sum of the loss function and the overfitting control term at the t iteration.

$$L^t = \sum_{i=1}^{n} l(y_i, (y_i')^{t-1} + f_t(x_i)) + \Omega(f_t)$$

Feature Selection

A workflow chart for the study is shown in Figure 1. To improve computational efficiency, a feature selection process was carried out to determine the features that would go into the model prior to the derivation phase. The selection process was divided into 2 stages: literature review and variance analysis. Features recognized to have an association with CKD in previous literature were extracted as risk factors. These factors included demographics, chronic disease history, abnormal laboratory test results, and medication prescriptions. Chronic disease history included primary or secondary diagnosis. Medication prescriptions referred to the number of prescriptions for a particular medicine during the past 1 year. Laboratory test results were labeled as abnormal or normal according to thresholds provided by each facility participating in the HIE network. The rest of the features were screened by chi-square test to filter out those not significantly associated with CKD outcome ($P > .05$). The target of this process was to exclude features having low discriminatory power. For example, features that were 0, No, or NA in most of the patient records would probably be removed.

Derivation Phase

The derivation cohort was divided into 2 subsets for training and calibration purposes. An initial model was derived with the training subset. The model input was the selected features that profiled the preceding 1-year clinical history from January 1, 2013, to December 31, 2013, and the model output was set to either 1 or 0 to indicate whether or not a patient was diagnosed with CKD during the period from January 1, 2014, to December 31, 2014.

A gradient tree-based boosting algorithm was used to develop the model [32]. The idea of the algorithm is to approach the output by an ensemble of classification trees. Assume the training subset had n samples ($x_i, y_i, i=1,\ldots, n$), a tree ensemble model developed with the training subset can be written according to the formula in Figure 2, where $f(x)$ is the predictive function of a tree and K is the maximum number of trees in algorithm ($K=500$ in this study). Overfitting was avoided by adding a term to penalize the complexity of the algorithm. Parameters were chosen to minimize the sum of loss function and the overfitting control term. See Figure 3 for the sum term at the t iteration, where l is the loss function, $(y_i')^{t-1}$ is the predictive value at the t-1 iteration, and Ω is the term that controls overfitting. Ω is a function of the number of trees and weights of each tree in the algorithm.

An approximate greedy algorithm was used as a splitting method to grow trees. Features on each node were sorted to propose a couple of candidates at percentiles. Splitting points were chosen to optimize purity at the next level. In this study, the maximum depth of each tree was set to 5. Each node was assigned with an estimated value. The final predictive estimate was summed for individual trees.

A calibration subset was used to convert predictive estimates of the model developed with the training subset to a measure of positive predictive values (PPVs), which provided a universal, standardized risk measure. PPV for each predictive estimate y' was calculated as the proportion of incident CKD events in a subset of samples having predictive estimates higher than y'.

In this way, all the predictive estimates were mapped to the calculated PPVs. The PPVs were defined as scores that described the probability of having a new diagnosis of CKD within the next 1 year. We grouped all patients into 3 categories: low risk, intermediate risk, and high risk, based on the scores.

The scores after calibration were converted to relative risks. The relative risk of each individual was calculated by dividing the score of the individual by the mean score of all patients in the cohort. The relative risk measured the ratio of the probability of having CKD to the baseline. The higher the relative risk, the higher the probability of receiving a diagnosis of CKD in the next 1 year.

Validation Phase

A validation cohort of patients with clinical history from January 1, 2014, to December 31, 2014, was assembled to test the model performance on predicting the risk of CKD from January 1, 2015, to December 31, 2015. Predicted score and relative risk to the baseline were calculated for each patient. The c-statistic, relative risk distribution, and incidence of CKD diagnosis in each risk category were estimated to assess the performance of the model on the validation phase. The performance of the model was also evaluated in subgroups of patients using receiver operating characteristic (ROC) curves and c-statistics. Characteristics and clinical patterns of patients in each risk category were compared. Model errors were described by false positives (labeling a patient with no CKD in next 1 year as high risk) and false negatives (labeling a patient with CKD in next 1 year as low or intermediate risk), and clinical patterns of these patients were discussed.

Survival analysis was performed to track the timing of CKD diagnosis in different risk categories. Kaplan-Meier curves were plotted separately for each risk category to compare the probabilities of being diagnosed of CKD at the same time point. The analysis was not censored. A Kruskal-Wallis test was performed to compare the curves between the 3 risk categories. A temporal comparison of the CKD prediction date (ie, the time point when a high-risk patient was identified by the model) and CKD recognized date (ie, the time point when the patient was assigned ICD-9-CM diagnosis codes of CKD) was performed to evaluate the predictive power of the model in the time domain. All analyses were performed using R software (The R Foundation).

Results

Study Cohort

The final cohort included 1,310,363 patients for model derivation, 7448 of whom received a new CKD diagnosis in the next 1 year (from January 1, 2014, to December 31, 2014) and 1,430,772 patients for model validation, 8299 of whom had CKD diagnosed in the next 1 year (from January 1, 2015, to December 31, 2015). A cohort construction diagram is shown in Multimedia Appendix 3.

Table 1. Baseline characteristics.

Characteristic	Derivation cohort N=1,310,363 n (%)	Validation cohort N=1,430,772 n (%)
Age (years)		
≥65	269,355 (20.56)	299,893 (20.96)
50-65	288,645 (22.03)	312,456 (21.83)
40-50	163,792 (12.50)	172,877 (12.08)
<40	588,571 (44.92)	645,546 (45.12)
Female	690,714 (52.71)	748,867 (52.34)
Race		
White	1,090,046 (83.19)	1,194,478 (83.48)
Black	18,233 (1.39)	21,770 (1.52)
Asia	9,082 (0.69)	10,677 (0.75)
Other[a]/unknown[b]	193,002 (14.73)	203,847 (14.25)
Diabetes	54,366 (4.15)	60,631 (4.24)
Hypertension	121,413 (9.27)	133,328 (9.32)
Heart disease	49,684 (3.79)	52,780 (3.69)
Obesity	37,734 (2.88)	40,765 (2.85)

[a]Other refers to patients labeled as other race, multirace, or mixed.

[b]Unknown refers to patients labeled as unknown, undetermined, not applicable, or declined to answer.

Baseline Characteristics

The baseline characteristics of patients in derivation and validation cohorts are shown in Table 1. Both cohorts exhibited similar patterns of demographics and clinical conditions. The study involved patients of all ages and was gender balanced. In both cohorts, elderly patients (age ≥65 years) composed around 21% of the cohort, while young adults (<40 years) made up around 45% of the total; 18% of patients were pediatric (<18 years). The majority of patients were white. A history of diabetes or hypertension, 2 well-established risk factors of CKD, was present in approximately 4% and 9%, respectively, of the cohorts. Heart disease and obesity were present in almost 4% and 3%, respectively, in the cohorts.

Feature Selection

There are 14,680 features to profile each patient's clinical history in HIE dataset. The literature review identified a total of 153 clinical features as conventional risk factors of CKD, including 10 demographic features, 11 socioeconomic characteristics, 46 diagnostic diseases and conditions, 30 laboratory tests, and 56 medications. In parallel, 399 clinical features were selected after screening by chi-square test. These features, plus 184 chronic conditions identified by Clinical Classifications Software for classifying diagnoses and procedures into clinically meaningful categories (Healthcare Cost and Utilization Project, US Agency for Healthcare Research and Quality), constituted a set of 736 features for model derivation (Multimedia Appendix 4). The derivation process identified 146 features with non-zero weight as the final predictors of the model, including 6 demographic features, 2 socioeconomic characteristics, 36 diagnostic diseases and

conditions, 17 laboratory tests, 78 medication prescriptions, and 7 utilization variables (Multimedia Appendix 5). The top 50 features and their weights and odds ratios are listed in Multimedia Appendix 6. The following features played an important role in the model: age; history of diabetes, renal diseases, and heart diseases; history of diabetes and blood pressure medications; and health care resource utilization including length of stay in the hospital, total number of medications, and total number of laboratory tests with abnormal results.

Derivation Phase

We grouped all patients into 3 categories: low risk (score < 0.005), intermediate risk (score 0.005-0.05), and high risk (score

≥ 0.05). Model outcomes in the derivation phase are shown in Table 2. The model had a c-statistic of 0.916 in the derivation cohort. Patients diagnosed with CKD in the next 1 year (n=7448) had a median relative risk of 12.5, meaning that the model predicted these patients to have a probability of having CKD 12.5 times more than the baseline. Of these patients, 16.22% (1208/7448) were classified as low risk, 21.17% (1577/7448) as intermediate risk, and 62.61% (4663/7448) as high risk. The percentage of CKD cases and relative risk had a monotonic increase from low-risk (0.10%, 0.017) to high-risk categories (11.82%, 25.4).

Table 2. Comparison of the model outcome in derivation and validation cohorts.

Outcome	Derivation cohort N=1,310,363	Validation cohort N=1,430,772
Diagnosed with CKD[a] in the next 1 year, n (%)	7448 (0.57)	8299 (0.58)
Risk score model		
Baseline score, mean (SD)	0.0050 (0.034)	0.0044 (0.018)
Baseline score for those diagnosed with CKD in the next 1 year, median (1st quartile, 3rd quartile)	0.063 (0.013, 0.29)	0.049 (0.0079, 0.092)
Relative risk[b] for those diagnosed with CKD in the next 1 year, median (1st quartile, 3rd quartile)	12.5 (2.6, 57.3)	11.1 (1.8, 21.0)
CKD diagnosis by risk category: low/intermediate/high	1208/1577/4663	1778/2334/4177
Percent incidence of CKD diagnosis (95% CI)		
Low (score 0-0.005)	0.10 (0-0.30)	0.14 (0-0.45)
Intermediate (score 0.005-0.05)	1.73 (1.15-2.60)	2.10 (1.10-2.90)
High (score ≥ 0.05)	11.82 (10.10-13.80)	7.94 (6.50-10.10)
Relative risk to the population baseline (95% CI)		
Low (score 0-0.005)	0.017 (0.012-0.023)	0.011 (0.0067-0.017)
Intermediate (score 0.005-0.05)	3.2 (3.0-3.3)	4.1 (3.9-4.2)
High (score ≥ 0.05)	25.4 (23.9-27.2)	18.3 (17.8-19.0)

[a]CKD: chronic kidney disease.

[a]Relative risk of each patient was defined as the ratio of the risk score of the patient to the baseline score (ie, the mean risk score of total population).

Validation Phase

The performance of the model was slightly reduced in the validation cohort, with a c-statistic of 0.871, but had similar results (Table 2). The median relative risk of patients diagnosed with CKD in the next 1 year (n=8299) was 11.1, and 50.33% (4177/8299) of these patients were labeled as high risk. The total numbers of low-, intermediate-, and high-risk patients were 1,266,893, 111,195, and 52,594, respectively, 0.14%, 2.10%, and 7.94%, respectively, of whom had a diagnosis of CKD within the next 1 year.

The cutoff of the high-risk patients (score ≥ 0.05) gave a sensitivity of 62.61% (95% CI 61.50%-63.71%) and a specificity of 97.33% (95% CI 97.30%-97.36%) in the derivation cohort

and sensitivity of 50.33% (95% CI 49.25%-51.41%) and a specificity of 96.60% (95% CI 96.57%-96.63%) in the validation cohort. A 2-by-2 contingency table is shown in Multimedia Appendix 7. A list of sensitivities, specificities, and PPVs as a function of cutoffs is shown in Multimedia Appendix 8. Reduction of the specificity from 96.60% to 87.88% will increase the sensitivity from 50.33% to 80.33%, but it will also reduce PPV from 7.94% to 3.72%.

As shown in Figure 4, the model had effective discriminatory power within patient subgroups. C-statistics for patients with no chronic disease history (741,703/1,430,772, 51.84%), those aged ≥65 years (280,787/1,430,772, 19.62%), and those <65 years (1,149,985/1,430,772, 80.38%) were 0.804, 0.819, and 0.734, respectively.

Figure 4. Receiver operating characteristic curves and c-statistics of the model prediction.
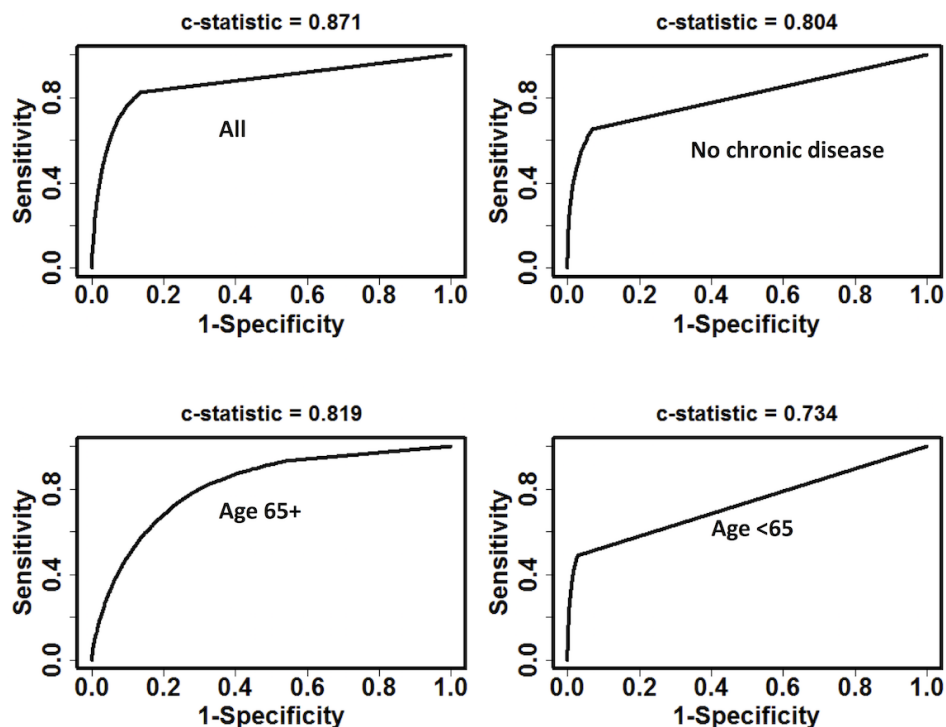
Table 3. Clinical patterns of patients by risk categories in the validation cohort.

Characteristic	Low risk N=1,266,983	Intermediate risk N=111,195	High risk N=52,594
Age, years, median (1st quartile, 3rd quartile)	39 (20, 56)	75 (68, 82)	79 (71, 85)
Female, n (%)	667,440 (52.68)	55,717 (50.11)	25,710 (48.88)
Race, n (%)			
White	1,031,954 (81.45)	110,303 (99.20)	52,221 (99.29)
Black	21,151 (1.67)	424 (0.38)	195 (0.37)
Asian	10,332 (0.82)	253 (0.23)	92 (0.17)
Other/unknown	203,546 (16.07)	215 (0.19)	86 (0.16)
Diabetes, n (%)	22,025 (1.74)	19,271 (17.33)	19,335 (36.76)
Hypertension, n (%)	60,970 (4.81)	39,564 (35.58)	32,794 (62.35)
Heart disease, n (%)	17,388 (1.37)	16,156 (14.53)	19,236 (36.57)
Obesity, n (%)	29,308 (2.31)	6686 (6.01)	4771 (9.07)
Blood pressure medication, n (%)	64,974 (5.13)	42,096 (37.86)	34,183 (64.99)
Diabetes medication, n (%)	26,533 (2.09)	17,045 (15.33)	15,553 (29.57)
Abnormal diabetes test, n (%)	575 (0.05)	388 (0.35)	618 (1.18)
Abnormal urine albumin-to-creatinine ratio, n (%)	155 (0.01)	90 (0.08)	171 (0.33)
Total costs, median (1st quartile, 3rd quartile)	170 (0, 925)	850 (170, 2455)	1700 (510, 4530)
Outpatient visits, median (1st quartile, 3rd quartile)	1 (0, 3)	4 (1, 8)	8 (4, 15)
Total counts of medications, median (1st quartile, 3rd quartile)	0 (0, 3)	7 (0, 31)	32 (7, 75)
Total counts of laboratory tests, median (1st quartile, 3rd quartile)	0 (0, 0)	0 (0, 29)	6 (0, 81)

Figure 5. Distribution of false positive patients (top) and false negative patients (bottom) in validation cohort.
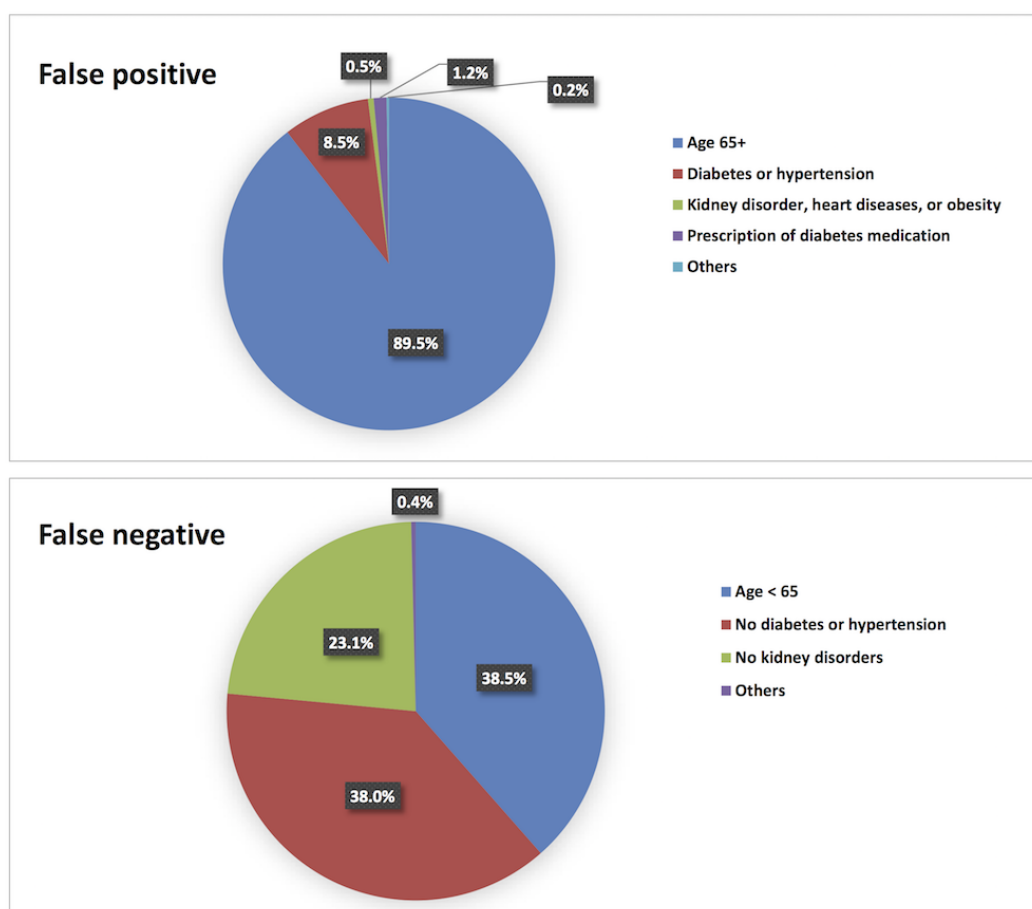
Clinical patterns were compared among the low-, intermediate-, and high-risk categories in the validation cohort (Table 3). There was a significant difference (*P*<.001) in age distribution between low- and high-risk patients: 99.86% (644,610/645,546) of young adults (age <40 years) were classified as low risk, while 87.80% (46,175/52,594) of the high-risk patients were ≥65 years of age. Patients in the high-risk category featured more serious comorbidities and more resource consumption. Among high-risk patients, a history of diabetes, hypertension, heart diseases, and obesity was present in 36.76%, 62.35%, 36.57%, and 9.07%, respectively, much higher than in the overall cohort (4.24%, 9.32%, 3.69%, and 2.85%). High-risk patients also utilized the largest amount of resources in terms of total number of outpatient visits (median 8), medications (median 32), and laboratory tests (median 6) over the last 1 year, resulting in the highest annual costs (median $1700) among all 3 risk categories. The model tends to aggregate heavy users of health care resources and those with traditional risk factors of CKD (age, diabetes, and hypertension) into the high-risk category.

False Positives and False Negatives

The distribution of false positives and false negatives in the validation cohort is shown in Figure 5. Of false positives, 89.53% (43,346/48,417) were patients ≥65 years of age; 8.45% (4092/48,417) were <65 years but with diagnosis of hypertension and/or diabetes; 0.55% (265/48,417) did not have diabetes or

hypertension but had kidney disorders, heart diseases, or obesity; and 1.25% (603/48,417) were prescribed medications for diabetes or hypertension. Of the other 111 false positive patients, 110 had at least 1 medication prescription or 1 abnormal laboratory test result during the preceding year. Conversely, there were 4122 false negatives, patients with CKD in next 1 year who were missed by the model. Among these, 38.50% (1587/4122) were <65 years of age, 37.99% (1566/4122) were ≥65 years old but had no history of diabetes or hypertension, and 23.12% (953/4122) had diabetes or hypertension but no kidney disorder.

Temporal Analysis

Kaplan-Meier analysis was performed to estimate freedom from a new CKD diagnosis for patients in the 3 risk categories in the validation cohort (Figure 6). Significant differences (*P*<.001) were demonstrated between the risk categories.

Among high-risk patients in the validation cohort, 4177/52,594 received a new CKD diagnosis within the next 1 year. Figure 7 shows the distribution of the time intervals from the time point when a patient was identified as high risk by the model and the time point when the patient was assigned an ICD-9-CM CKD diagnosis code. Nearly half (48.24%) of the CKD cases were marked as high risk 6 months or more prior to assignment of a diagnosis code (ie, confirmatory diagnosis was made by physician).

Figure 6. Kaplan-Meier progression to chronic kidney disease for patients in low-, intermediate-, and high-risk categories of validation cohort.
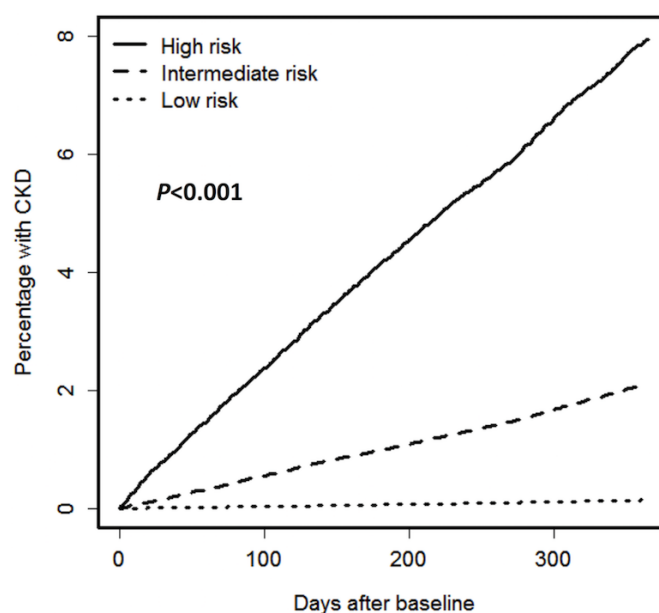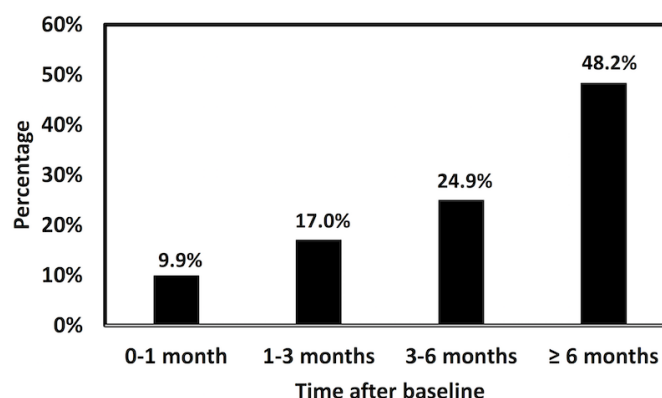

Figure 7. Distribution of high-risk patients in the validation cohort by time intervals between the model identification and coded chronic kidney disease diagnosis by International Classification of Disease, Ninth Revision, Clinical Modification.

Discussion

Principal Findings

We have derived and validated a risk model predictive of incident CKD diagnosis within the next 1 year across 1.3 million patients in the state of Maine. Through machine learning from preceding 1-year clinical profiles that were recorded in EMR databases, patients were classified into 3 risk categories (low, intermediate, and high risk), indicating the stratified probabilities of receiving CKD-related ICD-9-CM codes in the next 1 year. The model had similar performance in the derivation phase (c-statistic of 0.916) and validation phase (c-statistic of 0.871). Compared with other simplified score metrics [19,21,33], the model uses more predictors, giving a better result in classification (Multimedia Appendix 9). Performance of the model in subcohorts, especially those considered low risk by traditional risk factors (ie, age <65 years, no history of chronic disease) was fairly good (c-statistics 0.734 and 0.804, respectively), showing predictive power in patients with low awareness of CKD that traditional models tend to ignore. Model

outcomes (Table 2) and survival analysis (Figure 6) both showed the model to provide reasonable risk stratification.

We applied a 2-step feature reduction process; 736 features survived after the first step (filtered by $P<.05$ plus literature review), and 146 features survived after the second step (filtered by non-zero weight in algorithm). Features having smaller P value in the chi-squared screening might not have larger weight in the algorithm due to the different mechanisms of establishing the relationship between the outcome and the features in the 2 steps of feature reduction. With this consideration, we set up $P<.05$ as a threshold to enable more features that might contribute to the modeling to go into the next step.

Results of misclassification analysis (Figure 5) show that 97.98% of false positives were patients who were ≥65 years of age or had a history of diabetes or hypertension. These patients, although they did not receive a CKD diagnosis within the next 1 year, were still considered at higher risk for developing CKD or other adverse outcomes than the general population. Monitoring these patients would help identify signs of CKD at an early stage and may benefit their long-term outcomes. Among false negatives, 99.61% were patients who lacked one or more

major risk factor (eg, patients <65 years old or without a history of diabetes, hypertension, or kidney disorders), causing the model to identify them as low or intermediate risk.

Our model identified patients in other or unknown race categories as less likely to get CKD. The percentages of patients in the other race category were 16.07% (203,546/1,266,983) in the low-risk category and 0.16% (86/52,594) in the high-risk category in the validation cohort. Most patients (90.13%, 183,723/203,847) in the other or unknown race category actually had an unknown race marked in the dataset. It perhaps indicated a data quality issue that the race information was probably missing. Compared to the total studied population, patients in the unknown race category had a much lower rates of history of diabetes (0.16% vs 4.24%) and hypertension (0.38% vs 9.32%), and fewer outpatient visits (5.59% vs 56.04%). Lack of risk factors of CKD made the majority of these patients stratified to the low-risk group. However, such results didn't mean these patients were healthier than the average level of the total population. As described in the Multimedia Appendix 2, for a patient who didn't have any EMR, it is hard to tell whether this patient was healthy thus had never used care service or this patient did use care service but the records were missing. Such limitation was caused by the EMR storage format.

A main challenge of this study was that actual values of estimated glomerular filtration rate (eGFR) and albumin-to-creatinine ratio (ACR), the 2 parameters used to determine CKD stage [34,35], were not available in our data source. The total counts of abnormal creatinine blood test results and ACR over the preceding 1-year period were used instead. Moreover, as it was a study on the general population, most of the participants did not have abnormal test results related to eGFR or ACR. Therefore, unlike other studies in which eGFR and ACR played critical roles in CKD prediction, these parameters were not selected as top features by the model proposed in this study. The model, however, had performance comparable to studies using exact values of eGFR and ACR as predictors [21,22], indicating that CKD incidence can be predicted without knowledge of eGFR or ACR. These results support the potential value of EMR- or claims-based retrospective studies in which actual laboratory test results tend to be missing due to data quality issues or data sharing policies. An analysis of Medicare patients showed that even among patients older than 65 years, a group at high risk for CKD, less than 80% of patients had claims indicating serum creatinine testing and less than 20% had urine albumin testing [3]. Development and validation of a CKD risk model within a general population in which eGFR and ACR are frequently absent is extremely useful for its applicability in clinical practice as a routinely used assistant tool. It makes our model an economically feasible method for general population screening because it eliminates the time and costs of collecting eGFR and ACR during traditional screening tests of CKD [36-38]. A prescreening on general population using the proposed model followed by tests of urine albumin and serum creatinine on high-risk patients forms a cost-effective approach to identify risks of CKD.

Another challenge was that this study targeted prediction of CKD incidence within the next 1 year, which is a short time horizon compared with other studies of CKD prediction in which the follow-up periods were several years [21-23]. Such a short time frame resulted in a low incidence (0.57% in the derivation cohort and 0.58% in validation cohort), which increased the difficulty of prediction. The complex model with multiple predictors allowed identification of a group of patients with a high 1-year incidence of CKD (derivation phase 11.82%, 20.7 times higher than the baseline; validation phase 7.94%, 13.7 times higher than the baseline). These patients were labeled as high risk and are good targets for administration and intervention plans. Traditional risk factors (age, history of diabetes and hypertension) identified a group of patients with a 1-year incidence of 1.95% in the derivation phase and 1.97% in the validation phase, only about 3 times higher than the baseline.

Interpretation of Predictors

The feature selection process that combined both data-driven methodology and domain knowledge resulted in a list of predictors composing the predictive algorithm (Multimedia Appendix 5). Traditional risk factors of CKD remained highly important. Age and the use of furosemide were 2 predictors of top importance. This observation makes sense, as age is considered a common risk factor of CKD, while furosemide is a medication used in patients with congestive heart disease, kidney disorders, and high blood pressure, all of which are correlated with CKD. The link between cardiovascular diseases and CKD has been reported in many studies, and the role of cardiovascular diseases in the development and progression of CKD was found [39,40]. CKD was found in over half of patients with heart failure [41]. CKD and cardiovascular diseases share common risk factors, and a bidirectional pathway was noticed between the progression of cardiovascular disease and CKD [39]. Medical history of furosemide, which is commonly used to treat congestive heart failure, therefore may indicate a risk of CKD initiation. What's more, furosemide is a commonly used preventive and therapeutic drug for acute kidney injury (AKI) [42]. The benefits of furosemide in reducing hypertension and improving eGFR show its potential role in reducing the risk of AKI. Compared with other diuretics for kidney diseases such as bumetanide, hydrochlorothiazide, and spironolactone that were predictors of our model, furosemide is more powerful and less expensive. The biological link between AKI and CKD has been established, and AKI is considered as an independent risk factor of CKD.

In addition, the model identified a group of previously prescribed medications as predictors, primarily drugs for diabetes (insulin glargine, insulin isophane, glipizide, insulin fetemir, etc.), blood pressure control (hydralazine, amlodipine besylate, metoprolol tartrate, etc.), heart diseases (isosorbide mononitrate, valsartan, amiodarone, etc.), and kidney disorders (allopurinol). Such medication histories indicate patients either at risk for or living with diseases that might lead to CKD. Prescriptions for medications used for inflammatory processes (prednisone and colchicine), bone disease (febuxostat), anemia (folic acid), and hypokalemia (potassium chloride) were also identified as predictors, illustrating their contribution to the disease network. Abnormal results of metabolic panel, glucose test, coagulation test, and therapeutic drug monitoring were predictors in the laboratory test category, which indicates disease

states such as diabetes. History of hypertension, renal disorders, heart diseases, anemia, and diabetes were top important diagnostic features that were highly correlated to CKD.

In addition to the clinical features, variables indicative of high resource consumption (eg, health care costs, total counts of medications, laboratory and radiology tests, outpatient visits, and inpatient length of stay) were also considered risk factors by the model. This pattern identifies heavy users of health care services (eg, older patients or patients with multiple chronic morbidities) to have a higher probability of developing CKD, which makes sense as CKD has been considered as a complication of complex chronic diseases that are associated with large health resource expenditures [3].

In all, senior patients and heavy users of care resources with chronic conditions like diabetes and hypertension that are highly correlated to CKD tend to be classified as high risk for incident CKD by the model.

Beyond Risk Estimation

Several previous studies have reported the development and validation of CKD risk scores. The predictors, modeling process, validation, and accuracy of the scores were well presented, but little effort was made to translate the risk scores to patient care action plans. Those studies addressed whether the risk of CKD onset or progression can be predicted but did not address what actions should be taken for high-risk patients [43]. The widespread application of EMR in the state of Maine has enabled us to develop risk scores for the Maine residents [27,28,44-46] in terms of future resource utilization and clinical conditions. The meaningful use of EMR data, however, is not only to forecast the health status in the future but also to guide the health care providers to make decisions in the present. There are already established guidelines in CKD preventive care to address both nonmodifiable and modifiable risk factors. For example, CKD screening is recommended on a regular basis for patients with nonmodifiable risk factors (eg, older patients) to identify CKD at an early stage.

For patients with modifiable risk factors such as concurrent chronic conditions, life styles, and medications, there are quite a few targeted intervention options to reduce the risk. Nutritional treatments such as a low-protein diet together with sufficient and regular exercise should be initiated on patients with obesity, hypertension, or diabetes to prevent or slow CKD progression [47,48]. Medications that may reduce renal function or cause complications, such as angiotensin-converting enzyme and nonsteroidal anti-inflammatory drugs, should be prescribed with careful consideration and monitoring plans if necessary [49,50]. Advice to stop smoking and limit alcohol should be given to smokers and alcohol users to improve overall health and reduce the risk of CKD [51,52] for those individuals. The modifiable risk factors are even more important than nonmodifiable predictors as they offer an opportunity to both clinicians and patients to be proactive to the disease by implementing interventions before deterioration.

In all, a combination of a single scalar score and longitudinal clinical profile including chronic disease history, current problem list, and therapies and medications will help clinicians develop a personalized action plan with modifiable risk factors for each high-risk individual. It is the subsequent actions rather than an isolated risk score that help improve health status, outcomes, and resource utilization [53]. The ultimate goal of this study is to confirm, modify, or disapprove care plans based on the risk prediction outcomes, leading to improved quality of care. Obtaining a risk score is not the end of the study but the first step of translating predictive analytics into prescriptive solutions, a proactive approach to prevent or delay deterioration in health.

Implications for Treatment and Prognosis

A chart showing time intervals between identification of high-risk patients and receiving a CKD diagnosis code in Figure 7 reveals clinical implications for treatment and prognosis of CKD. Certain interventions at an early stage can reduce the risk of developing CKD or progression to end-stage disease. For example, clinical trials showed that patients receiving blood pressure control treatment had significantly reduced proteinuria within the first 4 months compared with those had no blood pressure control, suggesting a reduced risk of CKD development and progression [54,55]. A meta-analysis reported that lifestyle modifications for 3 months decreased the risk for diabetes from the end of intervention up to 10 years later [56], which in turn correlated to attenuated risk of developing CKD, as diabetes has been recognized as an important predisposing factor for CKD.

In our validation cohort, the model identified 72.90% (3045/4177) of high-risk patients at least 3 months before the confirmatory diagnosis was made by physicians. Of those patients, 41.02% (1249/3045) had diabetes or an abnormal glucose test result at the time they were identified by the model to be at high risk for CKD. Implementation of lifestyle modifications at that time has the potential to mitigate adverse outcomes in those patients. Moreover, 64.59% (2698/4177) of high-risk patients were identified by the model at least 4 months prior to confirmatory diagnosis, and 9.82% (265/2698) of those patients were not taking any blood pressure medication and did not have a diagnosis of hypertension. Blood pressure monitoring and necessary control in these patients can help to reduce the risk of CKD. Such explorations highlight potential meaningful use of the model in clinical practice, in that it can help to initiate decision making and timely intervention.

The predictive model and risk scores can benefit health care organizations at multiple levels. For health care managers who take charge of the population management at the whole department or hospital, the population stratification by risk scores will help with budget planning, as high-risk patients tend to require more resources. For physicians, the model can be used as an assistant tool for decision making. High-risk patients without eGFR or ACR parameters available can be referred to the CKD screening test to decide whether or not the patients have CKD already. The risk stratification will also give physicians ideas of treating patients at high risk of CKD for other concurrent clinical problems, especially in the situation where the current medical or surgical treatments can help with the existing problems but accelerate CKD progression in patients. Clinicians can also drill down to see what information

is driving the risk scores, which provides the clinical background they need to trust and act on the risk scores.

Study Limitations

There are several limitations of this study. First, uncoded CKD cases could be outliers of the model and affect accuracy. Patients with undiagnosed CKD might not be excluded from the study cohort. Computed false positives from undiagnosed CKD patients missing diagnosis codes are actually true positives, especially for those who were over age 65 years. eGFRs of older patients tend to be lower and thus may further complicate the diagnosis. Patients who were waiting a random urine test for confirmative diagnosis but didn't have an ICD-9-CM code assigned during the study period could confound the model as well. Maine HIE went live in 2009, so patients with CKD diagnosed before 2009 might not be documented in the EMR database. These patients could be treated as false positives during the performance evaluation. Second, there might be a delay of the assignment of an ICD-9-CM code that was longer than the transition of kidney function from a normal state to a disease state. It might explain why the incidence rate (0.568%-0.580%) in our study cohort was higher than reported by other studies [57,58], as some of the patients who received an ICD-9-CM code might already have undiagnosed stage 1 to 2 CKD. Assignment of an ICD-9-CM code doesn't always mean a new case showing up. Another possible reason for the high incidence was that the study cohort had a slight age bias (20.96% in the validation cohort vs 15.64% in the overall population in Maine for percentage of patients at 65 years and over), and age is an independent risk factor of CKD. Third, unlike other CKD risk models, our model does not include exact values of eGFR or ACR as predictors due to lack of such data, and it is possible that including eGFR could further improve the model performance. Fourth, all laboratory test variables were labeled as either normal or abnormal in the data source. A detailed classification of laboratory test results would help to construct a deeper understanding of clinical conditions of patients and enhance model performance. Fifth, due to the nature of EMR storage format, we cannot differentiate the situation where a particular medical record was missing, although it would happen at a very low probability. Sixth, the cutoff point (score ≥ 0.05) for high-risk categories was selected to optimize the PPV with a fair value of sensitivity. In the production dashboard we deployed at the Maine HIE, there is an option to allow each provider user to set up its own cutoffs on our real-time population health care surveillance platform. Seventh, all the study participants were from the state of Maine, and recalibration as well as other necessary adjustments would be needed before leveraging the model to health care management for populations in other states. Geographical, environmental, and racial disparities may contribute to population characteristics, and additional risk factors should be considered if necessary.

Conclusions

A risk model that estimated the probability of receiving CKD diagnosis within the next 1 year was developed and validated in this study. Through the statistical learning of the EMRs of over 1.3 million patients in the state of Maine, the model was able to assign each individual a risk score based on the preceding 1-year clinical history. The whole population was stratified into 3 risk categories according to the score, where the high-risk category had a CKD incidence 13.7 times higher than the baseline. A *c*-statistic of 0.871 was achieved in the validation phase. Identification of patients at high risk of receiving CKD diagnosis will help to promote care plans of monitoring and intervention, which will ultimately benefit the outcomes of patients.

Authors' Contributions

SH, TF (Tianyun Fu), QW, BJ, CZ, ZH, YG, YZ, and YY carried out the initial analysis and interpretation of data and drafted the initial manuscript. FS, KGS, EW, DBM, and XBL conceptualized and designed the study and critically reviewed and revised the manuscript. TF (Terry Fouts) and PN contributed to the interpretation of data and critically reviewed and revised the manuscript. DSC and STA coordinated and supervised data acquisition and critically reviewed and revised the manuscript. All authors have read and approved this submission for publication. All authors have agreed to be accountable for all aspects of the work.

Conflicts of Interest

The authors have the following interests: KGS, EW, and XBL are cofounders and equity holders of HBI Solutions, Inc, which is currently developing predictive analytics solutions for health care organizations. BJ, CZ, FS, and EW are employed by HBI Solutions, Inc. From the departments of cardiothoracic surgery and surgery, Stanford University School of Medicine, Stanford, California, KGS and XBL conducted this research as part of a personal outside consulting arrangement with HBI Solutions, Inc. The research and research results are not in any way associated with Stanford University. There are no patents, further products in development, or marketed products to declare. This does not alter our adherence to all the journal policies on sharing data and materials as detailed online in the guide for authors.

Multimedia Appendix 1

Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist for reporting derivation and validation predictive model.

[PDF File (Adobe PDF File), 43KB - medinform_v5i3e21_app1.pdf]

Multimedia Appendix 2

Missing data handling.

[[PDF File (Adobe PDF File), 11KB](#) - [medinform_v5i3e21_app2.pdf](#)]

Multimedia Appendix 3

Construction of derivation and validation cohorts.

[[PDF File (Adobe PDF File), 24KB](#) - [medinform_v5i3e21_app3.pdf](#)]

Multimedia Appendix 4

List of 736 features considered for model derivation.

[[PDF File (Adobe PDF File), 59KB](#) - [medinform_v5i3e21_app4.pdf](#)]

Multimedia Appendix 5

List of 146 features (final predictors) and their weights in the model.

[[PDF File (Adobe PDF File), 28KB](#) - [medinform_v5i3e21_app5.pdf](#)]

Multimedia Appendix 6

Top 50 features (ie, predictors) in the final model: weight, log odds ratio, and 0.95 confidence interval.

[[PDF File (Adobe PDF File), 25KB](#) - [medinform_v5i3e21_app6.pdf](#)]

Multimedia Appendix 7

A 2-by-2 contingency table for the derivation cohort and the validation cohort with a cutoff score of 0.05.

[[PDF File (Adobe PDF File), 22KB](#) - [medinform_v5i3e21_app7.pdf](#)]

Multimedia Appendix 8

Relationships between sensitivities, specificities, and positive predictive values of the model on the validation cohort.

[[PDF File (Adobe PDF File), 170KB](#) - [medinform_v5i3e21_app8.pdf](#)]

Multimedia Appendix 9

Comparison of c-statistics using other clinical prediction scores.

[[PDF File (Adobe PDF File), 30KB](#) - [medinform_v5i3e21_app9.pdf](#)]

References

1. Centers for Disease Control and Prevention. Chronic Kidney Disease Surveillance System—United States. URL: http://www.cdc.gov/ckd [accessed 2017-07-13] [WebCite Cache ID 6q8p85gyg]
2. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, et al. Prevalence of chronic kidney disease in the United States. JAMA 2007 Nov 07;298(17):2038-2047. [doi: 10.1001/jama.298.17.2038] [Medline: 17986697]
3. Saran R, Li Y, Robinson B, Abbott KC, Agodoa LYC, Ayanian J, et al. US Renal Data System 2015 annual data report: epidemiology of kidney disease in the United States. Am J Kidney Dis 2016 Mar;67(3 Suppl 1):S1-S305. [doi: 10.1053/j.ajkd.2015.12.014] [Medline: 26925525]
4. Honeycutt AA, Segel JE, Zhuo X, Hoerger TJ, Imai K, Williams D. Medical costs of CKD in the Medicare population. J Am Soc Nephrol 2013 Sep;24(9):1478-1483 [FREE Full text] [doi: 10.1681/ASN.2012040392] [Medline: 23907508]
5. Centers for Disease Control and Prevention. National chronic kidney disease fact sheet: general information and national estimates on chronic kidney disease in the United States. 2014. URL: https://www.cdc.gov/diabetes/pubs/pdf/kidney_factsheet.pdf [accessed 2017-07-13] [WebCite Cache ID 6rwNuJWIN]
6. Roderick P, Jones C, Drey N, Blakeley S, Webster P, Goddard J, et al. Late referral for end-stage renal disease: a region-wide survey in the south west of England. Nephrol Dial Transplant 2002 Jul;17(7):1252-1259. [Medline: 12105249]
7. Lin C, Chuang F, Wu C, Yang C. Early referral as an independent predictor of clinical outcome in end-stage renal disease on hemodialysis and continuous ambulatory peritoneal dialysis. Ren Fail 2004 Sep;26(5):531-537. [Medline: 15526911]

8.   Kinchen KS, Sadler J, Fink N, Brookmeyer R, Klag MJ, Levey AS, et al. The timing of specialist evaluation in chronic kidney disease and mortality. Ann Intern Med 2002 Sep 17;137(6):479-486. [Medline: 12230348]

9.   Jungers P, Massy ZA, Nguyen-Khoa T, Choukroun G, Robino C, Fakhouri F, et al. Longer duration of predialysis nephrological care is associated with improved long-term survival of dialysis patients. Nephrol Dial Transplant 2001 Dec;16(12):2357-2364. [Medline: 11733627]

10.   Plantinga LC, Boulware LE, Coresh J, Stevens LA, Miller ER, Saran R, et al. Patient awareness of chronic kidney disease: trends and predictors. Arch Intern Med 2008 Nov 10;168(20):2268-2275 [FREE Full text] [doi: 10.1001/archinte.168.20.2268] [Medline: 19001205]

11.   Guessous I, McClellan W, Vupputuri S, Wasse H. Low documentation of chronic kidney disease among high-risk patients in a managed care population: a retrospective cohort study. BMC Nephrol 2009 Sep 16;10:25 [FREE Full text] [doi: 10.1186/1471-2369-10-25] [Medline: 19758452]

12.   Akbari A, Swedko PJ, Clark HD, Hogg W, Lemelin J, Magner P, et al. Detection of chronic kidney disease with laboratory reporting of estimated glomerular filtration rate and an educational program. Arch Intern Med 2004 Sep 13;164(16):1788-1792. [doi: 10.1001/archinte.164.16.1788] [Medline: 15364673]

13.   Agrawal V, Ghosh AK, Barnes MA, McCullough PA. Perception of indications for nephrology referral among internal medicine residents: a national online survey. Clin J Am Soc Nephrol 2009 Feb;4(2):323-328 [FREE Full text] [doi: 10.2215/CJN.03510708] [Medline: 19218472]

14.   Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, et al. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension 2003 Dec;42(6):1206-1252 [FREE Full text] [doi: 10.1161/01.HYP.0000107251.49515.c2] [Medline: 14656957]

15.   American Diabetes Association. Executive summary: standards of medical care in diabetes—2012. Diabetes Care 2012 Jan;35 Suppl 1:S4-S10 [FREE Full text] [doi: 10.2337/dc12-s004] [Medline: 22187471]

16.   Su S, Lin C, Kao S, Wu C, Lu K, Lai C, et al. Risk factors and their interaction on chronic kidney disease: a multi-centre case control study in Taiwan. BMC Nephrol 2015 Jun 16;16:83 [FREE Full text] [doi: 10.1186/s12882-015-0065-x] [Medline: 26077152]

17.   Levin A. Identification of patients and risk factors in chronic kidney disease—evaluating risk factors and therapeutic strategies. Nephrol Dial Transplant 2001;16 Suppl 7:57-60. [Medline: 11590259]

18.   Kazancıoğlu R. Risk factors for chronic kidney disease: an update. Kidney Int Suppl (2011) 2013 Dec;3(4):368-371 [FREE Full text] [doi: 10.1038/kisup.2013.79] [Medline: 25019021]

19.   Thakkinstian A, Ingsathit A, Chaiprasert A, Rattanasiri S, Sangthawan P, Gojaseni P, et al. A simplified clinical prediction score of chronic kidney disease: a cross-sectional-survey study. BMC Nephrol 2011 Sep 26;12:45 [FREE Full text] [doi: 10.1186/1471-2369-12-45] [Medline: 21943205]

20.   Taal MW. Predicting renal risk in the general population: do we have the right formula? Clin J Am Soc Nephrol 2011 Jul;6(7):1523-1525 [FREE Full text] [doi: 10.2215/CJN.04200511] [Medline: 21700818]

21.   O'Seaghdha CM, Lyass A, Massaro JM, Meigs JB, Coresh J, D'Agostino RB, et al. A risk score for chronic kidney disease in the general population. Am J Med 2012 Mar;125(3):270-277 [FREE Full text] [doi: 10.1016/j.amjmed.2011.09.009] [Medline: 22340925]

22.   Mocroft A, Lundgren JD, Ross M, Law M, Reiss P, Kirk O, et al. Development and validation of a risk score for chronic kidney disease in HIV infection using prospective cohort data from the D:A:D study. PLoS Med 2015 Mar;12(3):e1001809 [FREE Full text] [doi: 10.1371/journal.pmed.1001809] [Medline: 25826420]

23.   Hippisley-Cox J, Coupland C. Predicting the risk of chronic kidney disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. BMC Fam Pract 2010 Jun 21;11:49 [FREE Full text] [doi: 10.1186/1471-2296-11-49] [Medline: 20565929]

24.   Mani S, Chen Y, Elasy T, Clayton W, Denny J. Type 2 diabetes risk forecasting from EMR data using machine learning. AMIA Annu Symp Proc 2012;2012:606-615 [FREE Full text] [Medline: 23304333]

25.   Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. J Am Med Inform Assoc 2009;16(3):371-379 [FREE Full text] [doi: 10.1197/jamia.M2846] [Medline: 19261943]

26.   Amarasingham R, Velasco F, Xie B, Clark C, Ma Y, Zhang S, et al. Electronic medical record-based multicondition models to predict the risk of 30 day readmission or death among adult medicine patients: validation and comparison to existing models. BMC Med Inform Decis Mak 2015 May 20;15:39 [FREE Full text] [doi: 10.1186/s12911-015-0162-6] [Medline: 25991003]

27.   Hao S, Jin B, Shin AY, Zhao Y, Zhu C, Li Z, et al. Risk prediction of emergency department revisit 30 days post discharge: a prospective study. PLoS One 2014;9(11):e112944 [FREE Full text] [doi: 10.1371/journal.pone.0112944] [Medline: 25393305]

28.   Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based real-time case finding for the population health management of patients with diabetes mellitus: a prospective validation of the natural language processing-based algorithm with statewide electronic medical records. JMIR Med Inform 2016 Nov 11;4(4):e37 [FREE Full text] [doi: 10.2196/medinform.6328] [Medline: 27836816]

29. HealthInfoNet. URL: http://hinfonet.org [accessed 2017-05-01] [WebCite Cache ID 6q8so13wR]

30. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 07;350:g7594 [FREE Full text] [Medline: 25569120]

31. Fleet JL, Dixon SN, Shariff SZ, Quinn RR, Nash DM, Harel Z, et al. Detecting chronic kidney disease in population-based administrative databases using an algorithm of hospital encounter and physician claim codes. BMC Nephrol 2013 Apr 05;14:81 [FREE Full text] [doi: 10.1186/1471-2369-14-81] [Medline: 23560464]

32. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2939785: ACM; 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco p. 785-794.

33. Hemmelgarn BR, Culleton BF, Ghali WA. Derivation and validation of a clinical index for prediction of rapid progression of kidney dysfunction. QJM 2007 Feb;100(2):87-92. [doi: 10.1093/qjmed/hcm001] [Medline: 17277316]

34. Gansevoort RT, Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, et al. Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. A collaborative meta-analysis of general and high-risk population cohorts. Kidney Int 2011 Jul;80(1):93-104 [FREE Full text] [doi: 10.1038/ki.2010.531] [Medline: 21289597]

35. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis 2002 Feb;39(2 Suppl 1):S1-S266. [Medline: 11904577]

36. Manns B, Hemmelgarn B, Tonelli M, Au F, Chiasson TC, Dong J, et al. Population based screening for chronic kidney disease: cost effectiveness study. BMJ 2010 Nov 08;341:c5869 [FREE Full text] [Medline: 21059726]

37. Komenda P, Ferguson TW, Macdonald K, Rigatto C, Koolage C, Sood MM, et al. Cost-effectiveness of primary screening for CKD: a systematic review. Am J Kidney Dis 2014 May;63(5):789-797. [doi: 10.1053/j.ajkd.2013.12.012] [Medline: 24529536]

38. Qaseem A, Hopkins RH, Sweet DE, Starkey M, Shekelle P. Screening, monitoring, and treatment of stage 1 to 3 chronic kidney disease: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2013 Dec 17;159(12):835-847. [doi: 10.7326/0003-4819-159-12-201312170-00726] [Medline: 24145991]

39. Levey AS, Atkins R, Coresh J, Cohen EP, Collins AJ, Eckardt K, et al. Chronic kidney disease as a global public health problem: approaches and initiatives—a position statement from Kidney Disease Improving Global Outcomes. Kidney Int 2007 Aug;72(3):247-259. [doi: 10.1038/sj.ki.5002343] [Medline: 17568785]

40. Ahmed A, Campbell RC. Epidemiology of chronic kidney disease in heart failure. Heart Fail Clin 2008 Oct;4(4):387-399 [FREE Full text] [doi: 10.1016/j.hfc.2008.03.008] [Medline: 18760751]

41. Shiba N, Shimokawa H. Chronic kidney disease and heart failure—bidirectional close link and common therapeutic goal. J Cardiol 2011 Jan;57(1):8-17 [FREE Full text] [doi: 10.1016/j.jjcc.2010.09.004] [Medline: 21030212]

42. Ho KM, Power BM. Benefits and risks of furosemide in acute kidney injury. Anaesthesia 2010 Mar;65(3):283-293 [FREE Full text] [doi: 10.1111/j.1365-2044.2009.06228.x] [Medline: 20085566]

43. Levin A, Stevens PE. Early detection of CKD: the benefits, limitations and effects on prognosis. Nat Rev Nephrol 2011 Jun 28;7(8):446-457. [doi: 10.1038/nrneph.2011.86] [Medline: 21712852]

44. Hu Z, Jin B, Shin AY, Zhu C, Zhao Y, Hao S, et al. Real-time Web-based assessment of total population risk of future emergency department utilization: statewide prospective active case finding study. Interact J Med Res 2015;4(1):e2 [FREE Full text] [doi: 10.2196/ijmr.4022] [Medline: 25586600]

45. Hu Z, Hao S, Jin B, Shin AY, Zhu C, Huang M, et al. Online prediction of health care utilization in the next six months based on electronic health record information: a cohort and validation study. J Med Internet Res 2015 Sep 22;17(9):e219 [FREE Full text] [doi: 10.2196/jmir.4976] [Medline: 26395541]

46. Hao S, Wang Y, Jin B, Shin AY, Zhu C, Huang M, et al. Development, validation and deployment of a real time 30 day hospital readmission risk assessment tool in the Maine Healthcare Information Exchange. PLoS One 2015;10(10):e0140271 [FREE Full text] [doi: 10.1371/journal.pone.0140271] [Medline: 26448562]

47. Stump CS. Physical activity in the prevention of chronic kidney disease. Cardiorenal Med 2011;1(3):164-173 [FREE Full text] [doi: 10.1159/000329929] [Medline: 22258539]

48. Giovannetta S. A low-nitrogen diet with proteins of high biological value for severe chronic uraemia. Lancet 1964 May 09;1(7341):1000-1003. [Medline: 14129799]

49. Plantinga L, Grubbs V, Sarkar U, Hsu C, Hedgeman E, Robinson B, et al. Nonsteroidal anti-inflammatory drug use among persons with chronic kidney disease in the United States. Ann Fam Med 2011;9(5):423-430 [FREE Full text] [doi: 10.1370/afm.1302] [Medline: 21911761]

50. Mason NA. Angiotensin-converting enzyme inhibitors and renal function. DICP 1990 May;24(5):496-505. [Medline: 2188438]

51. White SL, Polkinghorne KR, Cass A, Shaw JE, Atkins RC, Chadban SJ. Alcohol consumption and 5-year onset of chronic kidney disease: the AusDiab study. Nephrol Dial Transplant 2009 Aug;24(8):2464-2472. [doi: 10.1093/ndt/gfp114] [Medline: 19307230]

52. Orth SR, Hallan SI. Smoking: a risk factor for progression of chronic kidney disease and for cardiovascular morbidity and mortality in renal patients--absence of evidence or evidence of absence? Clin J Am Soc Nephrol 2008 Jan;3(1):226-236 [FREE Full text] [doi: 10.2215/CJN.03740907] [Medline: 18003763]

53. Mollins P. KnowledgeTree. 2015. Predictive versus prescriptive URL: https://www.knowledgetree.com/blog/2015/10/predictive-versus-prescriptive/# [accessed 2017-07-13] [WebCite Cache ID 6q8s2qsMc]

54. Ravera M, Re M, Deferrari L, Vettoretti S, Deferrari G. Importance of blood pressure control in chronic kidney disease. J Am Soc Nephrol 2006 Apr;17(4 Suppl 2):S98-S103 [FREE Full text] [doi: 10.1681/ASN.2005121319] [Medline: 16565257]

55. Peterson JC, Adler S, Burkart JM, Greene T, Hebert LA, Hunsicker LG, et al. Blood pressure control, proteinuria, and the progression of renal disease. The Modification of Diet in Renal Disease Study. Ann Intern Med 1995 Nov 15;123(10):754-762. [Medline: 7574193]

56. Schellenberg ES, Dryden DM, Vandermeer B, Ha C, Korownyk C. Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. Ann Intern Med 2013 Oct 15;159(8):543-551. [doi: 10.7326/0003-4819-159-8-201310150-00007] [Medline: 24126648]

57. Fox CS, Larson MG, Leip EP, Culleton B, Wilson PWF, Levy D. Predictors of new-onset kidney disease in a community-based population. JAMA 2004 Feb 18;291(7):844-850. [doi: 10.1001/jama.291.7.844] [Medline: 14970063]

58. Drey N, Roderick P, Mullee M, Rogerson M. A population-based study of the incidence and outcomes of diagnosed chronic kidney disease. Am J Kidney Dis 2003 Oct;42(4):677-684. [Medline: 14520617]

Abbreviations

ACR: albumin-to-creatinine ratio
AKI: acute kidney injury
CKD: chronic kidney disease
eGFR: estimated glomerular filtration rate
EMR: electronic medical record
HIE: Health Information Exchange
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
NHANES: National Health and Nutrition Examination Survey
PPV: positive predictive value
ROC: receiver operating characteristic
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Original Paper

Is There Evidence of Cost Benefits of Electronic Medical Records, Standards, or Interoperability in Hospital Information Systems? Overview of Systematic Reviews

Zilma Silveira Nogueira Reis[1*], MD, PhD; Thais Abreu Maia[2*], MBA, MSc; Milena Soriano Marcolino[3*], MD, PhD; Francisco Becerra-Posada[4*], MD, DrPH; David Novillo-Ortiz[4*], MLIS, PhD; Antonio Luiz Pinho Ribeiro[5*], MD, PhD

[1]Informatics Center in Health, Obstetrics and Gynecology Department, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[2]State Health Secretariat of Minas Gerais, Planning Advisory, Belo Horizonte, Minas Gerais, Brazil

[3]Faculty of Medicine, Internal Medicine Department, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[4]Pan American Health Organization, Washington, DC, DC, United States

[5]Medical School, Telehealth Center, Hospital das Clínicas, Internal Medicine Department, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[*]all authors contributed equally

Corresponding Author:
Zilma Silveira Nogueira Reis, MD, PhD
Informatics Center in Health
Obstetrics and Gynecology Department
Universidade Federal de Minas Gerais
Av. Prof. Alfredo Balena, 190
Belo Horizonte, Minas Gerais, 30130100
Brazil
Phone: 55 3134099648
Email: zilma.medicina@gmail.com

Abstract

Background: Electronic health (eHealth) interventions may improve the quality of care by providing timely, accessible information about one patient or an entire population. Electronic patient care information forms the nucleus of computerized health information systems. However, interoperability among systems depends on the adoption of information standards. Additionally, investing in technology systems requires cost-effectiveness studies to ensure the sustainability of processes for stakeholders.

Objective: The objective of this study was to assess cost-effectiveness of the use of electronically available inpatient data systems, health information exchange, or standards to support interoperability among systems.

Methods: An overview of systematic reviews was conducted, assessing the MEDLINE, Cochrane Library, LILACS, and IEEE Library databases to identify relevant studies published through February 2016. The search was supplemented by citations from the selected papers. The primary outcome sought the cost-effectiveness, and the secondary outcome was the impact on quality of care. Independent reviewers selected studies, and disagreement was resolved by consensus. The quality of the included studies was evaluated using a measurement tool to assess systematic reviews (AMSTAR).

Results: The primary search identified 286 papers, and two papers were manually included. A total of 211 were systematic reviews. From the 20 studies that were selected after screening the title and abstract, 14 were deemed ineligible, and six met the inclusion criteria. The interventions did not show a measurable effect on cost-effectiveness. Despite the limited number of studies, the heterogeneity of electronic systems reported, and the types of intervention in hospital routines, it was possible to identify some preliminary benefits in quality of care. Hospital information systems, along with information sharing, had the potential to improve clinical practice by reducing staff errors or incidents, improving automated harm detection, monitoring infections more effectively, and enhancing the continuity of care during physician handoffs.

Conclusions: This review identified some benefits in the quality of care but did not provide evidence that the implementation of eHealth interventions had a measurable impact on cost-effectiveness in hospital settings. However, further evidence is needed to infer the impact of standards adoption or interoperability in cost benefits of health care; this in turn requires further research.

XSL·FO
RenderX

Introduction

Information technology (IT) applied to health care, or electronic health (eHealth) [1], ostensibly offers numerous benefits to the quality of health information, particularly in its recording, retrieval, and use. Patients can benefit directly from safe and accessible electronic clinical information for better decision making [2]. However, demographics and patient data are highly fragmented and distributed across multiple unintegrated systems [3]. Comprehensive and consistent health care, leading to effective use of services, requires the computerization of health data for more efficient communication. To achieve this, standardized information channels are needed to make syntactic interoperability possible among electronic records systems. Semantic interoperability is necessary to guarantee the consistency of information, as health information models require adopting standards to support communication [2]. Even if the standardization of electronic health records (EHRs) in eHealth systems is accomplished, health data sharing will continue to be a global challenge. Few publications exist concerning the impact of medical records and interoperability among health systems in cost and benefits of patient care.

Improvements in health and economic indicators are relevant metrics to justify IT investments. Indeed, planning and investing in IT is necessary for the efficient use of information that not only advances health care but also holds financial, social, cultural, and ethical benefits. Comparative cost-effectiveness studies guide agencies and institutions in choosing the best option for desired clinical outcomes and costs, which is the key to ensuring the sustainability of government health systems and their welfare programs [3,4].

This review analyzes systematic reviews addressing the cost benefit and effectiveness of electronic medical records (EMR), standards adoption, or interoperability to discuss the benefits, drawbacks, and lessons learned from the implementation of actions related to eHealth and serves as a reference for government representatives and stakeholders. The assessment of the involvement of government and private health institutions in the implementation and maintenance of eHealth interventions that were tested and valuated worldwide is also of interest. The study was directed by 2 questions: What evidence exists regarding the impact of computerizing applications, standards, health information exchange, or interoperability to support the quality of care or patient outcomes in hospital settings? What critical cost-benefit evidence is published to provide a clear understanding of the value of eHealth implementations?

Methods

Basic Concepts and International Standards

On eHealth

The terms used here to describe eHealth technologies are available in Multimedia Appendix 1. Despite different meanings, some papers use the terms electronic medical record (EMR) and electronic health record (EHR) synonymously. A relevant dissimilarity exists between health information exchange for systems integration and interoperability. The former refers to organizational framework for the dissemination of electronic health care information or clinical data across health-related institutions and systems to enhance patient care [5]. The latter relates to the the ability or capability of two or more systems to exchange information and use the exchanged information, which may support a longitudinal record widely available across institutions and over life spans [6]. Additionally, in a more specific context, "interoperability means the ability of health information systems to work together within and across organizational boundaries in order to advance the effective delivery of health care for individuals and communities" [7].

It is also important to emphasize that interoperability is usually divided into (1) syntactic interoperability: the capability of two or more systems to communicate and exchange data through specified data formats and communication protocols, and (2) semantic interoperability: the ability for data shared by systems to be understood at the level of fully defined domain concepts [8].

Worldwide coordinated efforts resulted in the development of standards to define an EHR as one or more repositories of actionable information by computers. The European Committee for Standardization (CEN), health level seven (HL7), International Organization for Standardization (ISO), and openEHR Foundation are nonprofit organizations dedicated to providing frameworks and standards. Terminologies, EHR specifications, and information models are proposed by these international standards organizations that support the exchange, integration, interoperability, and retrieval of electronic health information [6].

To better represent the meaning of standards in the primary selected systematic reviews, we adopted the generic definition for the term as: "A document adopted by consensus by a recognized entity, that provides rules, guidelines and/or features for common use, in order to obtain an optimal level of performance in a given context…" [9].

On Economic Analysis

Economic analysis supports health care policy and organizational decision making. However, it encounters some difficulties with eHealth systems, which are as follows: constantly changing technology, inconsistent study design to manage inadequate sample sizes, the inappropriateness of

conventional techniques of economic evaluation, and the problem of placing value on health and nonhealth outcomes [10]. Consequently, five methods have been used to calculate the cost-effectiveness of traditional and eHealth interventions: cost-minimization analysis, cost-benefit analysis, cost-effectiveness analysis, cost-utility analysis, and cost-consequence analysis [11].

Data Sources and Search Strategy

This review of systematic reviews has been conducted in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement [12] and the recommended methodological considerations when using existing systematic review as described by Whitlock et al [13].

On February 22, 2016, electronic searches were conducted on the MEDLINE, Cochrane Library, LILACS, and IEEE Library databases. To identify the EHR concept, standards for interoperability, and health information and its cost benefits, the search strategy was:

(("Electronic health records"[MeSH Terms] OR "Health Information Exchange"[MeSH Terms] OR ("Health Information Management"[MeSH Terms] OR ("Medical Informatics" [MeSH Terms]) AND ("Interoperability" OR "Standard of Information")) AND ("Cost-Benefit Analysis"[MeSh] OR "Evaluation Studies"[Publication Type] OR "Program Evaluation"[MeSh] OR impact or effectiveness)

The search was limited by language of publication (English, Spanish, French, Italian, and Portuguese), studies in humans, type of study (systematic reviews and meta-analyses), and year of publication (since 2005). Two systematic reviews that satisfied the criteria were identified manually. To better define certain eHealth technology descriptions, additional sources of evidence were considered.

Study Selection

The inclusion criteria were as follows:

1. Primary impact: EMR, standards, or interoperability on cost-benefit, or
2. Secondary impact: EMR, standards, or interoperability on quality of care (clinical outcomes), and
3. Real-life reviews about interventions in in-hospital settings.

Studies in primary or secondary care scenarios, studies without the primary or secondary impact of eHealth actions, and duplications were excluded. Titles and abstracts of retrieved papers were independently screened and evaluated by 2 investigators (ZSNR and TAM). Abstracts providing insufficient information were retrieved for independent, full-text evaluation by 2 investigators to determine study eligibility. Disagreements were resolved by consensus. Additional publications were identified using the reference lists of selected manuscripts.

Data Extraction and Quality Assessment

ZSNR prepared electronic data with paper contents abstracted using StArt software (Systematic Review System) to organize the analysis [14]. The data extraction of full-text analysis included the following: study design, number of studies evaluated, objectives, type of interventions/clinical data sources, eHealth interventions and terminology, interface/health information exchanges, duration of follow-up, cost-effectiveness, impact on quality of care, main results control group, potential bias, limitations, and lessons learned. The results were summarized into two subgroups according to the modality of intervention:

- Subgroup 1: eHealth systems implementation without health information exchange
- Subgroup 2: eHealth systems with health information exchange functionalities

The methodological quality assessment was based on the AMSTAR (a measurement tool to assess systematic reviews) checklist [15].

Results

A total of 288 papers were identified during the initial research phase, which decreased to 273 after removing 15 duplicates. After applying our criteria, only six systematic reviews were included in the final analysis and data-abstraction phase. The review process is represented in Figure 1, according PRISMA Statement [14].

The primary cause for excluding the 20 studies was mixed or outpatient settings for eHealth interventions (11 papers of 14 excluded, 79%). The Pan American Health Organization (PAHO) conducted a review of the implementation and effective use of standards to achieve interoperability in Latin American and Caribbean countries but without direct or indirect outcomes analysis [6]. Multimedia Appendix 2 presents a detailed summary of the 14 full-text excluded systematic reviews.

Characteristics and Quality of the Selected Studies

Evidence of the cost-effectiveness of eHealth interventions that met the criteria was identified. Only one systematic review of the six performed a meta-analysis [16]. The quality assessment of the included studies followed AMSTAR (a measurement tool to assess systematic reviews) methodology and resulted in wide variability of the quality score. Two studies were classified with a moderate rating of quality with 5 positive points among 11 items [16,17], whereas other reports neglected many AMSTAR criteria [18-21]. Table 1 summarizes the quality assessment ratings, the study design, and the funding or support of the six included systematic reviews.
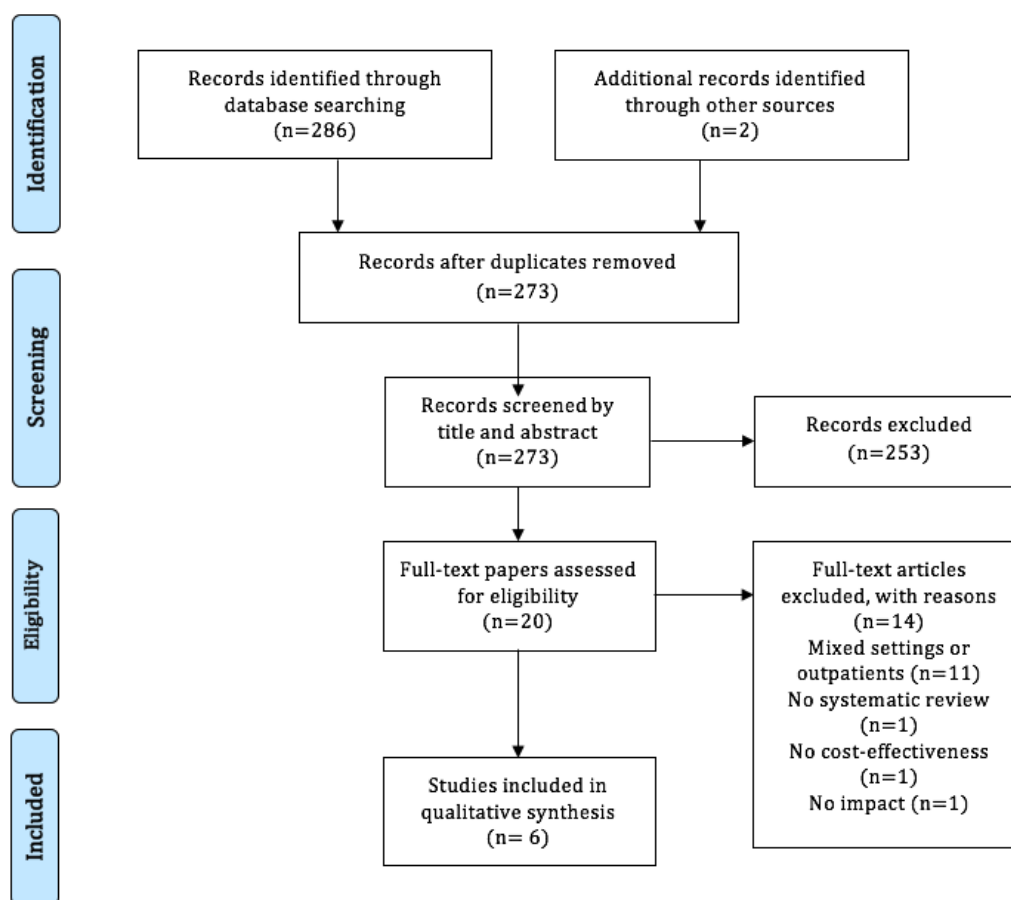
Figure 1. Flow of information through the different phases of the systematic review.

Table 1. Quality assessment ratings and characteristics of the six included systematic reviews.

Study	AMSTAR[a] score				Funding or support	Study design	Number of studies evaluated	Control group (most frequent)	Meta-analysis
	Y[b]	N[c]	CA[d]	N/A[e]					
Thompson et al 2015 [16]	4	7	0	0	Y	RCT[f], pre-post studies, descriptive studies	45 total/Meta-analysis: 26	Pre-post implementation (paper vs system)	Y
Cheung et al 2015 [17]	5	4	0	2	NC[g]	RCT, quasi-experimental studies, descriptive studies	18	Pre-post implementation	N
de Bruin et al 2014 [20]	2	7	0	2	NC	Quasi-experimental	26	True infection detection by infection control experts	N
Mapp et al 2013 [21]	1	7	1	2	NC	Observational, Pilot studies	9	No control	N
Li et al 2013 [18]	4	5	0	2	NC	RCT, quasi-experimental studies	6	Patient not reported in written notes or before system	N
Govindan et al 2010 [19]	5	4	0	2	Y	Observational: accuracy of the automated method with a gold standard method	43	Standard chart review	N

[a]AMSTAR: a measurement tool to assess systematic reviews.

[b]Y: yes.

[c]N: no.

[d]CA: cannot answer.

[e]N/A: not applicable.

[f]RCT: randomized controlled trial.

[g]NC: not commissioned.

Table 2 summarizes the objective, the type of intervention/clinical data sources, eHealth intervention and terminology, interface/health information exchange, and duration of follow-up of the six included systematic reviews.

Summarized Outcomes

Among the included systematic reviews, only one was classified as showing an effect on eHealth implementation without electronic health information exchange (Subgroup 1), and the other five were ranked as showing effects of systems implementation with incorporated health information exchange among other electronic data sources (Subgroup 2).

Subgroup 1

Considering eHealth systems implementation without health information exchange, the review of Thompson et al [16] reported a parallel to advances in digital technology and how different forms of eHealth systems have been developed and implemented (Table 3).

Types and Functions of Technology Systems

The selected review stated a mix of electronic interventions: EHR, EMR, computerized decision support systems (CDSS), computerized provider order-entry (CPOE) and surveillance systems used by physicians, nurses, allied health professionals, and managers of health services evaluating evidence from pre-and postsystems implementation. The analysis synthesized 46 publications about systems for diagnosis, treatment, and clinical monitoring. The study included a meta-analysis extracted from 26 publications to evaluate the effects of different types of systems regarding health IT in the inpatient of intensive care unit (ICU) setting on mortality, length of stay (LOS), and cost.

Effects on Quality or Efficiency of Care

Not enough evidence showed that electronic interventions can improve quality and safety of health care. The goals for secondary outcomes were the effects of health IT in the inpatient and ICU on mortality or LOS. The quality of included studies and interventions varied significantly, which was highlighted as the major limitation. Despite this, the surveillance systems had a pooled odd ratio (OR) of 0.85 (95% CI 0.76-0.94) with moderate heterogeneity, I^2 of 59%.

Effects on Costs

Costs were unable to be evaluated quantitatively because the primary studies presented mixed and inconclusive results, leaving us unable to draw a definitive conclusion about cost-effectiveness. The analysis of costs was more limited than the evidence on quality and efficiency.

Subgroup 2

EHR implementation with health information exchange is a recent worldwide trend in hospital settings. A summary of the results of the systematic reviews included in subgroup 2 is presented in Table 4.

XSL•FO
RenderX

Table 2. Descriptive summary of the systematic reviews included in electronic medical records (EMRs)/Interoperability review.

Study	Objective	Type of intervention/ Clinical data sources	eHealth intervention and terminology	Interface/health information exchange	Duration of follow-up
Thompson et al 2015 [16]	To evaluate effects of health IT[a] in the inpatient and ICU[b] on mortality, LOS[c], and cost	Multiple health IT interventions on diagnosis, treatment, monitoring, cost reduction/No reference	EHR[d], EMR[e], CDSS[f], CPOE[g], Surveillance system	No reference	No reference
Cheung et al 2015 [17]	To evaluate the effects of an information system integrated to PDMS[h] on organizational and clinical outcomes, in ICU[i]/Operating room	Integrating bedside equipment to an information system/vital signs, patient monitor, ventilator, anesthesia machine, dialysis machine, IV pump, lab values, hospital information system, admission, discharge and transfer	CDSS, PDMS, health information exchange	PDMS to an information system/no mention about direction of data exchange	1 day to 1 week; 11 months to 4 years
de Bruin et al 2014 [20]	To evaluate recent trends in use of electronically available patient data by electronic surveillance systems for HAIs[j] and identify consequences for system effectiveness	HAIs that utilize EHR available in hospitals to surveillance the HAIs/Medico-administrative data procedures or discharge reports, free text reports, biochemistry, microbiology, and radiology laboratory test results, pharmacy dispensing records, radiology free-text records, vital signs, electronic discharge summary	Automated detection by HAI systems: EHR, health information exchange, using ICD[k]-9, ICD-10, discharge coding, ATC[l] code	EHR to HAI systems/no mention about direction of data exchange	No reference
Mapp et al 2013 [21]	To examine early warning scoring systems and their effectiveness in predicting a patient's potential for deterioration and considers whether these scoring systems prevent unplanned ICU admissions and/or death	Instruments and clinical support systems available to assist health care personnel in recognizing early clinical deterioration/Vital signs, SpO2[m], LOC[n], UOP[o], nurse/family concerns, complaints, lab values	EMR, CDSS, health information exchange based on SBAR[p] communication	Early warning scoring systems that interface with EMRs and are supplemented with decision aides (algorithms) and clinical support systems/no mention about direction of data exchange	Seven studies: 3 to15 months/two studies: over 24 months to 8 years
Li et al 2013 [18]	To evaluate the impact of the CHTs[q] on the quality of physician handoff, patient care, and physician work efficiency	Decision support/training, emergency referrals, supervision, alerts and reminders, client education, data collection, medicine dosing/Patient demographics, medications, diagnosis, problem lists, comment line, vital signs, to-do list, LOS, free daily notes, lab values	CHTs, EMR, CDSS, health information exchange. Allergy Code	Clinical information exchange using CHTs for physician handoff for hospitalized patients CHTs/mixed (no interface, unidirectional or bidirectional interface exchange)	1 to 6 months
Govindan et al 2010 [19]	To identify, describe, and evaluate the effectiveness of automated inpatient harm-detection methods	Automated harm detection on EMR. Gold standard: chart review	Automated detection by surveillance systems: EMR, health information exchange, using ICD-9, procedure codes, billing codes	Automated harm detection on EMR, using field-defined systems, natural language-processing/Unidirectional retrospective	No reference

<superscript>a</superscript>IT: information technology.

<superscript>b</superscript>ICU: intensive care unit.

<superscript>c</superscript>LOS: length of stay.

<superscript>d</superscript>EHR: electronic health record.

<superscript>e</superscript>EMR: electronic medical record.

<superscript>f</superscript>CDSS: computerized decision support systems.

<superscript>g</superscript>CPOE: computerized provider order-entry.

<superscript>h</superscript>PDMS: Patient data management system.

<superscript>i</superscript>ICU: intensive care unit.

<superscript>j</superscript>HAIs: health care–associated infections systems.

<superscript>k</superscript>ICD: international classification of disease.

<superscript>l</superscript>ATC: anatomical therapeutic chemical.

<superscript>m</superscript>SpO$_2$: oxygen saturation.

<superscript>n</superscript>LOC: level of consciousness.

<superscript>o</superscript>UOP: urine output.

<superscript>p</superscript>SBAR: situation, background, assessment, recommendation.

<superscript>q</superscript>CHTs: computerized physician handoff tools.

Table 3. Descriptive summary of the results of systematic reviews included in electronic medical records(EMRs)/Interoperability review. Subgroup 1: electronic health (eHealth) systems implementation without health information exchange.

Study	Primary impact: Cost-effectiveness	Secondary impact: Quality of care/ Clinical outcome	Main results	Potential bias	Lessons
Thompson et al 2015 [16]	Mixed and inconclusive	Mortality: overall CPOE[a] systems did not show a significant effect (OR[b]: 0.91, 95% CI 0.75-1.10; I^{2c} 66%), nor EHR[d] alone (OR: 0.96, 95% CI 0.77-1.19). CDSS[e](OR 0.96, 95% CI 0.77-1.19). The surveillance systems had a pooled OR of 0.85 (95% CI 0.76-0.94) with moderate heterogeneity, I^{2}59% LOS: CPOE trended toward a reduction in LOS (mean decrease, 0.67 days, 95% CI –2.07 to 0.73), though with significant heterogeneity (I^{2}82%). Neither CDSS nor surveillance systems trended toward changes in hospital LOS, and the net-pooled effect was not significant.	Electronic interventions were not shown to have a substantial effect on mortality, LOS[f], or cost.	Selection, measurement	There is not enough evidence to confidently state that electronic interventions have the ability to achieve the goal of improving quality and safety.

<superscript>a</superscript>CPOE: computerized provider order-entry.

<superscript>b</superscript>OR: odds ratio.

<superscript>c</superscript>I^2: measure of heterogeneity.

<superscript>d</superscript>EHR: electronic health record.

<superscript>e</superscript>CDSS: computerized decision support systems.

<superscript>f</superscript>LOS: length of stay.

Table 4. Descriptive summary of the results of systematic reviews included in the electronic health record (EHR)/Interoperability review. Subgroup 2: electronic health (eHealth) systems implementation with information exchange.

Study	Primary impact: Cost-effectiveness	Secondary impact: Quality of care/ Clinical outcome	Main results	Potential bias	Lessons
Cheung et al 2015 [17]	Not evaluated[a]	PDMS[b] reduced charting time, increased time spent on direct patient care and reduced the occurrence of errors (medication errors, intravenous and ventilation incidents). The effect on documentation was mixed. Improvement in clinical outcomes when PDMS was integrated with a CDSS[c], but scarce literature is available.	The effect on documentation was mixed. Qualitative analysis showed a significant decrease in time spent on documentation. Clinical outcomes: inconclusive.	Selection, measurement	Improvement in clinical outcomes when PDMS was integrated with a CDSS, but there is scarce literature available. Organizational advantages included improved accuracy, legibility, data accessibility, and decision support. Such integration may improve clinical outcomes, although further studies are required for validation.
de Bruin et al 2014 [20]	Not evaluated[a]	Electronic surveillance achieves equal or better sensitivity than manual surveillance. Several studies also reported time savings of 60% to 99.9% or a reduction in chart reviews of 40% to 90.5%.	Driven by the increased availability of electronic patient data, electronic HAIs[d] surveillance systems use more data, making systems more sensitive yet less specific but also allow systems to be tailored to the needs of health care institutes' surveillance programs.	Selection	HAIs detection systems use increasingly more EHR[e] and patient data as more data sources become available. Thus, systems tend to become more sensitive and less specific.
Mapp et al 2013 [21]	Not evaluated[a]	An increase occurred in the number of rapid response calls by nursing staff, a decrease in unplanned ICU[f] admissions, and a decrease in hospital mortality.	Improvement in clinical outcomes when using early warning scoring systems.	Selection	Early warning scoring systems can be more effective with the integration of algorithms and clinical support systems.
Li et al 2013 [18]	Not evaluated[a]	Impact on physician work efficiency (self-reported time spent on handing copying patient information; 50%) and proportionally more time to see patients. Time on each patient during rounding decreased by 1.5 min. Impact on quality on physician handoff: completeness and consistency of the handoff document has improved.	Completeness and consistency of the handoff document has improved. Accuracy of information about patients during physician handoff.	Selection, measurement	CHTs[g] could potentially enhance work efficiency and continuity of care during physician handoff, but the role in improving quality is less clear. The information available was often not sufficient to help on-call physicians make patient care decisions.

Study	Primary impact: Cost-effectiveness	Secondary impact: Quality of care/ Clinical outcome	Main results	Potential bias	Lessons
Govindan et al 2010 [19]	Not evaluated[a]	Sensitivities of different methods ranged from 0.10 to 0.94, specificity from 0.10 to 0.94, PPV[h] from 0.03 to 0.84, and NPV[i] from 0.70 to 0.96. The field-defined methods of automated harm detection will prove superior to natural language processing, particularly if information about harm is accurately documented.	Automated harm detection has the potential to positively influence clinical practice.	Selection, measurement	Automated harm detection has the potential to positively influence clinical practice. Another potential benefit is the reduction of person-hour required to harm surveillance.

[a]Not evaluated in the selected study.

[b]PDMS: Patient data management system.

[c]CDSS: computerized decision support systems.

[d]HAIs: health care–associated infections systems.

[e]EHR: electronic health record.

[f]ICU: intensive care unit.

[g]CHTs: computerized physician handoff tools

[h]PPV: positive predictive value.

[i]NPV: negative predictive value.

Types and Functions of Technology Systems

Most of the reviews use ICUs as settings for eHealth intervention analysis. However, the objectives of interventions were quite heterogeneous. Two studies reported the effect of surveillance systems on harm detection [19] and health care–associated infections [20]. Bedside data integration in an information system [17], continuity of care using physician handoff tools [18], and prediction of death or unexpected ICU admission [21] were the proposals of the other reviews. Regarding application users, two studies focused on patient outcome results for health care managers [19,20]. Some focused directly on health care professionals to improve clinical practice [18,19,21]. On the direction of electronic health information exchange, one review described it as unidirectional [19], three did not clarify whether the exchange was bidirectional [17,20,21], and one summarized mixed studies including systems without interfaces [18]. None mentioned interoperability among electronic health systems. Regarding standards for the exchange of clinical data, four studies reported the use of terminologies such as International Classification of Disease (ICD) and anatomical therapeutic chemical (ATC) code [18-21].

Effects on Quality or Efficiency of Care

Among reviews focused on improving clinical practice, inconclusive results in direct patient care were reported by Cheung et al [17]. Mapp et al [21] highlighted an increase in nursing staff efficiency regarding rapid calls response, a decrease in unplanned ICU admissions, and hospital mortality. Li et al [18] presented a positive impact on continuity of inpatient care. With regard to indirect results on patient care, two studies highlighted the improvement of health data quality in terms of accuracy, legibility, completeness, and consistency of documents [17,18]. The other reviews focused on electronic surveillance. The results showed that systems tend to become more sensitive and less specific than manual monitoring to detect infection [20]. With respect to inpatient harm detection, the automated systems allowed rapid scanning of a vast number of patient records with minimal effort and may identify events as they occur in real time [19]. Most automated surveillance systems were retrospective, but some real-time surveillance alerts that informed physicians and pharmacists of adverse events were reported [19].

Effects on Costs

None of the reviews evaluated effects of eHealth interventions on costs.

Discussion

Principal Findings

This study found preliminary benefits in the use of electronically available inpatient data systems on the quality of care. Despite the limited number of studies that met the eligibility criteria, the heterogeneity of electronic systems reported, and different interventions on hospital routines, the identification of preliminary secondary benefits on patient mortality was possible [16]. eHealth systems with information exchange functionalities also showed potential impact on quality of care or patient outcomes. From five studies, one had inconclusive results on direct patient care [17] and four presented partial effects, as nursing staff efficiency led to a faster call response, a decrease in unplanned ICU admissions and hospital mortality [18], improvement of health data quality [17,18], and more efficient

surveillance programs inside hospitals [19,20]. It is expected that the systems able to share health information would improve care at the time and point of attention, especially the surveillance systems and those that use common terminologies and vocabularies to support consistency in information collection [6,19,20,22].

However, no substantial review regarding the impact of electronic interventions on cost-effectiveness was identified. Among the six analyses included, only Thompson et al reported that some preliminary studies have identified decreases in cost, but the heterogeneity and the absence of information of follow-up impaired a proper analysis of cost-effectiveness [16]. Immediate cost savings are not anticipated for organizations when choosing to adopt eHealth strategies because the high cost of implementation limits the transition from paper-based to electronic systems and represents a significant challenge to their widespread adoption [23]. Regardless, medium and long-term positive results are expected, and the World Health Organization (WHO) recognized overall eHealth as cost- effective and secure [24]. Potential indirect cost saving was mentioned as a secondary outcome in three studies, with the reduction of person-hours harming surveillance and the increase in time spent on direct patient care [17,18,24].

Unfortunately, no study about interoperability, in the sense of syntactic and semantic meaning, on cost benefit was identified. Importantly, none of the studies in this review properly defined EHR concept as a longitudinal health record with entries by health care practitioners in multiple sites of care or mentioned interoperability applications among electronic systems. However, taking the antecedent step toward full interoperability, an effective information sharing between stakeholders and systems can be attained through the use of standards [6]. Standards adoption for the exchange of clinical data was mentioned in four studies [18-21], mostly terminologies adoptions, but the potential impact of such tools on continuity of care or costs remains an open question that needs investigation. Although within the limits of hospital systems, the analysis confirmed the potential to positively impact physician practice organizations, as previously reported [23].

Further longitudinal research is needed to determine the actual impact of eHealth adoption on health care costs and clinical outcomes.

Limitations

The current results should be interpreted as a whole with the study limitations. Only four major databases were searched and gray literature sources were not evaluated. Additionally, the limitation to English, Spanish, French, Italian, and Portuguese languages prevented the capture of all relevant studies. Furthermore, the quality of included studies was poor, and they varied regarding the type of eHealth interventions, follow-up time, and goals. This systematic review summarized primary and secondary outcomes from different classes of intervention from which to draw results, analysis, and conclusions. Due to the variation in scenarios and lack of numeric goals, a meta-analysis was considered inappropriate.

Conclusions and Lessons Learned

This review identified some benefits on the quality of care but did not provide evidence that the eHealth interventions had a measurable impact on cost-effectiveness, mortality, or LOS in hospital settings. Preliminary evidence indicates that the use of eHealth interventions with information exchange may improve clinical process outcomes. The absence of studies precludes the assessment of impact of interoperability on benefits of health care or cost, and this aspect needs further research. Technological barriers might influence eHealth solutions implementation and data exchange for systems integration or interoperable interfaces. There are also issues with the lack of standardization of most aspects of health information and misuse of terms in the scientific publications. Authors should be explicit when they are using interfacing syntactic interoperability or semantic interoperability to reduce the confusion with different health information exchange possibilities. Further research with long-term follow-up is needed to determine the actual impact of eHealth adoption on health care costs to demonstrate (1) value for money (including clinical impacts) and (2) the clinical impact of semantic and synthetic interoperability.

Conflicts of Interest

None declared.

Multimedia Appendix 1

eHealth concepts and definitions.

XSL·FO

RenderX

[PDF File (Adobe PDF File), 49KB - medinform_v5i3e26_app1.pdf]

Multimedia Appendix 2

Descriptive summary of the 14 full-text excluded systematic reviews.

[PDF File (Adobe PDF File), 41KB - medinform_v5i3e26_app2.pdf]

References

1. World Health Organization. Apps.WHO. 2006. Building foundations for eHealth : progress of member states : report of the Global Observatory for eHealth URL: http://apps.who.int/iris/handle/10665/43599 [accessed 2017-07-13] [WebCite Cache ID 6rw1isUOl]

2. World Health Organization. Apps.WHO. 2006. Electronic health records: manual for developing countries URL: http://apps.who.int/iris/handle/10665/207504 [accessed 2017-07-13] [WebCite Cache ID 6sm2rrOne]

3. Shekelle PG, Morton SC, Keeler EB. Costs and benefits of health information technology. Evid Rep Technol Assess (Full Rep) 2006 Apr(132):1-71. [Medline: 17627328]

4. World Health Organization. WHO. Geneve: WHO Press; 2016. Atlas of eHealth country profiles: the use of eHealth in support of universal health coverage: based on the findings of the third global survey on eHealth URL: http://www.who.int/goe/publications/atlas_2015/en/ [accessed 2017-07-14] [WebCite Cache ID 6rF0c5lHC]

5. ncbi.nlm. 2017. Health Information Exchange URL: https://www.ncbi.nlm.nih.gov/mesh/?term=health+information+exchange [accessed 2017-06-20] [WebCite Cache ID 6rMNCkzRG]

6. Pan American Health Organization. Webcitation.: Pan American Health Organization; 2016. eHealth in Latin Americathe Caribbean: interoperability standards review URL: http://www.webcitation.org/6rwrfeqBJ [accessed 2017-07-14] [WebCite Cache ID 6rwrfeqBJ]

7. Webcitation. 2005. Interoperability Definition and Background URL: http://www.webcitation.org/6rKvIpuvO [accessed 2017-07-14] [WebCite Cache ID 6rKvIpuvO]

8. International Organization for Standarization - ISO. ISO. Health informatics -- Requirements for an electronic health record architecture URL: https://www.iso.org/standard/52823.html [accessed 2017-06-19] [WebCite Cache ID 6rL0KEbKn]

9. International Organization for Standarization - ISO. Webcitation. Standardization and related activities -- General vocabulary URL: http://www.webcitation.org/6rKvaBnAJ [accessed 2017-07-14] [WebCite Cache ID 6rKvaBnAJ]

10. Mistry H. Systematic review of studies of the cost-effectiveness of telemedicine and telecare. Changes in the economic evidence over twenty years. J Telemed Telecare 2012 Jan;18(1):1-6. [doi: 10.1258/jtt.2011.110505] [Medline: 22101609]

11. Rojas SV, Gagnon M. A systematic review of the key indicators for assessing telehomecare cost-effectiveness. Telemed J E Health 2008 Nov;14(9):896-904 [FREE Full text] [doi: 10.1089/tmj.2008.0009] [Medline: 19035798]

12. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche P, Ioannidis J. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventionsxplanation and elaboration. Br Med J 2009;339:b2700 [FREE Full text] [doi: 10.1371/journal.pmed.1000100]

13. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med 2008 May 20;148(10):776-782. [Medline: 18490690]

14. Hernandes E, Zamboni A, Di Thommazo A, Fabbri S. Avaliação da ferramenta StArt utilizando o modelo TAM e o paradigma GQM. In: Proceedings of 7th Experimental Software Engineering Latin American Workshop (ESELAW 2010). Goiânia: UFG/INF/FUNAPE; 2010 Presented at: 7th Experimental Software Engineering Latin American Workshop (ESELAW 2010); 12/10/2010; Goiânia URL: http://www.portal.inf.ufg.br/~eselaw2010/proceedings-eselaw2010.pdf#page=37

15. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007 Feb 15;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]

16. Thompson G, O'Horo JC, Pickering BW, Herasevich V. Impact of the electronic medical record on mortality, length of stay, and cost in the hospital and ICU: a systematic review and meta-analysis. Crit Care Med 2015 Jun;43(6):1276-1282. [doi: 10.1097/CCM.0000000000000948] [Medline: 25756413]

17. Cheung A, van Velden FH, Lagerburg V, Minderman N. The organizational and clinical impact of integrating bedside equipment to an information system: a systematic literature review of patient data management systems (PDMS). Int J Med Inform 2015 Mar;84(3):155-165. [doi: 10.1016/j.ijmedinf.2014.12.002] [Medline: 25601332]

18. Li P, Ali S, Tang C, Ghali WA, Stelfox HT. Review of computerized physician handoff tools for improving the quality of patient care. J Hosp Med 2013 Aug;8(8):456-463. [doi: 10.1002/jhm.1988] [Medline: 23169534]

19. Govindan M, Van Citters AD, Nelson EC, Kelly-Cummings J, Suresh G. Automated detection of harm in healthcare with information technology: a systematic review. Qual Saf Health Care 2010 Oct;19(5):e11. [doi: 10.1136/qshc.2009.033027] [Medline: 20671081]

20. de Bruin JS, Seeling W, Schuh C. Data use and effectiveness in electronic surveillance of healthcare associated infections in the 21st century: a systematic review. J Am Med Inform Assoc 2014;21(5):942-951 [FREE Full text] [doi: 10.1136/amiajnl-2013-002089] [Medline: 24421290]

XSL·FO
RenderX

21. Mapp ID, Davis LL, Krowchuk H. Prevention of unplanned intensive care unit admissions and hospital mortality by early warning systems. Dimens Crit Care Nurs 2013;32(6):300-309. [doi: 10.1097/DCC.0000000000000004] [Medline: 24100432]

22. Hammond WE, Bailey C, Boucher P, Spohr M, Whitaker P. Connecting information to improve health. Health Aff (Millwood) 2010 Feb;29(2):284-288 [FREE Full text] [doi: 10.1377/hlthaff.2009.0903] [Medline: 20348075]

23. Police RL, Foster T, Wong KS. Adoption and use of health information technology in physician practice organisations: systematic review. Inform Prim Care 2010;18(4):245-258 [FREE Full text] [Medline: 22040851]

24. Asamblea Mundial de la Salud. Apps.WHO. 2005. Cibersalud URL: http://apps.who.int/iris/handle/10665/23104 [accessed 2017-07-14] [WebCite Cache ID 6smbcRPBE]

25. Kawamoto K, Houlihan C, Balas E, Lobach D. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Br Med J 2005;330(7494):765. [doi: 10.1136/bmj.38398.500764.8F]

26. Mitchell J. From Telehealth to E-health: The Unstoppable Rise of E-health. Canberra, NSW: Commonwealth Department of Communications, Information Technology and the Arts; 1999:064275036.

27. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]

28. Pan American Health Organization. PAHO. Washington, DC: PAHO HQ Library Cataloguing-in-Publication; 2011. Epidemiological Surveillance of Healthcare Associated Infections URL: http://new.paho.org/hq/dmdocuments/2011/ENG_Modulo_I_final.pdf [accessed 2017-07-14] [WebCite Cache ID 6rwqNuepz]

29. Main C, Moxham T, Wyatt J, Kay J, Anderson R, Stein K. Computerised decision support systems in order communication for diagnostic, screening or monitoring test ordering: systematic reviews of the effects and cost-effectiveness of systems. Health Technol Assess 2017 Jul 14;14(48):1-227 [FREE Full text] [doi: 10.3310/hta14480] [Medline: 21034668]

30. 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 16;144(10):742-752. [Medline: 16702590]

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

32. Wollersheim D, Sari A, Rahayu W. Archetype-based electronic health records: a literature review and evaluation of their applicability to health data interoperability and access. HIM J 2009;38(2):7-17. [Medline: 19546483]

33. Chan AJ, Chan J, Cafazzo JA, Rossos PG, Tripp T, Shojania K, et al. Order sets in health care: a systematic review of their effects. Int J Technol Assess Health Care 2012 Jul;28(3):235-240. [doi: 10.1017/S0266462312000281] [Medline: 22980699]

34. 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]

35. Ojeleye O, Avery A, Gupta V, Boyd M. The evidence for the effectiveness of safety alerts in electronic patient medication record systems at the point of pharmacy order entry: a systematic review. BMC Med Inform Decis Mak 2013 Jul 01;13:69 [FREE Full text] [doi: 10.1186/1472-6947-13-69] [Medline: 23816138]

36. Walsh C, Siegler EL, Cheston E, O'Donnell H, Collins S, Stein D, Informatics Intervention Research Collaboration (I2RC). Provider-to-provider electronic communication in the era of meaningful use: a review of the evidence. J Hosp Med 2013 Oct;8(10):589-597 [FREE Full text] [doi: 10.1002/jhm.2082] [Medline: 24101544]

37. Tenório J, Cohrs F, Pisa I, de Fátima Marin H. Personal health record: a review on current models and experiments. J Health Inform 2013:91-97 [FREE Full text]

38. Moja L, Kwag KH, Lytras T, Bertizzolo L, Brandt L, Pecoraro V, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. Am J Public Health 2014 Dec;104(12):e12-e22. [doi: 10.2105/AJPH.2014.302164] [Medline: 25322302]

39. Nguyen L, Bellucci E, Nguyen LT. Electronic health records implementation: an evaluation of information system impact and contingency factors. Int J Med Inform 2014 Nov;83(11):779-796. [doi: 10.1016/j.ijmedinf.2014.06.011] [Medline: 25085286]

40. Fiander M, McGowan J, Grad R, Pluye P, Hannes K, Labrecque M, et al. Interventions to increase the use of electronic health information by healthcare practitioners to improve clinical practice and patient outcomes. Cochrane Database Syst Rev 2015 Mar 14(3):CD004749. [doi: 10.1002/14651858.CD004749.pub3] [Medline: 25770311]

Abbreviations

AMSTAR: a measurement tool to assess systematic reviews
ATC: anatomical therapeutic chemical
CDSS: computerized decision support systems
CEN: European Committee for Standardization
CHTs: computerized physician handoff tools
CPOE: computerized provider order-entry

XSL•FO
RenderX

CPR: computer-based patient record
EHR: electronic health record
EMR: electronic medical record
HAIs: health care–associated infections surveillance systems
HL7: Health level seven
I2: measure of heterogeneity
ICD: International Classification of Disease
ICU: Intensive Care Unit
ISO: International Organization for Standardization
IT: information technology
LOC: level of consciousness
LOS: length of stay
NPV: negative predictive value
OR: odds ratio
PAHO: Pan American Health Organization
PDMS: patient data management system
PPV: positive predictive value
PRISMA: preferred reporting items for systematic reviews and meta-analyses
SpO2: oxygen saturation
SBAR: situation, background, assessment, recommendation
UOP: urine output
WHO: World Health Organization

XSL•FO
RenderX

<u>Original Paper</u>

Prototype Development: Context-Driven Dynamic XML Ophthalmologic Data Capture Application

Peggy Peissig[1], PhD; Kelsey M Schwei[2], PhD; Christopher Kadolph[1], BS; Joseph Finamore[1], BS; Efrain Cancel[3], MD; Catherine A McCarty[4], MPH, PhD; Asha Okorie[3], MD; Kate L Thomas[1], MA; Jennifer Allen Pacheco[5], MS; Jyotishman Pathak[6], PhD; Stephen B Ellis[7], BS; Joshua C Denny[8], MS, MD; Luke V Rasmussen[9], MS; Gerard Tromp[10], PhD; Marc S Williams[11], MD; Tamara R Vrabec[12], MD; Murray H Brilliant[13], PhD

[1]Marshfield Clinic Research Institute, Biomedical Informatics Research Center, Marshfield, WI, United States
[2]Marshfield Clinic Research Institute, Center for Oral and Systemic Health, Marshfield, WI, United States
[3]Marshfield Clinic, Department of Ophthalmology, Marshfield, WI, United States
[4]Essentia Institute of Rural Health, Center for Research and Education, Duluth, MN, United States
[5]Center for Genetic Medicine, Northwestern University, Chicago, IL, United States
[6]Weill Cornell Medical College, Healthcare Policy and Research, Cornell University, New York, NY, United States
[7]Personalized Medicine Institute, Mount Sinai, New York, NY, United States
[8]School of Medicine, Biomedical Informatics, Vanderbilt University, Nashville, TN, United States
[9]Division of Health and Biomedical Informatics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[10]Autism and Developmental Medicine Institute (ADMI), Geisinger, Danville, PA, United States
[11]Genomic Medical Institute, Geisinger, Danville, PA, United States
[12]Department of Ophthalmology, Geisinger, Danville, PA, United States
[13]Marshfield Clinic Research Foundation, Human Genetics, Marshfield, WI, United States

Corresponding Author:
Peggy Peissig, PhD
Marshfield Clinic Research Institute
Biomedical Informatics Research Center
1000 N Oak Ave
Marshfield, WI, 54449
United States
Phone: 1 715 221 8322
Fax: 1 715 221 6402
Email: peissig.peggy@mcrf.mfldclin.edu

Abstract

Background: The capture and integration of structured ophthalmologic data into electronic health records (EHRs) has historically been a challenge. However, the importance of this activity for patient care and research is critical.

Objective: The purpose of this study was to develop a prototype of a context-driven dynamic extensible markup language (XML) ophthalmologic data capture application for research and clinical care that could be easily integrated into an EHR system.

Methods: Stakeholders in the medical, research, and informatics fields were interviewed and surveyed to determine data and system requirements for ophthalmologic data capture. On the basis of these requirements, an ophthalmology data capture application was developed to collect and store discrete data elements with important graphical information.

Results: The context-driven data entry application supports several features, including ink-over drawing capability for documenting eye abnormalities, context-based Web controls that guide data entry based on preestablished dependencies, and an adaptable database or XML schema that stores Web form specifications and allows for immediate changes in form layout or content. The application utilizes Web services to enable data integration with a variety of EHRs for retrieval and storage of patient data.

Conclusions: This paper describes the development process used to create a context-driven dynamic XML data capture application for optometry and ophthalmology. The list of ophthalmologic data elements identified as important for care and research can be used as a baseline list for future ophthalmologic data collection activities.

XSL•FO
RenderX

Introduction

Background

Capturing clinical information in a machine-interpretable format is challenging yet extremely critical to patient care, biomedical research, health care quality, and workflow efficiency initiatives [1-7]. However, without structured data capture, computers cannot be used to perform or enhance many tasks surrounding a patient encounter (eg, appointing, diagnostic test ordering, return of test results with interpretation, and billing) or to present treatment options to clinical staff [8,9]. In addition, some research, quality assessment, and process improvement activities may be cost prohibitive because of manual information review.

Electronic health records (EHRs) use a variety of mechanisms to capture structured (or coded) medical information to meet high-volume regulatory and billing requirements, but they often lack options for lower volume ancillary or practice-specific medical initiatives or research. As a result, health care institutions often supplement EHR data capture functionality with vendor software solutions, form development tools, or standard office tools (eg, spreadsheets and document editors) to create such capability [10-12]. These solutions usually have EHR integration limitations, are platform or device dependent, are difficult to maintain, lack security, and are limited in the ability to capture and use a broad range of clinical data capture methods. Not all of these solutions have exposed application program interfaces (APIs) that can be used to interface with EHRs, and there are challenges of maintaining functionality with the upgradation of the form system or EHR. Additionally, there are Web-based subscriptions or paid services that provide varying implementations of Web forms, but these have limited feature sets resulting in the inability to create custom questions, are unable to support system integration and illustration functionality requirements, and lack regulatory assurances needed for handling medical data (eg, Health Insurance Portability & Accountability Act [HIPAA] Security; [13-15]).

Processing information found within a document can be accomplished using a flexible machine and human interpretable markup language referred to as extensible markup language (XML). XML can be used to create Web-based data capture forms supporting a variety of common data formats, including text, numeric, coded results (eg, multiple choice), voice clips, and images. There are a number of Web-based data capture applications reported in the literature that have been used for research [16-19] and clinical care [20,21]. Notable innovations such as the dynamic generation of structured data entry forms using metadata stored in databases [20] or XML schemas, XML schema designer tools [18,22], or the integration of EHR data into an XML form [17] are reported and could be used.

Optometry and ophthalmology are specialty areas in health care that deal with the anatomy, function, and diseases of the eye and its surrounding structures. These specialties have been slow to adopt EHR or electronic data capture technology because of characteristically high-volume practices with complex workflows (eg, office visit to surgical suite transitions) and image-intensive documentation requirements [23-26]. Optical Coherence Tomography (OCT) specialty testing images, for example, can be captured by an external camera and scanned into the EHR as a PDF, thereby trapping numeric data in images as opposed to discretely computable fields. Much of the clinically relevant information is captured with hand-drawn illustrations of the eye, thus making it unavailable for subsequent clinical or research use unless manually interpreted [27]. In some countries, procedure reimbursement of nonphotographic retinal images relies on examiner-generated illustrations [28], and as many electronic drawing tools are not yet sophisticated enough to execute retinal illustrations, they reinforce the use of paper in clinical practice.

Objective

Finding or developing software solutions that can easily integrate with and supplement basic EHR data capture functionality for complicated workflows are necessary for patient care, research, and quality management efforts [29]. Using the ophthalmologic areas as a use case for prototype development, we describe the process for identifying clinical- and research-relevant data elements and the creation of an open-source, context-driven, ophthalmologic data capture application that can be easily adapted to a variety of clinical workflows and EHR environments. This work takes advantage of previous research efforts [20] that use metadata encapsulated in a database or an XML schema to drive form generation and context-driven controls to populate the data capture application. We expand on these approaches by applying the technologies in an integrated fashion to the ophthalmologic areas and add ink-over Web controls to capture unstructured clinically relevant drawings and notes. Ink-over, often referred to as digital ink, refers to technology that digitally represents handwritten notes and drawings. In this study, we describe the process used to develop an ophthalmologic data capture prototype application and its heuristic evaluation.

Methods

Goal

The overall goal of this research was to develop a context-driven dynamic XML ophthalmologic prototype that enables efficient data capture of ophthalmologic and optometric data and increases the collection of discrete (structured) data while preserving the ability to capture handwritten notes and drawings where needed.

The specific objectives included: (1) determining clinical- and research-relevant ophthalmologic data elements (both structured and image-based); (2) documenting system requirements for an ancillary data capture application; and (3) designing, developing,

and evaluating a data capture prototype that is context-aware, modifiable, and can integrate into an EHR.

Environment

Marshfield Clinic Health System's CattailsMD EHR has been used for clinical care since the late 1980s, serving clinicians throughout Central and Northern Wisconsin. It uses a variety of data-gathering techniques to capture and code patient encounter information, including diagnoses, laboratory results, procedures, medications, and vital sign measurements such as height, weight, and blood pressure. Clinical narratives and illustrations are stored in textual and/or image-based unstructured formats and made available for viewing via the EHR. Medical staff use tablet personal computers (PCs) to interact with EHR applications that have been optimized to run on these devices. Ophthalmology and Optometry departments use these EHR applications, but they also supplement clinical data capture by using paper-based forms that are scanned into the EHR.

Process

We used a participatory process to develop the ophthalmologic data capture prototype. A design team comprising stakeholders (3 physicians, 2 administrators, 3 medical assistants, 2 researchers, 2 informaticians, and 2 programmers) conducted a series of face-to-face meetings to (1) gather and understand user and technology requirements and workflows within ophthalmology and optometry and (2) define and analyze existing and proposed data elements for capture. The latter prompted content analysis of existing forms and a survey. Both of these activities are described in the following section. Requirements identified from the design team discussions were developed into use cases for prototype development. For example, the physicians identified the need to capture drawings of the eye within the prototype. An overview of the prototype development process is shown in Figure 1. An iterative clinical review process was used to refine the prototype. The developed prototype was then reviewed and considered for integration into the EHR via a prioritization process.

Figure 1. Process for developing the ophthalmologic data capture prototype.
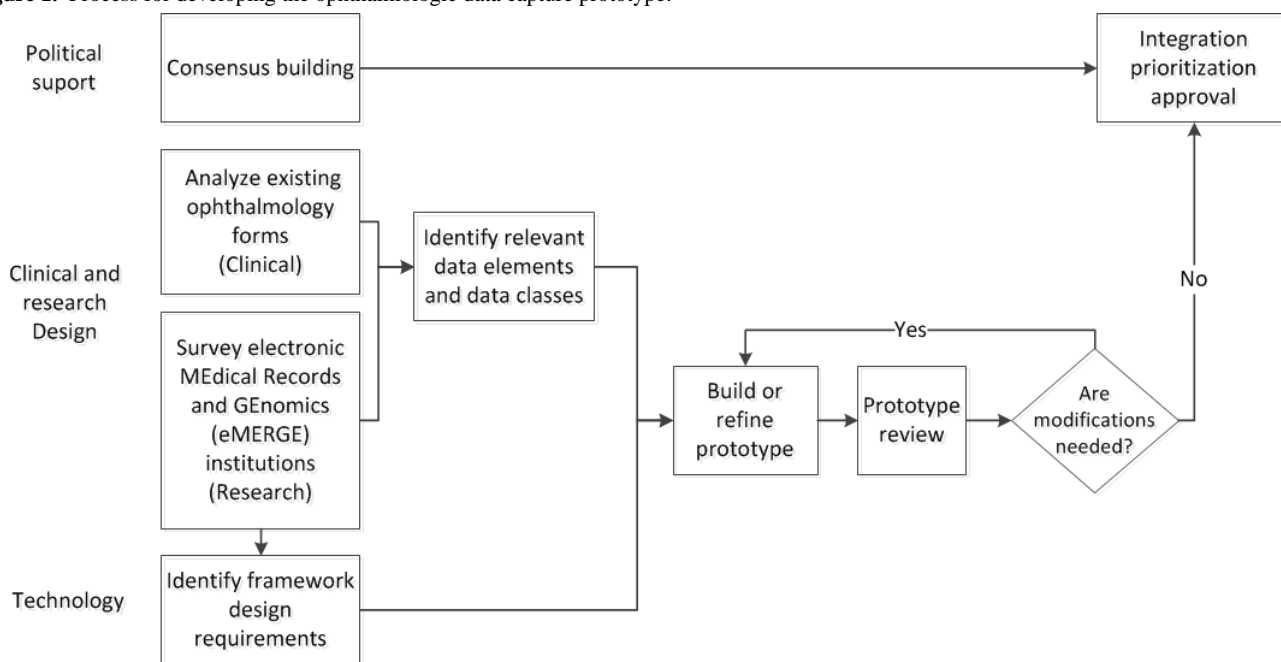
Clinical and Research Requirements

We conducted content analysis on 30 different handwritten unstructured paper forms used for documenting patient encounters within optometry and ophthalmology. Data elements nominated for prototype inclusion were prioritized in the following order: (1) data elements found on multiple forms or used by multiple practices and/or specialists; (2) data elements required for Meaningful Use [30]; and (3) data elements found on the Comprehensive Adult Medical Eye Evaluation developed by the American Academy of Ophthalmology (AAO) [25]. This list was then reviewed and prioritized for inclusion by the design team. In addition, each data element was defined and then grouped into a logical data class (ie, visit information, medical history, family history, examination, slit-lamp examination, specialty testing, and miscellaneous).

In a parallel effort, we distributed an EHR ophthalmologic data availability survey to institutions participating in the electronic MEdical Record and GEnomics (eMERGE) network [31] to gain an understanding of data elements that other institutions capture and to evaluate the generalizability and common types of ophthalmologic data elements collected across institutions and the mechanisms used for their capture. The survey can be found in Multimedia Appendix 1.

Technology and Architecture

A prototype of the ophthalmologic data capture application was developed on an extensible, open architecture using Microsoft's ASP.NET MVC 3.0, JavaScript, and jQuery. The development environment and database chosen for prototype development was based on the familiarity and experience of the programming team. An overview of the architecture is shown in Figure 2, and the database schema to support the data capture application can

be found in Multimedia Appendix 2. This architecture supports a *form specification database* (FSD), which includes form design specifications, actions, and Web control definitions. Form specifications can also be stored in XML configuration files if a database is not available or for lightweight implementations. Each data collection form is constructed using Web controls (hereafter referred to as a *control*), which allows a user to interact with the Web form . The FSD has many control records in the database to define a form both in appearance and actions. The FSD schema can store multiple forms and is extensible to accommodate additional control types that may be added at a later date.

The overall process for rendering a Web form is outlined in Figure 2 and is described below:

- A Web server generates the Web form based on specifications stored in the FSD and brokers data requests between the *EHR* or *flexible back-end database* (FBD) and the Web browser device (steps 1, 2, and 4).

- The Web server then requests patient identifiable information (PII) from the EHR or FBD to fill the Web form (step 3).
- The Web server generates the Web form, includes the PII, and presents it to the user (steps 4 and 5).
- User data and ink-over annotated drawings are captured and passed back to the EHR or FBD for storage (steps 6-8).

A collection of control classes are used to parse the XML input (retrieved from either the FSD or an XML configuration file) and then deliver the form content to the user via the Web browser. A control can be interactive (a form element that the user can act upon—CheckBox, TextBox, etc) or passive (eg, a container for grouping controls in a visual manner). Each control stored in the FSD contains a number of properties and methods as described in Table 1. These properties and methods define how the form will display and act when presented to the user. The Web form can dynamically change by modifying the controls and/or control properties and methods defined in the FSD or XML configuration file.

Table 1. Web control properties and methods.

Type	Web control class	Description
Properties		
	Name of control	A unique identifier for the specific element on the form. Multiple controls of the same type will have different names.
	Label	Human readable text to display to the user (not required).
	Control type	The type of the control (CheckBox, TextBox, ComboBox, image, etc). Once a <control/> element is found in the extensible markup language (XML), type property is checked to determine the control class to use.
	Value	The current value of the property the control represents. This value can be preloaded from the database or assigned as a result of the form submission.
	Children	The collection of child controls whose visibility is dependent on the value of the current control. For example, an ink-over image control can be displayed if a CheckBox is checked.
	Functions	The collection of events that define the logic for displaying or hiding child controls. This can be further expanded to handle non–child-related events (notifying the user, requesting more information from the server, etc).
Methods		
	Render hypertext markup language (HTML)	Each control is responsible for rendering itself. A CheckBox type control only knows how to render a CheckBox. This is also applicable for a TextBox, ComboBox, or any other basic or complex control type in the library. This means that a control is ignorant of its parent and sibling controls (if any) and only cares *where* its children are rendered, not *how*. This allows flexibility of form layout and control hierarchy.
	Render JavaScript	Whereas the HTML and JavaScript work hand-in-hand, the two are not usually rendered at the same place on the HTML page, so the two must be separated to meet the needs of a Web-based application.
	Render XML	After a user provides results and submits the form, the results are rendered back into XML format. That XML can then be sent to a service/database for storage or simply saved as a file on the local system. Similar to the Render HTML method details state, each control is responsible for how it's rendered in the resulting XML, whereas a CheckBox control's value may only be *checked* or *unchecked*; some other complex control will have a collection of results.

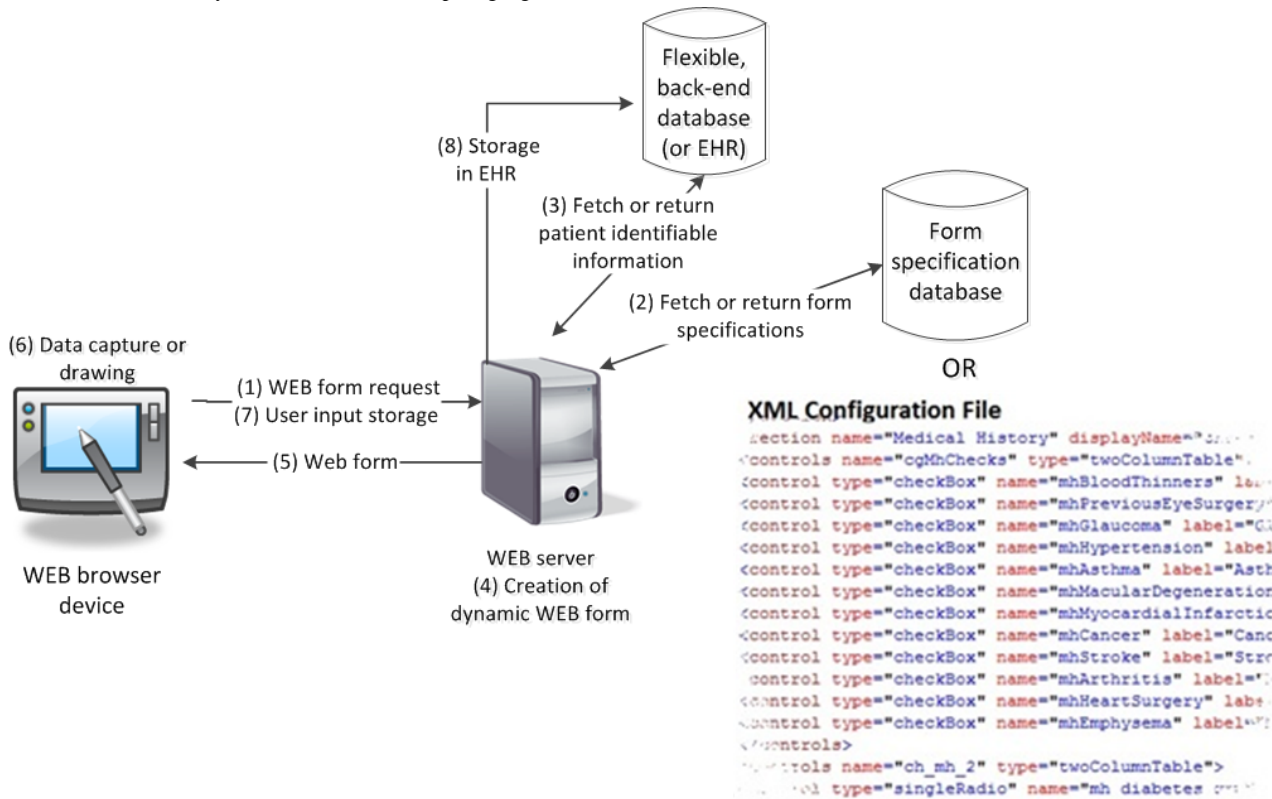**Figure 2.** Context-driven dynamic extensible markup language (XML) architecture.

Figure 3. Ophthalmology prototype with expanded slit-lamp exam section.

The ophthalmology data capture prototype is Web enabled to run on a variety of devices, including an iPad. The form is divided into seven logical sections for data entry. All the sections are initially collapsed for easy form navigation and then expanded as data collection ensues. Figure 3 shows the Web form with the section detail hidden, with the exception of the slit-lamp examination section.

Context-Driven Controls

The inclusion of context-driven controls is considered an important component of the application for several reasons. First, it reduces the need for a user to review and answer unnecessary questions. This helps to both guide and support a streamlined workflow for data entry. Figure 4 shows one section of the ophthalmology prototype that supports hidden controls. When a clinician selects a specialty test, a control is displayed and readied for handwritten comments. The corresponding XML used to generate the form is pictured in Figure 4. Second, it helps to keep the data entry form as compact as possible, removing unnecessary questions from the form, based on the answers to prior questions. For example, in Figure 5, we have a control for describing the macula or central retina. If the patient has a normal macula, Figure 5 (1) is only shown with no additional data entry required. The selection of a drop-down value for *Macula* (2) causes another form to appear, prompting the user to define the type of macular disorder, which, in this example, has been shown as age-related macular degeneration (AMD). The selection of AMD prompts additional controls (3) based on the attribute selected for *AMD Type*, thus resulting in additional relevant data capture. Currently, the application allows for very basic context-driven behavior (showing/hiding child controls based on a selected value). This is done by making use of a dependency property associated with the control and using naming conventions within the Form Specification Database (Figure 2). Figure 5 also shows the corresponding XML code that demonstrates this concept. To display a control when a value from a drop-down list is selected, the name of that control needs to match the value of the item on which it is dependent. The XML code is automatically generated based upon the records found in the Form Specification Database (Figure 2). Users are able to customize and/or add new form features by either adding or removing records or changing the attributes of a data element within the database.

Figure 4. Specialty control and corresponding extensible markup language (XML) code.

Figure 5. Example for adding the show/hide event to a dependent control and corresponding extensible markup language (XML) code.


Ink-Over Capability

Ophthalmologists require tools to illustrate ocular abnormalities [28]. A key feature of this application is the ability to draw on a Web form or canvas. The canvas area is overlaid with a transparent image, and the user can annotate as they see fit on the transparent image as shown in Figure 6. The images can be specific to the context of the form, as in the ink-over Web control for annotating the eye or simply a series of horizontal lines for writing text annotations (Figure 6). Given Marshfield

Clinic Health System's information technology environment (Tablet PCs and Windows 7/Internet Explorer [IE] 8), we focused on embedding the Microsoft.Ink library into the Web form to provide native ink-over functionality to the users. For users accessing the form on a device with browsers such as IE 9+, Chrome, Firefox, Safari, and others that do not have the Microsoft.Ink libraries installed, HTML5 Canvas was used. The resultant images are saved along with the coded data to the EHR or FBD (Figure 2).

Figure 6. Examples of ink-over controls.
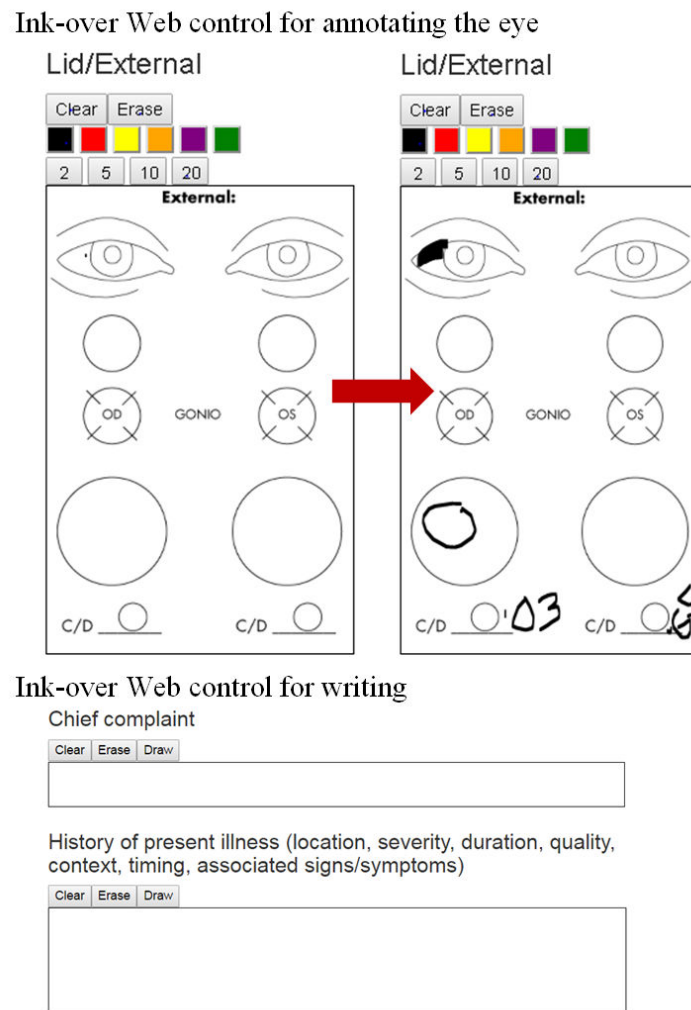
Prototype Evaluation

We conducted a heuristic evaluation of the software following the development of the prototype. Heuristic evaluation is a usability inspection method employed to discover issues with a user interface. A usability analyst at the Marshfield Clinic Research Institute reviewed the ophthalmology data capture prototype multiple times, comparing it with a list of established heuristics, recording and reporting all conflicts with the heuristics [32].

Results

The most important design criteria prioritized by clinical, research, and technical stakeholders during the interview and survey processes are presented in Table 2. Ophthalmologists considered the most important requirements for an interface, which included an interface that is (1) easy to navigate, (2) integrated into clinical workflow, (3) organized into a consolidated view and meaningful groups (meaning all ophthalmologic information is located in one place so clinicians can view it easily, and groupings represent visit reason, medical and family history, examination, and specialty testing), (4) EHR integrated to prefill the forms with relevant patient information, and (5) ink-over form writing capable to support ocular illustrations. The important technical requirements focused on integration and maintenance functionality. Research requirements highlighted coded data capture and sharing of the prototype technology with other collaborators. All stakeholders mentioned the need for a secure data sharing and authentication architecture. Our solution addresses all of these requirements.

Table 2. Stakeholder requests that influenced prototype design.

Item	Stakeholder	Description
Consolidated form view	Clinical	Consolidate the 30 forms into a single form, with logical groupings for medical and family history, visit information, examination, and slit-lamp examinations. Collect data once and display in multiple views.
Electronic health record (EHR) integration	Clinical	Prefill form with EHR data, including patient identifiers and patient and family history. Data captured in the form should be sent to the clinical data repository of the EHR for use in patient care and research.
Multiple data capture formats	Clinical and Research	System supports Boolean (true/false), numeric, coded, textual, and analog (graphical) data inputs. Allows for flexibility of data capture while still maintaining the ability to capture discrete and disparate data types.
Context-driven controls	Clinical	Behavior of a form control can be determined by the values of another control. Facilitates conditional questions to be enabled, as necessary, based on the answers to other questions and gives the form a neat and organized appearance.
Metadata driven	Technical	Form specifications should be easily amendable to accommodate new or modified data capture requirements and form layout changes among practices. Properties of the dynamic Web controls can be easily modified in the database, thus allowing changes to the extensible markup language (XML) specifications.
Dynamic form generation	Technical	Dynamic generation of the form, based on a form definition. Storage of multiple types or versions of forms.
Flexible, extensible back-end database	Technical	Back-end database will support storage and retrieval of multiple data types, allowing for capture and storage of discrete, disparate data types.
Service-oriented architecture (SOA) and database agnostic	Technical	Utilize SOA. The SOA layer will allow for enhanced security and consistency in data transfers.
System portability	Research	Architecture used for this development can either be a stand-alone system or have the ability to integrate into an EHR.
Open source	Research	Architecture should be shareable with other research sites that do not have access to ophthalmologic data collection mechanisms.
Secure data sharing/Authentication	All	Architecture should support integrated user login and secure data sharing transparency between the application and EHR.

Content Analysis

Unique data elements, identified from over 30 different clinical ophthalmology forms, were reviewed by clinicians and specialty leadership. Several collaborative discussions were held to organize the data elements into seven general data classes to support common ophthalmologic workflows as shown in Table 3. Ten data elements important to research were identified within these data classes (indicated with superscript letter *a* within Table 3). Medical and family history related data elements will likely require EHR access to display previously collected information. An API to the EHR can be used to collect extant data to prefill the form.

eMERGE Data Availability Survey

Out of 9 eMERGE sites, 6 sites (66%; Geisinger Health System, Marshfield Clinic, Mayo Clinic, Mount Sinai, Northwestern University, and Vanderbilt University) completed the *Ophthalmology/Optometry Data Availability in the Electronic Health Record* survey found in Multimedia Appendix 1. Out of these 6 eMERGE sites, 4 sites (66%) used commercially developed EHRs (Epic and General Electric's Centricity), and 2 sites used in-house developed EHRs (Vanderbilt and Marshfield's CattailsMD). A summary of the survey responses can be viewed in Table 4. The ophthalmologic information from the respondents' EHRs is captured in a variety of data formats, including coded, XML, text, and images. Many of the data elements identified by Marshfield are also being collected at other eMERGE institutions. As expected, there is only a limited amount of coded ophthalmologic data captured in the EHR among the eMERGE institution respondents.

Table 3. Data class groupings and data elements important to ophthalmologic health care and research.

Data class	Types of data elements		
Visit information	Chief complaint	History of present illness	
Medical history	Blood thinners	Asthma	Diabetes
	Previous eye surgery	Cancer	Smoking
	Macular degeneration	Stroke	Alcohol
	Glaucoma	Arthritis	Mental status
	Hypertension	Heart surgery	Occupation
	Myocardial infarction	Emphysema	
Family history	Blindness	Glaucoma	Macular degeneration
Examination	Visual acuity[a]	Medial rectus; oculus dexter (OD or right eye)	Intraocular pressure[a]
	Pinhole	Medial rectus; oculus sinister (OS or left eye)	Visual field defect[a]
	Current glasses	Lid/External	Pupils
	Current contacts		
Slit-lamp exam general	Conjunctiva	Anterior chamber	Posterior vitreous
	Corneal epithelium	Iris	Cup-disc ratio/grade[a]
	Corneal stroma	Lens-cataract grade and type[a]	Disc color
	Corneal endothelium	Anterior vitreous	Disc classification
	Corneal tear film		
Slit-lamp exam fundus	View	Vessels	Periphery
Slit-lamp exam macula	Normal	Diabetic retinopathy[a]	Age-related macular degeneration includes presence and type of drusen[a]
Specialty testing	Schirmer test[a]	Rose Bengal staining[a]	Tear breakup time[a] (TBUT)
Miscellaneous	Impression	Return to clinic	Additional workflow coordination notes
	Recommendation		

[a]indicates that the data element is deemed important for research activities by research stakeholders.

Prototype Evaluation

A heuristic evaluation of the prototype revealed that the form most often displayed inconsistent use of controls. When given the choice of the same number and types of options, the form contained both radio buttons and drop-down controls. For consistency and efficiency of use, the usability analyst recommended implementing radio buttons in all cases so that direct selection with a stylus was possible. The analyst also recommended that all text boxes have an erase toggle, rather than the existing erase button so that users can go back and forth between writing and erasing. The form originally prevented users from continued writing in text boxes after clicking the erase button. It required that the text box be cleared before accepting input. Minor recommendations included highlighting selected functionality, increasing padding around buttons and controls (radio buttons, drop-down boxes, text boxes, etc), and indenting dependent child controls to show relation to their parent control.

Table 4. The electronic MEdical Record and GEnomics (eMERGE) institution responses to ophthalmologic data availability in their electronic health records (n=6).

Data elements captured for right/left eyes	EHRs that capture data element, %	EHRs that capture coded/extensible markup language (XML) data elements, %	Data capture formats
Visual acuity	83	50	Coded or XML, image, text
Intraocular pressure	66	33	Coded or XML, image, text
Fundus exam	66	33	Coded or XML, image, text
Visual field exam	66	16	Image, text
Optical coherence tomography (OCT)	66	16	XML, image, text
Cup-disc ratio	83	50	Coded or XML, image, text
Presence of drusen	83	33	Coded, image, text
Soft drusen	83	33	Coded, image, text
Hard drusen	83	33	Coded, image, text
AMD (age-related macular degeneration) staging severity	50	16	Image, text
Severity of diabetic retinopathy	66	33	XML, image, text
Macular edema	83	16	Coded or XML, image, text
Severity of cataract	33	0	Image, text
Brightness acuity	50	16	Coded, image, text
Schirmer test (value)	50	16	Coded, image, text
Rose Bengal staining	50	0	Image, text
Tear breakup time (BUT)	33	0	Image, text
BUT measurement method	33	0	Image, text

Discussion

Principal Findings

This research describes the creation of a context-driven, dynamic XML data capture prototype for ophthalmological care and research. Its open architecture allows the use of a service-oriented architecture (SOA), which can facilitate integration with a variety of EHRs for retrieval of patient-specific information and the transfer of newly collected information back to the EHR, thereby making the data available for other uses. The architecture also supports Web-based forms that can be created dynamically from a database or an XML schema, context-based controls for efficient data entry, and ink-over forms for illustrating abnormalities of the eye. The source code, along with a demonstration version of the ophthalmology prototype, can be found on the Marshfield Clinic Research Institute's website [33].

Comparison With Prior Work

We investigated several well-known data capture solutions to use for ophthalmology prototype development [10-15,19]. Many of the solutions had limitations with inbound and/or outbound EHR data flows, supporting conditional logic, ink-over drawing capability, and licensing for operational use [19]. With the unique challenges presented in ophthalmology, we chose Web-based technology for prototype development because it minimized problems when running the application on multiple types of devices. The choice of this technology also provided the ability to share the prototype with other institutions by leveraging SOA for EHR integration.

We encountered several challenges. First, ophthalmologic workflows are complicated, as indicated by the current use of over 30 different paper-based forms. Creating a single, easy-to-navigate, Web-based user interface required analyzing over 140 unique data elements and employing several previously described methods to reduce the number of items and to determine logical groupings of data classes for efficient workflow. Form design was problematic because of the large number of data elements required for data capture. We developed a series of context-based controls to hide data capture complexity.

Second, the XML architecture used to support context-driven controls must be adaptable to changing data collection needs and new controls. To address this, we added attributes to each control within the database, indicating whether a control was dependent on another and the action needed to invoke the control. Using this approach enables one to easily modify form actions by changing database control configurations.

Third, providing Web-enabled drawing capability for eye abnormalities introduced some additional challenges. Our design required provisions for native ink-over drawing functionality in a large variety of Web-enabled devices, supporting a wide range of client configurations. We utilized Microsoft.Ink library

XSL•FO

RenderX

for devices within Marshfield Clinic Health System's computing environment but had to investigate other options for non-Marshfield devices with other operating system or browser requirements. We designed the ink-over capability to be device agnostic (functional on tablets, laptops, desktops, and mobile phones) as long as the administrative user defines a form with the appropriate ink-over controls. A significant effort was expended to support browser version detection and Microsoft.Ink compatibility.

The ophthalmology data capture prototype is currently packaged as a stand-alone application for demonstration purposes. We envision that the prototype will be packaged as an EHR add-on for use in ophthalmology and optometry departments. The prototype is built on SOA, and the architecture promotes context awareness and supports the transition of data between the application and EHR in a secure manner. This prototype currently uses a flexible back-end database—MySQL (Figure 2). This stand-alone database could easily be transitioned into an EHR data repository with the development of a wrapping Web service to broker EHR data exchange to and from the services of this prototype system. This level of integration would allow patient information from the EHR to be prefilled in the application's forms and minimize data entry for the user.

Limitations

The data capture application was reviewed by several ophthalmologists throughout the development process, and suggestions were provided for terminology and logical groupings of data elements and form flow. During the prototype development, we did not conduct a formal usability evaluation, but we did conduct a heuristic evaluation. Heuristic evaluations are usually hampered by the fact that reviewers who conduct them are not experts in the field that a user interface covers. In this case, a subject-matter expert was involved throughout the data gathering and design phases of development, so this evaluation may be less limited than others because of this previously contributed expertise. The modular nature of our application architecture permitted us to address the findings of the heuristic evaluation quickly. We were able to modify the type of controls used to provide consistent controls, and the enhancement to the text entry control was instantly applied across all forms using it.

Data elements collected using this application can be defined and annotated using standards set forth by the AAO, Clinical Data Interchange Standards Consortium (CDISC), or health care data standards consortium. Future application enhancements could include APIs that interface with common terminology databases or management systems.

Finally, we did not implement audit trails or identification management or explore the use of data entry error–checking algorithms within this prototype, as it was a proof-of-concept project and meant to explore the possibilities of dynamic form generation and context awareness for data collection. Future development will include this functionality.

Conclusions

This research describes the creation of an open-source, context-driven, structured data capture dynamic XML ophthalmologic data capture application that can be integrated into a variety of EHRs. Relevant ophthalmologic and optometric data elements were identified for clinical care and research. Data entry was streamlined using context-driven controls and ink-over capabilities for illustrating eye abnormalities.

Conflicts of Interest

None declared.

Multimedia Appendix 1

eMERGE Survey - Ophthalmology/Optometry data availability in the electronic medical record.

[PDF File (Adobe PDF File), 297KB - medinform_v5i3e27_app1.pdf]

Multimedia Appendix 2

Database schema used to support the extensible markup language (XML)-driven data capture framework.

[PDF File (Adobe PDF File), 183KB - medinform_v5i3e27_app2.pdf]

References

1. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. Biomed Res Int 2015;2015:370194 [FREE Full text] [doi: 10.1155/2015/370194] [Medline: 26229957]

2. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful optimization of large-scale health information technology. J Am Med Inform Assoc 2017 Jan;24(1):182-187. [doi: 10.1093/jamia/ocw037] [Medline: 27107441]

3. HealthIT. System interfaces URL: https://www.healthit.gov/safer/guide/sg005 [accessed 2017-02-07] [WebCite Cache ID 6o6Bw0Dig]

4. Cresswell K, Sheikh A. Health information technology in hospitals: current issues and future trends. Future Hosp J 2015 Feb;2(1):50-56 [FREE Full text] [doi: 10.7861/futurehosp.15.015]

5. Clarke S, Wilson ML, Terhaar M. Using clinical decision support and dashboard technology to improve heart team efficiency and accuracy in a transcatheter aortic valve implantation (TAVI) program. Stud Health Technol Inform 2016;225:98-102. [Medline: 27332170]

6. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]

7. 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 16;144(10):742-752. [Medline: 16702590]

8. Chen J, Dougherty E, Demir S, Friedman C, Li CS, Wong S. Grand challenges for multimodal bio-medical systems. IEEE Circuits Syst Mag 2005;5(2):46-52. [doi: 10.1109/MCAS.2005.1438739]

9. Coulter A, Locock L, Ziebland S, Calabrese J. Collecting data on patient experience is not enough: they must be used to improve care. BMJ 2014 Mar 26;348:g2225. [doi: 10.1136/bmj.g2225]

10. SoftwareGeek. Formtools URL: http://www.softwaregeek.com/download/form_tools.html [accessed 2017-08-10] [WebCite Cache ID 6scXbiHdn]

11. Sybase. PowerBuilder web forms applications URL: http://infocenter.sybase.com/help/index.jsp?topic=/com.sybase.infocenter.dc00586.1250/doc/html/hfr1272397208427.html [accessed 2017-08-10] [WebCite Cache ID 6scXe8wnS]

12. Oracle. Oracle forms services 11g URL: http://www.oracle.com/technetwork/developer-tools/forms/overview/index-098877.html [accessed 2017-08-10] [WebCite Cache ID 6scXfBehr]

13. Arclab. Arclab web form bilder 3 URL: https://www.arclab.com/en/webformbuilder/updates.html [accessed 2017-08-10] [WebCite Cache ID 6scXoRS12]

14. FormDocs. 2014. FormDocs healthcare and medical forms URL: http://www.formdocs.com/medical-healthcare-patient-form.htm [accessed 2017-08-10] [WebCite Cache ID 6scXrXZxI]

15. Zoho. 2014. Online forms for every business URL: https://www.zoho.com/forms/ [accessed 2017-08-10] [WebCite Cache ID 6scXtPjiO]

16. Basch E, Artz D, Iasonos A, Speakman J, Shannon K, Lin K, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. J Am Med Inform Assoc 2007;14(3):264-268 [FREE Full text] [doi: 10.1197/jamia.M2177] [Medline: 17329732]

17. El FA, Lucas N, Rance B, Verplancke P, Lastic P, Daniel C. The REUSE project: EHR as single datasource for biomedical research. Stud Health Technol Inform 2010;160(Pt 2):1324-1328. [Medline: 20841899]

18. Mathura VS, Rangareddy M, Gupta P, Mullan M. CliniProteus: a flexible clinical trials information management system. Bioinformation 2007;2(4):163-165 [FREE Full text] [Medline: 21670796]

19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

20. Collins M, Ross E, Meropol NJ, Lazev AB. Using metadata to generate web-based electronic data capture forms. AMIA Annu Symp Proc 2006:896 [FREE Full text] [Medline: 17238515]

21. Pyles LA, Hines C, Patock M, Schied M, Chase J, Jamrozek K, et al. Development of a web-based database to manage American college of emergency physicians/American academy of pediatrics emergency information forms. Acad Emerg Med 2005 Mar;12(3):257-261 [FREE Full text] [doi: 10.1197/j.aem.2005.01.001] [Medline: 15741591]

22. Garwood KL, Taylor CF, Runte KJ, Brass A, Oliver SG, Paton NW. Pedro: a configurable data entry tool for XML. Bioinformatics 2004 Oct 12;20(15):2463-2465 [FREE Full text] [doi: 10.1093/bioinformatics/bth251] [Medline: 15073025]

23. Chiang MF, Boland MV, Brewer A, Epley KD, Horton MB, Lim MC, American Academy of Ophthalmology Medical Information Technology Committee. Special requirements for electronic health record systems in ophthalmology. Ophthalmology 2011 Aug;118(8):1681-1687. [doi: 10.1016/j.ophtha.2011.04.015] [Medline: 21680023]

24. Safir A. Editorial: computers in ophthalmology. Invest Ophthalmol 1976 Mar;15(3):163-168. [Medline: 767282]

25. Boland MV, Chiang MF, Lim MC, Wedemeyer L, Epley KD, McCannel CA, American Academy of Ophthalmology Medical Information Technology Committee. Adoption of electronic health records and preparations for demonstrating

meaningful use: an American Academy of Ophthalmology survey. Ophthalmology 2013 Aug;120(8):1702-1710. [doi: 10.1016/j.ophtha.2013.04.029] [Medline: 23806425]

26. Sanders DS, Lattin DJ, Read-Brown S, Tu DC, Wilson DJ, Hwang TS, et al. Electronic health record systems in ophthalmology: impact on clinical documentation. Ophthalmology 2013 Sep;120(9):1745-1755. [doi: 10.1016/j.ophtha.2013.02.017] [Medline: 23683945]

27. Rasmussen LV, Peissig PL, McCarty CA, Starren J. Development of an optical character recognition pipeline for handwritten form fields from an electronic health record. J Am Med Inform Assoc 2012 Jun;19(e1):e90-e95 [FREE Full text] [doi: 10.1136/amiajnl-2011-000182] [Medline: 21890871]

28. Corcoran S. Ophthalmology Management. 2010. Billing for extended ophthalmoscopy URL: http://www.ophthalmologymanagement.com/issues/2010/october-2010/coding-reimbursement [accessed 2017-08-10] [WebCite Cache ID 6scYCEtxO]

29. Zvornicanin J, Zvornicanin E, Sabanovic Z. Ophthalmology and information technology in tuzla canton health care system. Acta Inform Med 2012 Jun;20(2):90-93. [doi: 10.5455/aim.2012.20.90-93] [Medline: 23322959]

30. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]

31. McCarty C, Chisholm R, Chute C, Kullo I, Jarvik G, Larson E, eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics 2011 Jan 26;4:13 [FREE Full text] [doi: 10.1186/1755-8794-4-13] [Medline: 21269473]

32. Molich R, Nielsen J. Improving a human-computer dialogue. Commun ACM 1990;33(3):338-348. [doi: 10.1145/77481.77486]

33. owf.mfldclin. 2017. Ophthalmology Web Form (OWF) URL: http://owf.mfldclin.edu/ [accessed 2017-08-23] [WebCite Cache ID 6svXiITpP]

Abbreviations

AAO: American Academy of Ophthalmology
AMD: age-related macular degeneration
API: application program interface
CDISC: Clinical Data Interchange Standards Consortium
CDXA: context-driven dynamic XML architecture
CTSA: Clinical and Translational Science Award
EHRs: electronic health records
eMERGE: electronic MEdical Record and Genomics
FBD: flexible back-end database
FSD: form specification relational database HIPAA: Health Insurance Portability & Accountability Act
HTML: hypertext markup language
NCATS: National Center for Advancing Translational Sciences
NHGRI: National Human Genome Research Institute
NIH: National Institutes of Health
OCT: optical coherence tomography
PCs: personal computers
PII: patient identifiable information
SOA: service-oriented architecture
TBUT: tear breakup time
XML: extensible markup language

XSL•FO

RenderX

Original Paper

Use of Simulation Based on an Electronic Health Records Environment to Evaluate the Structure and Accuracy of Notes Generated by Medical Scribes: Proof-of-Concept Study

Robert Pranaat[1], MD; Vishnu Mohan[1], MD; Megan O'Reilly[2], MD; Maxwell Hirsh[3], BS; Karess McGrath[4], BA; Gretchen Scholl[3], BS; Deborah Woodcock[1], BS, MBA; Jeffrey A Gold[4], MD

[1]Medical Informatics, Oregon Health & Sciences University, Portland, OR, United States
[2]Obstetrics and Gynecology, Oregon Health & Sciences University, Portland, OR, United States
[3]School of Medicine, Oregon Health & Sciences University, Portland, OR, United States
[4]Pulmonary Critical Care, Oregon Health & Sciences University, Portland, OR, United States

Corresponding Author:
Jeffrey A Gold, MD
Pulmonary Critical Care
Oregon Health & Sciences University
3181 SW Sam Jackson park rd
Portland, OR, 97239
United States
Phone: 1 5034181496
Email: goldje@ohsu.edu

Abstract

Background: The increasing adoption of electronic health records (EHRs) has been associated with a number of unintended negative consequences with provider efficiency and job satisfaction. To address this, there has been a dramatic increase in the use of medical scribes to perform many of the required EHR functions. Despite this rapid growth, little has been published on the training or assessment tools to appraise the safety and efficacy of scribe-related EHR activities. Given the number of reports documenting that other professional groups suffer from a number of performance errors in EHR interface and data gathering, scribes likely face similar challenges. This highlights the need for new assessment tools for medical scribes.

Objective: The objective of this study was to develop a virtual video-based simulation to demonstrate and quantify the variability and accuracy of scribes' transcribed notes in the EHR.

Methods: From a pool of 8 scribes in one department, a total of 5 female scribes, intent on pursuing careers in health care, with at least 6 months of experience were recruited for our simulation study. We created three simulated patient-provider scenarios. Each scenario contained a corresponding medical record in our simulation instance of our EHR. For each scenario, we video-recorded a standardized patient-provider encounter. Five scribes with at least 6 months of experience both with our EHR and in the specialty of the simulated cases were recruited. Each scribe watched the simulated encounter and transcribed notes into a simulated EHR environment. Transcribed notes were evaluated for interscribe variability and compared with a gold standard for accuracy.

Results: All scribes completed all simulated cases. There was significant interscribe variability in note structure and content. Overall, only 26% of all data elements were unique to the scribe writing them. The term *data element* was used to define the individual pieces of data that scribes perceived from the simulation. Note length was determined by counting the number of words varied by 31%, 37%, and 57% between longest and shortest note between the three cases, and word economy ranged between 23% and 71%. Overall, there was a wide inter- and intrascribe variation in accuracy for each section of the notes with ranges from 50% to 76%, resulting in an overall positive predictive value for each note between 38% and 81%.

Conclusions: We created a high-fidelity, video-based EHR simulation, capable of assessing multiple performance indicators in medical scribes. In this cohort, we demonstrate significant variability both in terms of structure and accuracy in clinical documentation. This form of simulation can provide a valuable tool for future development of scribe curriculum and assessment of competency.

XSL•FO
RenderX

Introduction

The electronic health record (EHR) is a vital tool in the delivery of clinical care. The EHR adoption rates have grown rapidly largely because of government programs such as the Health Technology for Economic and Clinical Health (HITECH) Act of 2009 [1]. However, physician dissatisfaction with EHRs remains high, a phenomenon probably linked to the perception that EHRs do not improve efficiency (42%), do not decrease workload (72%), have increased total operating costs (54%), and have yet to overcome operating challenges (43%) [2].

One key factor that contributes to the dissatisfaction is the paradigm of "information chaos" resulting from EHR use that can lead to impaired situational awareness and increased mental workload [3]. To amplify this paradigm, a number of studies conducted by our group and others have suggested that providers across multiple professions have difficulty in using the EHR as manifested by issues with data finding, recognition of patient safety issues, and impairment in clinical decision making [4,5]. Additionally, recent studies revealed that problems associated with clinicians' selective data gathering or selective data interpretation can lead to increased patient harm, a phenomenon that has also been identified and replicated in simulation exercises [6,7]. These issues are not just isolated to physicians: recent work from our group has suggested that the phenomenon affects nurses and pharmacists at all levels of training, implying a global problem related to human EHR interfaces [5,8].

Growing concerns with EHR usability and efficiency have been mirrored by concomitant increased utilization of medical scribes. To alleviate challenges associated with EHR data entry, physicians have increasingly incorporated scribes into clinic and hospital workflows. Though studies lauding their potential benefits have been present for nearly 30 years, recently the scribe workforce has demonstrated a significant and rapid growth; there were approximately 10,000 scribes working in 2014 with a projection of 20,000 scribes in the workforce by 2016 [9,10]. However, whereas the number of scribes has increased dramatically, there still exists no standardized approach for training and assessing scribes. Before being embedded within a practice, scribes have varied levels of clinical exposure and disparate degrees of training varying from formal EHR training by employers or scribe organizations to Web-based courses by commercial scribe solution organizations to ad hoc training conducted by clinicians to no training at all. This often creates an interesting paradox: most physicians feel that their own training with the EHR is inadequate and their need for utilizing scribes arises from their inability to use the EHR in a safe and efficient manner [2,11,12]. Yet, these physicians may then be responsible for training and assessing scribes who have had often little to no direct health care experience themselves.

Scribes who use the EHR may find the complex interface and usability constraints of the EHR potentially even more challenging than physicians do because they lack clinical learning and EHR-specific workflow training. In essence, this paradigm adds another layer of physician responsibility but does not eliminate the errors inherent with poor EHR use.

These issues are further magnified by the fact that scribes do not necessarily just engage in data entry activities during the clinical encounter but may also have a variable and expanded role at the discretion of the provider they are scribing for [13,14]. Currently, the only defined regulatory guidance for scribe use comes from The Joint Commission, which deems that medical scribes are to "chart at the direction of their provider" and should not place orders. Furthermore, physicians are required by the Joint Commission to authenticate, or attest to, all notes written by scribes [14].

To ensure that standardized activities are accomplished, scribes require appropriate training that directly links their learning needs with measured outcomes. This can be accomplished through training regimens that evaluate individual competencies pertinent to accurate EHR documentation. Training should maintain Health Insurance Portability and Accountability Act (HIPAA) compliance and ensure patient safety. Given the relationship between communication errors and patient safety [15], scribes' role in EHR documentation stands to benefit from training that does not endanger patient well-being.

On the basis of these concerns, it is imperative that methodology exists to ensure that scribes can be effectively trained and their competency assessed for safe and effective use of EHR in the appropriate clinical settings. Simulation has been a means of evaluating complicated systems, while posing no risk to patients, and providing high-fidelity standardized subject experiences [4,5]. Recently, we demonstrated that EHR-based simulation could be used to assess the creation and accuracy of both intern progress notes and admission history and physicals [16,17]. Given that high-fidelity simulation is effective with regard to facilitating improved EHR use for multiple clinical professions such as physicians, nurses, and pharmacists, it seems logical that similar techniques would also be effective for scribes, whose role as EHR documentation experts essentially replaces these same skills by physicians. Therefore, our hypothesis is that through the use of high-fidelity simulated provider-patient encounters and integrated EHR, it is possible to assess scribes' EHR use in similar fashion.

Methods

The study was approved by the institutional review board of the Oregon Health & Science University. All data were deidentified and stored securely.

Simulation Creation and Materials

Three Obstetrics-Gynecology (Ob-Gyn) scenarios were created by a clinical subject matter expert (Ob-Gyn attending physician) to represent standard ambulatory encounters. We created a replica of each clinical case in our simulation instance of EpicCare (Epic Systems) using techniques we have described in previous publications [4,18]. Briefly, the EHR instance

utilized for simulation activities is created from a "clone" of the clinical system, maintaining all user customizations, shortcuts, and macros. The instance contains only patient charts representing the simulation; it does not contain protected health information of real patients in our health care system. Given the need for any simulation-based training exercise to be both scalable and accessible from a variety of clinical environments, we decided to use a virtual video-based simulation. For each scenario, we video-recorded a standardized patient-provider encounter, with medical personnel serving in roles of both physician and patient. Once recorded, each video was cropped and edited to ensure adequate audio and video quality. On the basis of the script of each scenario, a "gold-standard" note was created for each case to allow for assessment of accuracy of content of individual scribe notes.

Subject Recruitment and Characteristics

A list of all medical scribes was collected from the Scribe Program Supervisor of the OHSU medical scribing program. Medical scribes working at the OHSU Center for Women's Health (CWH) were selected because they represented the largest proportion of all medical scribes working at OHSU. They were approached via email, phone texts, and phone calls to arrange simulation participation times. All scribes had a minimum of 1 year of scribe experience and minimum 6 months of experience scribing for CWH before study participation.

Simulation Procedure

In order for the simulations to accurately replicate scribes' work environment in real-world settings, the activity was conducted at the CWH, OHSU. For each simulated case, subjects were instructed to (1) familiarize themselves with each simulated patient chart before beginning the simulated physician-patient video, and (2) perform scribe activities in simulation just as they would during a real physician-patient interaction.. Simulations were performed in patient exam rooms at the CWH, OHSU that replicated real-world conditions accurately. Videos were displayed from a laptop computer on the exam table. Scribes used dedicated exam-room computers. The standardized narrative was read aloud to each scribe. Each simulation lasted between 6 and 18 min and scribes performed all three cases, in the same order.

Data Collection

Scribe- and physician-created notes were transferred from the Epic simulation environment into Pages (Apple Inc). Screenshots were taken of the Encounter, Labs, and Imaging tabs of Chart Review to determine whether the orders were pended. The gold-standard note was transferred from the Epic simulation environment into Pages in the same manner.

Data Analysis

Scribe notes were evaluated for note length, word economy, data elements, copy and paste blocks, pended orders, and attestations. These structural elements were compared with each other to determine interscribe variability. Structural elements were also compared with our gold-standard note to determine accuracy and positive predictive value (PPV). PPV was defined as the ratio of scribe's data elements also found in the gold-standard note to all those data elements included by the scribe. Data elements were defined as the individual positive and negative facts created by the scribe or gold standard from each of the patient-physician videos and provided resources. Data elements represented the interpretation of the scribe and the gold standard with respect to what was verbalized and performed during the encounter. Data elements were tabulated by note section, subjective, objective, or assessment and plan. The presence of copy and pasted blocks was determined using Plagiarism Checker X (Plagiarism Checker X, LLC), a plagiarism detection software package. Word economy was defined as the number of words required to create 1 data element or the number of words divided by data elements. Attestations were considered present if the medical-scribe included a statement at the end of their note signifying that they were a scribe working on behalf of the physician-provider.

Results

We first wanted to determine the general structure and interscribe variability determined by data elements, note length, word economy, pended orders, attestations, and the specific structure of each note section. A total of 150, 183, and 118 unique data elements were found in case 1, case 2, and case 3, respectively (Figure 1). Upon examining interscribe variability in elements, there was a 2- to 4-fold range in the number of data elements present for each range of data elements among the 5 scribes.

We next sought to determine the commonality of data elements between scribes. For each scribe, for a given element, we determined what fraction of the total cohort of scribes documented this element in their note for and individual case. Data from all three cases were then pooled for analysis. We further subdivided the analysis to the three main sections: Subjective, Physical exam, and Assessment and plan (Figure 2). Of interest, in the subjective section, less than 25% of data elements in an individual scribes' note were represented in all 5 of the notes, whereas almost 20% were unique to the individual scribe. Further, when analyzing the physical exam, scribe 3 and 4 documented elements that were not present in the simulation for case 3, explaining the inability of notes from the remaining scribes to have any elements present in 100% of the cohorts' note. Overall, 26% of all scribe-created data elements were unique to individual scribes, whereas 17% of all data elements created by scribes received complete agreement.

These differences in note elements were associated with significant variability in global note structure and content. There was almost an 87-fold difference in note length in case 1 between the high and low, 55-fold difference in case 2, and 115-fold difference in case 3. Of note, variance was observed across all structural domains of the note (Figure 3). In case 1, the shortest note was 37% (293/794) of the longest note, in case 2, it was 57% (251/440), and in case 3, the shortest note represented 31% (94/302) of the length of the longest note.

Finally, we wished to determine differences in the general structure of scribes' note with that of the gold-standard note. Errors of omission were demonstrated by calculating for accuracy, that is, the frequency by which scribes included all the data elements that were found in the gold-standard note.

Similarly, errors of commission were demonstrated through the use of PPV, whereby we were able to calculate how often scribes in our study included information that was not present, and therefore assumed to be inaccurate, in the gold-standard note. Individual scribe accuracy ranged from 50% to 76%, whereas the accuracy of subjective, objective, and assessment and plan was 72%, 60%, and 56%, respectively. For individual scribes the PPV ranged from 38% to 81%. When scribe notes were averaged, the PPV of subjective, objective, and assessment and plan was 54%, 52%, and 69%, respectively (Table 1).

Table 1. Accuracy and Positive Predictive Value (PPV) for each simulated case by structural element.

Note section	Case #1					Case #2					Case #3				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
True Positive															
Subjective	16	34	31	24	10	23	33	16	15	12	4	10	9	8	8
PE	14	15	4	1	0	12	13	4	8	6					
A&P	6	7	3	4	2	13	9	16	15	11	2	4	2	2	1
False Positive															
Subjective	6	11	13	10	3	7	12	15	9	8	0	1	9	6	1
PE	3	2	4	3	4	1	3	4	3	4					
A&P	3	2	14	0	4	3	2	3	5	2	7	6	5	7	3
False Negative															
Subjective	34	16	19	26	40	28	18	35	36	39	2	2	2	2	2
PE	2	1	12	15	16	2	1	10	6	8					
A&P	2	1	2	3	3	4	4	3	4	4	1	1	1	1	1
Accuracy															
Subjective	0.73	0.76	0.7	0.71	0.77	0.77	0.73	0.52	0.63	0.6	1	0.91	0.5	0.57	0.86
PE	0.82	0.88	0.5	0.25	0	0.92	0.81	0.5	0.73	0.6					
A&P	0.67	0.78	0.18	1	0.33	0.81	0.82	0.84	0.75	0.85	0.22	0.4	0.29	0.22	0.25
PPV															
Subjective	0.32	0.68	0.62	0.48	0.2	0.45	0.65	0.31	0.29	0.24	0.67	0.83	0.82	0.8	0.75
PE	0.88	0.94	0.25	0.06	0	0.86	0.93	0.29	0.57	0.43					
A&P	0.75	0.88	0.6	0.57	0.4	0.76	0.69	0.84	0.79	0.73	0.67	0.8	0.67	0.67	0.5

Figure 1. Distribution of data elements. Each of the 5 scribes completed 3 separate simulation exercises. The absolute number of data elements for each section of the note was tabulated for each individual scribe. Subjective (Panel A), Physical exam (Panel B), and Assessment and plan (Panel C).
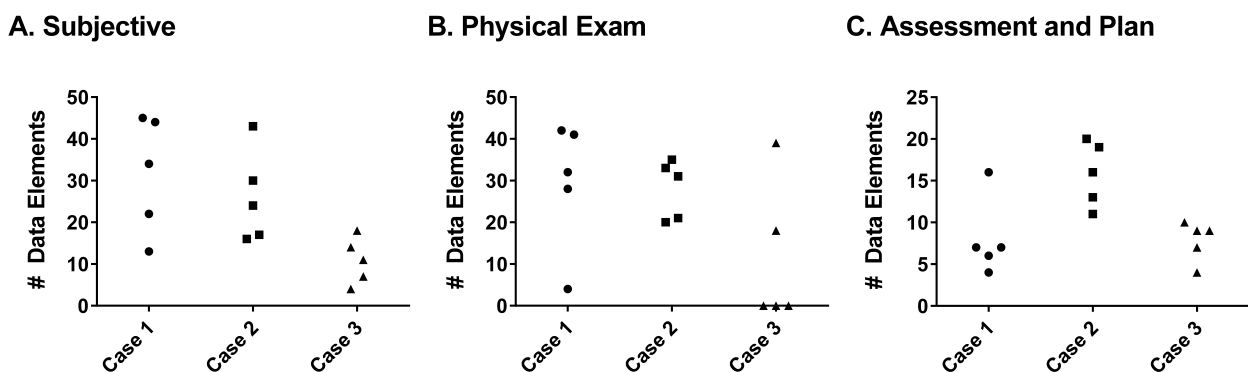

A. Subjective   B. Physical Exam   C. Assessment and Plan

Figure 2. Interscribe commonality in data elements. Each of the 5 scribes completed 3 separate simulation exercises. For each section of the note, Subjective (Panel A), Physical exam (Panel B), and Assessment and plan (Panel C), the fraction of data elements for each scribe in common among the other scribes for all three cases is presented.
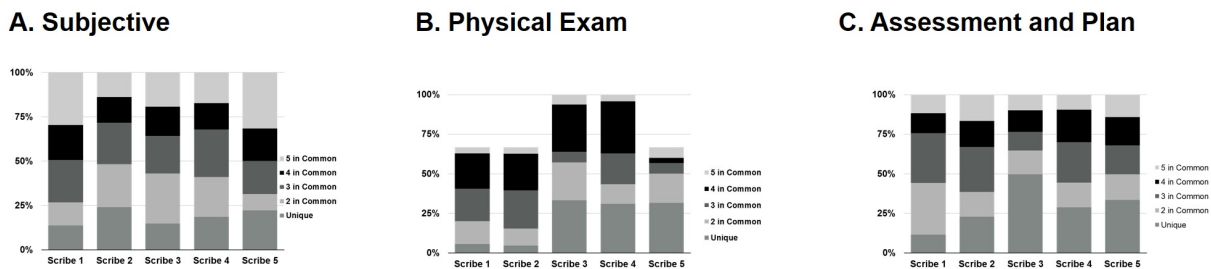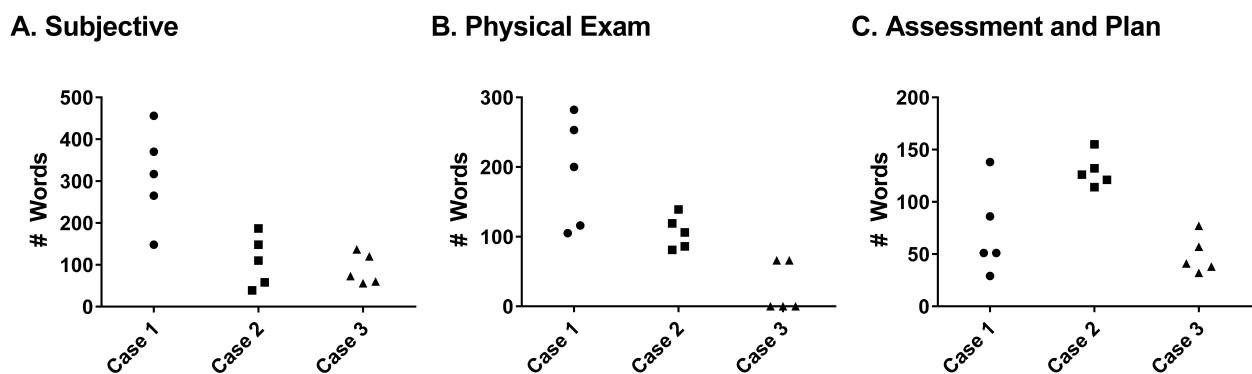

Figure 3. Distribution of Word Count. Five scribes each completed 3 separate simulation exercises. The absolute number of words for each section of the note was tabulated for each individual scribe. Subjective (Panel A), Physical Exam (Panel B), and Assessment and Plan (Panel C).

Discussion

Principal Findings

In this study, we created a novel virtual simulation to specifically assess scribe use and function. The use of a standardized video encounter carries the distinct advantage of untethering the simulation from a traditional simulation center, thereby improving accessibility of the training activity to multiple clinical environments. This represents a more scalable alternative, given how scribes are already reported to work in a variety of clinical environments and are deeply embedded in community clinics, many of which may not have access to traditional simulation. In addition, the use of a standardized video ensures consistency of the delivery of content, allowing for direct comparison of work-product between scribes and across practices.

With the standardization of the delivery of content and inclusion of the EHR as an integral part of the simulation activity, we were able to allow direct interscribe comparisons between notes, which revealed significant variability in note structure and length. There is a lack of clarity with respect to the extent of experience medical scribes require to attain any particular level of competency. Despite the fact that all of the scribes had at least 1 year of experience both in the specialty and with the EHR, there was almost a 3-fold difference in note length. Even more interesting was the difference in actual "note" elements between scribes. This is consistent with findings from studies showing discrepancies between physicians in the content and quality of documentation in notes [19,20]. Thus, whereas this phenomenon is most likely not unique to scribes, it does imply

that scribes may face the same issues that are found among other clinicians.

Although the simulation provides the basis to assess differences in note structure, we were also able to create a methodology to look at note content. We found evidence of errors of commission (incorrect data) and omission (missing data) by comparing the data elements found in notes written by scribes versus the notes written by an expert clinician. Notably, there was a paucity of overlap in content between the notes, with less than 40% of the documented plan items and diagnoses being common across the scribes. This is consistent with the observation that there is wide variability in the content of resident-physician-generated progress notes, where the primary author of the note (the resident) was also responsible for acquisition of the primary data and synthesizing that information into medical decision making [20]. This study suggests that similar issues may arise purely in the process of how our subjects communicate as members of an interprofessional team. However, this study does not delineate whether the differences observed are because of the individual scribe workflows, scribe deficits in medical knowledge, issues related to scribe training, or lacunae in scribe-physician communication. The use of a controlled simulated case may also explain the differences between our results and a recent study looking at actual scribe-generated notes in a practice setting [21]. In that study, scribe-generated notes for diabetes encounters, with medical assistants serving as scribes, created equally "readable" notes compared with physician-created notes. However, since each individual note corresponded to a unique patient encounter, there was no true "gold standard" for the information transmitted during that visit.

This highlights the power of using simulation as an objective tool for determining competency, by controlling for the actual clinical content verbalized. Given the variability among scribe training and experience, their ability is likely also variable. Through the use of high-fidelity simulation exercises, one can standardize their training to ensure that all scribes reach objective benchmarks required for clinical practice.

Limitations

It is important to note some important limitations to this study. Whereas this study focused on note creation, which is the primary role of the scribe, it did not address other scribe-specific activities such as data entry and data gathering [22,23]. Although we have previously demonstrated feasibility in integrating this into EHR-focused simulations, examining these other tasks will need to be the focus of future studies. Second, this study was a proof-of-concept study with a small number of scribes in a single specialty. Whereas the differences in note content and structure were noteworthy, a much larger cohort will be required to fully define the magnitude and scope of any potential safety

issues in documentation and EHR usage. This is even more important, given the wide spectrum in baseline scribe training and prior experience in medical care before functioning as a scribe. Third, even though the simulations were designed to be easily deployed across multiple environments, additional studies will be required to determine the quantity and content of training required for novice educators (eg, providers) to access, deploy, and assess the work output from these activities, especially in community and rural settings. Finally, in real-world workflow, scribe notes must be attested and signed by an attending physician. Thus, it is unclear how much of the variance observed in the note structure would persist in actual clinical care after this final, attending physician–level vetting.

Conclusions

In conclusion, our study highlights the variability of scribe documentation and the need for a more standardized approach to training. This proof-of-concept study demonstrated a means of effectively evaluating scribe performance.

Conflicts of Interest

None declared.

References

1. Mennemeyer ST, Menachemi N, Rahurkar S, Ford EW. Impact of the HITECH Act on physicians' adoption of electronic health records. J Am Med Inform Assoc 2016 Mar;23(2):375-379. [doi: 10.1093/jamia/ocv103] [Medline: 26228764]

2. Martineau M, Brookstone A, Streingham T, Hodgkins M. Physicians use of EHR systems 2014. Naples, FL: American EHR Partners; 2014.

3. Beasley JW, Wetterneck TB, Temte J, Lapin JA, Smith P, Rivera-Rodriguez AJ, et al. Information chaos in primary care: implications for physician performance and patient safety. J Am Board Fam Med 2011;24(6):745-751 [FREE Full text] [doi: 10.3122/jabfm.2011.06.100255] [Medline: 22086819]

4. Stephenson LS, Gorsuch A, Hersh WR, Mohan V, Gold JA. Participation in EHR based simulation improves recognition of patient safety issues. BMC Med Educ 2014 Oct 21;14:224 [FREE Full text] [doi: 10.1186/1472-6920-14-224] [Medline: 25336294]

5. Sakata KK, Stephenson LS, Mulanax A, Bierman J, Mcgrath K, Scholl G, et al. Professional and interprofessional differences in electronic health records use and recognition of safety issues in critically ill patients. J Interprof Care 2016 Sep;30(5):636-642. [doi: 10.1080/13561820.2016.1193479] [Medline: 27341177]

6. Zwaan L, Thijs A, Wagner C, Timmermans DR. Does inappropriate selectivity in information use relate to diagnostic errors and patient harm? the diagnosis of patients with dyspnea. Soc Sci Med 2013 Aug;91:32-38. [doi: 10.1016/j.socscimed.2013.05.001] [Medline: 23849236]

7. Gold JA, Tutsch AS, Gorsuch A, Mohan V. Integrating the electronic health record into high-fidelity interprofessional intensive care unit simulations. J Interprof Care 2015;29(6):562-563. [doi: 10.3109/13561820.2015.1063482] [Medline: 26652628]

8. Ancker JS, Kern LM, Edwards A, Nosal S, Stein DM, Hauser D, et al. How is the electronic health record being used? use of EHR data to assess physician-level variability in technology use. J Am Med Inform Assoc 2014;21(6):1001-1008 [FREE Full text] [doi: 10.1136/amiajnl-2013-002627] [Medline: 24914013]

9. Allred RJ, Ewer S. Improved emergency department patient flow: five years of experience with a scribe system. Ann Emerg Med 1983 Mar;12(3):162-163. [Medline: 6829994]

10. Gellert GA, Ramirez R, Webster SL. The rise of the medical scribe industry: implications for the advancement of electronic health records. JAMA 2015 Apr 07;313(13):1315-1316. [doi: 10.1001/jama.2014.17128] [Medline: 25504341]

11.   Underwood WS, Brookstone AJ, Barr MS. American EHR. 2011. The correlation of training duration with EHR usability and satisfaction: implications for meaningful us URL: http://www.americanehr.com/research/reports/correlation-training-ehr-meaningful-use.aspx [accessed 2017-08-31] [WebCite Cache ID 6t7cyBj5y]

12.   Yan H, Gardner R, Baier R. Beyond the focus group: understanding physicians' barriers to electronic medical records. Jt Comm J Qual Patient Saf 2012 Apr;38(4):184-191. [Medline: 22533131]

13.   Campbell LL, Case D, Crocker JE, Foster M, Johnson M, Lee CA, et al. Using medical scribes in a physician practice. J AHIMA 2012;83(11):64-69. [Medline: 23210302]

14.   -. Clarification: safe use of scribes in clinical settings. Joint Commission Perspectives 2011;31(6):4-5.

15.   Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. Qual Saf Health Care 2004 Oct;13 Suppl 1:i85-i90 [FREE Full text] [doi: 10.1136/qhc.13.suppl_1.i85] [Medline: 15465961]

16.   Mohan V, Scholl G, Gold JA. Intelligent simulation model to facilitate EHR training. AMIA Annu Symp Proc 2015;2015:925-932 [FREE Full text] [Medline: 26958229]

17.   Doberne JW, He Z, Mohan V, Gold JA, Marquard J, Chiang MF. Using high-fidelity simulation and eye tracking to characterize EHR workflow patterns among hospital physicians. AMIA Annu Symp Proc 2015;2015:1881-1889 [FREE Full text] [Medline: 26958287]

18.   March CA, Steiger D, Scholl G, Mohan V, Hersh WR, Gold JA. Use of simulation to assess electronic health record safety in the intensive care unit: a pilot study. BMJ Open 2013;3(4):- [FREE Full text] [doi: 10.1136/bmjopen-2013-002549] [Medline: 23578685]

19.   March CA, Scholl G, Dversdal RK, Richards M, Wilson LM, Mohan V, et al. Use of electronic health record simulation to understand the accuracy of intern progress notes. J Grad Med Educ 2016 May;8(2):237-240. [doi: 10.4300/JGME-D-15-00201.1] [Medline: 27168894]

20.   Rizvi RF, Harder KA, Hultman GM, Adam TJ, Kim M, Pakhomov SV, et al. A comparative observational study of inpatient clinical note-entry and reading/retrieval styles adopted by physicians. Int J Med Inform 2016 Jun;90:1-11. [doi: 10.1016/j.ijmedinf.2016.02.011] [Medline: 27103191]

21.   Misra-Hebert AD, Amah L, Rabovsky A, Morrison S, Cantave M, Hu B, et al. Medical scribes: how do their notes stack up? J Fam Pract 2016 Mar;65(3):155-159. [Medline: 27158687]

22.   Scheck A. The era of the scribe: lightening the EP's load. Emerg Med News 2004;26(2):1-6.

23.   Guglielmo WJ. What a scribe can do for you. Med Econ 2006 Jan 06;83(1):42, 44-42, 46. [Medline: 16509367]

Abbreviations

CWH: Center for Women's Health
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
HITECH: Health Technology for Economic and Clinical Health
NCATS: National Center for Advancing Translational Sciences
NIH: National Institutes of Health
Ob-Gyn: Obstetrics and Gynecology
OCTRI: Oregon Clinical and Translational Research Institute
OHSU: Oregon Health & Science University
PPV: positive predictive value

XSL•FO
RenderX

Original Paper

Impact of Electronic Health Records on Long-Term Care Facilities: Systematic Review

Clemens Scott Kruse[1], MBA, MHA, MSIT, PhD; Michael Mileski[1], DC, MPH, MSHEd, LNFA; Alekhya Ganta Vijaykumar[1], BDS; Sneha Vishnampet Viswanathan[1], BDS; Ujwala Suskandla[1], BDS; Yazhini Chidambaram[1], BDS

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

Corresponding Author:
Clemens Scott Kruse, MBA, MHA, MSIT, PhD
College of Health Professions
School of Health Administration
Texas State University
601 University Dr.
CHP Building, Room 250
San Marcos, TX, 78666
United States
Phone: 1 5122454462
Fax: 1 5122458712
Email: scottkruse@txstate.edu

Abstract

Background: Long-term care (LTC) facilities are an important part of the health care industry, providing care to the fastest-growing group of the population. However, the adoption of electronic health records (EHRs) in LTC facilities lags behind other areas of the health care industry. One of the reasons for the lack of widespread adoption in the United States is that LTC facilities are not eligible for incentives under the Meaningful Use program. Implementation of an EHR system in an LTC facility can potentially enhance the quality of care, provided it is appropriately implemented, used, and maintained. Unfortunately, the lag in adoption of the EHR in LTC creates a paucity of literature on the benefits of EHR implementation in LTC facilities.

Objective: The objective of this systematic review was to identify the potential benefits of implementing an EHR system in LTC facilities. The study also aims to identify the common conditions and EHR features that received favorable remarks from providers and the discrepancies that needed improvement to build up momentum across LTC settings in adopting this technology.

Methods: The authors conducted a systematic search of PubMed, Cumulative Index of Nursing and Allied Health (CINAHL), and MEDLINE databases. Papers were analyzed by multiple referees to filter out studies not germane to our research objective. A final sample of 28 papers was selected to be included in the systematic review.

Results: Results of this systematic review conclude that EHRs show significant improvement in the management of documentation in LTC facilities and enhanced quality outcomes. Approximately 43% (12/28) of the papers reported a mixed impact of EHRs on the management of documentation, and 33% (9/28) of papers reported positive quality outcomes using EHRs. Surprisingly, very few papers demonstrated an impact on patient satisfaction, physician satisfaction, the length of stay, and productivity using EHRs.

Conclusions: Overall, implementation of EHRs has been found to be effective in the few LTC facilities that have implemented them. Implementation of EHRs in LTC facilities caused improved management of clinical documentation that enabled better decision making.

(JMIR Med Inform 2017;5(3):e35) doi:10.2196/medinform.7958

KEYWORDS

electronic health record; long-term care; benefits; patient outcome; quality; nursing home; electronic medical record

XSL•FO

RenderX

Introduction

Background

While birth rates are falling, life expectancy is rising in many countries, and people are entering an age when they will most likely need care [1]. Seventy percent of older people live in low- or middle-income countries [1]. As age increases, so does the prevalence of chronic illness [2]. The new trend in societies today is smaller families and different residential patterns leading to a rising need for paid care [3]. Health care systems need to find innovative and sustainable ways to cope with these changing demographics accompanied by changes in familial social patterns. In most countries, a significant percentage of people in the older age group needing long-term care (LTC) services rely on services provided by unpaid caregivers [3]. Organization for Economic Cooperation and Development (OECD) estimates show that 80% of all older citizens in Austria and 82.2% in Spain are dependent on informal home care [3]. Approximately, 62.8% American men and women over the age of 65 years will need LTC by 2050, and so will the 39.8% Western Europeans in their respective countries [4,5]. This reflects an international issue. A study conducted by the US Department of Health and Human Services (HHS) showed that 4 out of every 10 people aged 65 years will be enrolled in a nursing home at some point in their lives, and roughly 10% of these will stay for 5 years or more [6,7].

Governments around the world have responded to the rising need for LTC at various echelons of care for a range of acute and emergent illness or disease. Western European countries are underfunding their LTC needs, relying on existing national systems to manage acute and emergent services, but their health systems are not prepared to care for the countries' dependent population for long periods [8]. For instance, most countries in Western Europe have a mechanism in place to fund formal care (50%-75% provided in the community), whereas Northern and continental European countries have arrangements to partially fund informal care [5]. Germany mandates LTC insurance, a program called *Pflegeversicherung*, to fund for LTC with equal contribution between the insured and their employers [8]. With the rising need for LTC and changing consumer expectations, some LTC facilities have been seen to adopt electronic health records (EHRs), despite the lack of funding opportunities, but overall the level of adoption of EHRs in the United States and Europe is low [8,9]. The EHR can improve quality of care in LTC facilities through a reduction in medication-related errors, improved clinical documentation and decision making, and through the Health Information Exchange. The latter point, which involves the Health Information Exchange, is particularly applicable to LTC because of the number of transfers and medical handoffs that accompany care of the elderly. These benefits are realized by the patient, the provider, and the organization. The EHR can also be associated with key qualities of both efficiency and effectiveness through improved data analysis and audits, coding and links to billing, going green or storage expenses, and record retention and proper safeguarding [8].

The Centers for Medicare and Medicaid Services (CMS) defines an EHR as "an electronic version of a patient's medical history, that is maintained by the provider over time, and may include all the key administrative clinical data relevant to that person's care under a specific provider, including demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports" [10]. Another organization within the CMS added the following to the definition: "allow access to evidence-based tools that providers can use to make decisions about a patient's care, as well as automate and streamline provider workflow" [11]. With the help of the EHR, providers can access care-related activities directly or indirectly through various interfaces such as evidence-based decision support, quality management, and outcomes reporting.

In the literature, the terms EHR and electronic medical record (EMR) are often used interchangeably. The CMS differentiates these two: the EMR is bound to one organization, and the EHR is compatible across organizational lines. Although we would prefer that all publications kept these distinct, we also realize that it is impractical. Some of the literature analyzed in this review refers to EMRs when they are really analyzing an EHR. In the interest of keeping the authors' words intact, we will not differentiate between them in our analysis. We are evaluating works about EHRs, EMRs, and some stand-alone components of the EHR or EMR for the same purpose of this review.

LTC is *a continuum of medical and social services designed to support the needs of people living with chronic health problems that affect their ability to perform everyday activities* [12]. LTC is an umbrella term that spans a large range of services. LTC services include traditional acute-care medical services, social services, skilled nursing facilities, rehabilitation facilities, assisted living, and other housing-based services. The goals of LTC are much more complicated and considerably more difficult to measure than the goals of acute medical care for the nonelderly.

The EHR enables providers to deliver better medical care to patients because of the availability of complete and accurate information [13]. Previous empirical studies conducted in other health settings consistently support that EHR can assist health care providers to minimize errors, improve safety and quality, and decrease costs [14]. The results from these empirical studies have influenced hospitals and other health care settings to implement and adopt EHRs actively; whereas LTC facilities, especially licensed nursing homes, have been slower in adopting and implementing EHRs [15]. This slower adoption pace is because of the lack of significant literature supporting the view that EHR implementation improves quality and decreases cost in the long run [16]. With the growth of aging population and LTC facilities providing care to this fast-growing segment, it seems important for these facilities to implement and use the EHR system meaningfully. Presence of the EHR in LTC could help meet the diverse needs of the dependent population and enable enhanced quality of care and coordination.

Objective

The purpose of this review was to address the knowledge gap and the lack of significant literature accounting for the

XSL•FO

RenderX

relationship between an EHR and LTC facilities. Do existing EHR implementations in current LTC facilities have positive outcomes? Do the users of these systems have positive experiences or observations that have been shared? The hypotheses are as follows:

H$_1$: There are positive experiences by users of existing EHR implementations in LTC facilities.

H$_0$: There are no positive experiences of users of existing EHR implementations in LTC facilities.

Methods

Eligibility Criteria

Our methods followed a measurement tool for the assessment of multiple systematic reviews (AMSTAR) [17]. The format of the review follows the preferred reporting items for systematic reviews and meta-analyses (PRISMA) [18]. Papers were eligible for selection in this systematic review if they were published in the last 10 years in academic (peer-reviewed) journals, in English, and whose full-text was available. We chose 10 years because we thought that 10 years would be a sufficient amount of time to collect information on technology. We limited the search to peer-reviewed journals to ensure some element of quality to the papers we were analyzing. We made the decision not to include other systematic reviews.

Information Sources

We queried three common research databases: MEDLINE (the Web-based component to the MEDical Literature Analysis and Retrieval System) by Web of Science, Cumulative Index of Nursing and Allied Health Literature (CINAHL), and PubMed. We used key terms from the US National Library of Medicine's Medical Subject Headings (MeSH) separated by Boolean terms. Searches were conducted from April 21 to April 24, 2017. The reason we chose to query MEDLINE by Web of Science is because we received different outputs when we queried MEDLINE in PubMed. We do not have a reason for the disparity. MEDLINE by Web of Science gave us more papers to choose from.

Search and Study Selection

Searches in each database were nearly identical: (EHR OR EMR OR "electronic health record" OR "electronic medical record") AND ("long term care" OR "long-term care" OR "nursing home") AND (outcome OR impact OR effect) NOT "patient portal" NOT "health information exchange." Due to the differences in indexing methods between the databases, we had to slightly modify the search string and filters for each. An exact listing of the search strings and filters is provided in Multimedia Appendix 1. We screened for date of publication to begin in 2007 and go through the end of April 2017. In PubMed and Web of Science, we were also able to screen out reviews. In both CINAHL and PubMed, we excluded MEDLINE because it was being collected separately from Web of Science. These 28 papers were placed into an Excel (Microsoft Corp) spreadsheet shared among the reviewers. The duplicates were removed.

Data Collection Process and Data Items

Reviewers agreed ahead of time what to look for in each abstract. We wanted to focus on papers that explained an experience within an LTC facility of an EHR or a major component of the EHR as defined by the CMS [8,11]. We also searched for papers that expressed the experience, positively or negatively, in terms of effectiveness, for example, outcomes and quality, and/or or efficiency or advantages in money saved or workflow [19]. The initial search resulted in 100 results. After removing duplicates and filtering, the remaining 28 abstracts were divided among the reviewers in a way that all were reviewed at least twice (overlapping sets), as outlined by AMSTAR [17]. Reviewers carefully read each abstract ensuring that our review objectives were being addressed. Independent notes were taken on a shared spreadsheet. Additionally, each reviewer examined the references of each paper to identify any salient papers that our search may have missed, which identified an additional nine papers to the review queue. Before a consensus meeting, the Excel spreadsheets of each reviewer were combined to show agreement or disagreement about whether or not the paper was germane to our objective. An initial kappa statistic was calculated at kappa=.79. Where there was disagreement, reviewers discussed what they observed and reached a consensus. One reviewer on the team served as the facilitator and made the final ruling after hearing the input. Through this process, an additional seven papers were removed from consideration because of lack of applicability to our topic. At the end of the consensus meeting, the final selection of papers was chosen for analysis (N=28).

Papers were assigned to reviewers in a way that each paper was read by at least two reviewers. Once again, reviewers recorded independent observations on their copy of the Excel spreadsheet, ensuring to capture the observations in terms of effectiveness, efficiency, or the negatives of the same. Reviewers were also asked to identify possible bias in each paper, loosely following the Cochrane Collaboration's risk-of-bias tool [20]. All observations were combined into one spreadsheet for discussion in the second consensus meeting. Reviewers were also asked to record any overarching themes that seemed to serve as a common thread between papers, as well as any significant levels of bias that could have been present in each study. This practice is in accordance with thematic analysis [21]. During the second consensus meeting, reviewers discussed their notes and observations (results, possible themes, and potential bias). No papers were discarded because of bias.

Summary Measures and Synthesis of Results

The summary measure used in this analysis was the expression of experience with EHRs or a major stand-alone component of an EHR, in an LTC facility, expressed in terms of effectiveness or efficiency, and a frequency of occurrence of the themes identified by the reviewers. A table of observations was created, and an affinity matrix was created to illustrate potential trends. Figure 1 illustrates the selection process with the inclusion and exclusion criteria. This figure strictly follows the PRISMA standard.

Figure 1. Paper selection process with inclusion and exclusion criteria.
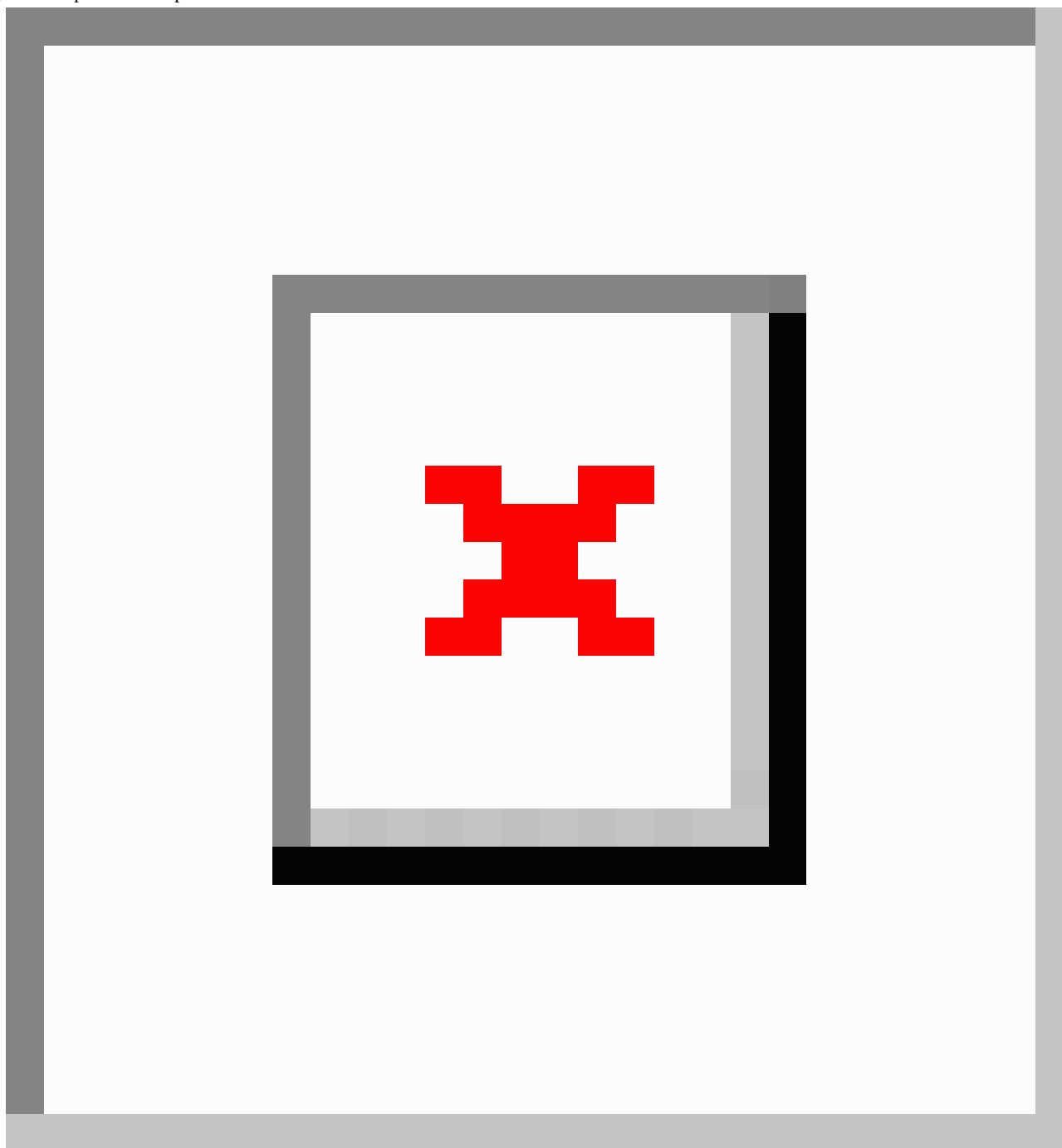
Results

Study Selection and Characteristics

As illustrated in Figure 1, 100 papers entered the screening process, 13 duplicates were removed, 60 were screened out using our selection and exclusion criteria, 7 papers were removed because after reading their abstracts they did not seem to be germane to our objectives, and 9 additional papers were added from the references of those remaining. The final sample for analysis included 28 papers. Observations from each paper were summarized into our spreadsheet and from that spreadsheet we created a summary of the studies (Multimedia Appendix 2). Reviewers recorded observations of positive and negative experiences with the EHR in LTC, as well as any miscellaneous observations relevant for discussion.

Additional Analysis

After the second consensus meeting, the overarching themes recorded by each reviewer were combined. We counted the number of times that a theme occurred in the literature and sorted by frequency of occurrence. These data were placed into an affinity matrix for further analysis (see Table 1). The total number of themes or attributes was 11, and the total number of occurrences was 44 [22-49].

The broad research criteria encouraged a thorough assessment of the implication of an EHR across various attributes: health outcomes, documentation, quality outcomes, length of stay,

productivity, accessibility, medication safety, cost, patient satisfaction, nurses' working time, physician satisfaction, and time consumption.

Among all the attributes, markedly 12 out of 28 papers addressed management of documentations, with approximately 27% of total occurrences of the attributes in the literature [24,27-33, 37,41,43,45,48]. Included in this attribute was data handling

and the use of the EHR. Among these 12 papers, eight papers [24,29,32,34,37,41,43,48] showed positive impact, that is, improved management of documentation using EHRs, and five papers [27-30,45] portrayed negative impact of an EHR on documentation management, that is, it either increased documentation time and burden or showed no results post implementation.

Table 1. The frequency of occurrence of attributes to assess the impact of electronic health records (EHRs) in long-term care (LTC) facilities.

Attributes	Occurrences	Frequency (%)
Management of documentation	24,27-33,37,41,43,45,48	12 (27)
Quality outcomes	22,24,26,31,35,36,39,44,47	9 (20)
Health outcomes	24,28,31,41	4 (9)
Time consumption	26,27,36,49	4 (9)
Access to patient data	26,28,36,41	4 (9)
Physician satisfaction	31,37,38	3 (7)
Medication safety	42,44	2 (4)
Cost	23,41	2 (5)
Patient satisfaction	36,48	2 (5)
Productivity	46	1 (2)

Nine out of 28 papers (32%) reported positive quality outcomes, accounting for 21% of the occurrences [22,24,25,31,35,36, 39,44,47]. Four of 28 papers (14%) showed improved health outcomes using EHRs in aged care settings and nursing homes, accounting for 9% of occurrences [24,28,34,41]. Four out of 28 papers reported impact of EHRs on time consumption, accounting for 9% of occurrences [25,27,36,48], and surprisingly three out of these four papers showed negative impact, that is, time spent on all activities either remained unchanged post implementation or increased [25,28,36], and one paper reported reduced time consumption in creating electronic medical charts [48]. Four out of 28 papers (14%) demonstrated improved access to clinical information and patient data using EHRs, accounting for 9% of occurrences [25,28,36,41].

Notably, only three out of 28 papers (11%) reported greater physician satisfaction using EHRs, as it improved working environment and reduced errors [34,37,38]. Also, three attributes were mentioned only twice out of 28 papers: patient satisfaction [36,48], medication safety [42,44], and cost [23,41], each of which represent 5% of total occurrences of attributes in the literature. Furthermore, one attribute, which increased productivity of the settings, was reported only once out of 28 papers [46] after implementing EHRs, which represent only 2% of total occurrences.

Management of documentation was identified as a common theme in 13 papers. Studies documented that the time consumed for management of documents in EHR compared with paper-based records was significantly less [36,41,48]. Few papers also recorded that the management of documentation was more comprehensive, better in quality, and reduced human errors such as repetition and neglecting to medicate a resident [24,29,32,34,37,43,49]. One paper also emphasized the ease of

documentation while using EHRs as compared with traditional paper-based documents [41].

Few papers mentioned that they could not observe much difference in the time consumed for documentation after implementing EHRs [28,33,45], and one among them mentioned that there was minimal difference initially which later increased the time taken 6 months after implementation but time taken increased 6 months after implementation [34]. One paper acknowledged the accuracy and comprehensibility of EHRs but also stated that these benefits were recorded in the first 6 months after implementing EHRs and were not sustained [37]. Reasons attributing to these unfavorable outcomes may, in part, be a result of the practice of documenting some information on paper and others on a computer. The lack of the staff's experience with computer systems and the unavailability of required resources largely contribute to such outcomes. A more complex and in-depth understanding of the staff's perception, documentation workflow, and information needs along with sufficient resources and training might help in overcoming these results [30,33].

Quality outcome was the second most commonly observed theme. Many papers stated that EHRs directly improved the quality of care [25,34,36,44,47]. Another paper reported that the use of EHR improved interprofessional integration, thereby improving the quality of care.

Health outcome was another commonly identified theme. Four different papers showed significant improvement in health outcomes by reducing the occurrence of infections, high-risk pressure sores, neurolepsis, improving activities of daily living (ADL), range of motion, and timely medication [41,34]. One study particularly emphasized that the likelihood of neglecting to medicate a resident decreased but also noted that there were unintended incidents of neglect to medicate because of energy

blackouts [24]. There was another study that mentioned that, when applied to delirium prevention strategies, EHRs failed to lower delirium rates among patients with hip fracture. Factors such as staff turnover, impact of organizational culture, personnel changes, and structure on the uptake of the delirium prevention strategies were the major factors that influenced the failure of this model. Furthermore, there were multiple challenges operating at different levels within the system [40].

The next most commonly occurring attribute was time consumption. In a few of the studies, the respondents have mentioned that EHR was time consuming because of reasons such as complexity in signing out of an EHR [25,36]. Another paper stated that there was no significant change in the proportion of time spent on activities and oral communication [27]. General physicians in a study had responded that the time taken to create electronic medication charts was much less compared with conventional charting [48]. Out of four papers that refer to time consumption, three state that there is no evidence of time saving as a result of using EHRs. This shocking observation calls for more research to address the time-consuming aspect of EHRs.

Access to patient data was another commonly occurring attribute [41,25,27,36]. Out of these papers, three mentioned that EHRs improve access to patient records by facilitating real-time availability and remote access [41,25,36]. One study stated that implementation of EHRs resulted in difficulty to access data [28].

Other common factors included cost, patient satisfaction, physician satisfaction, length of stay, and productivity. Studies mentioned that there was a marked increase in the cost incurred by facilities post implementation [41,35]. The authors recommend that further research should attempt to throw light on the factors contributing to increase in cost and evaluate ways by which the high upfront cost could be balanced with benefits in LTC facilities, as this would inspire more facilities to adopt EHRs.

Discussion

Summary of Evidence

EHRs are known to improve care coordination and health outcomes. Although LTC facilities have been slow to adopt EHR, they continue to be areas where the benefits of implementing EHR can be realized to its fullest potential. This review tries to identify the established outcomes in various LTC facilities that have adopted this technology. For this review, we analyzed papers, studies, and other summaries of experiences relating to our topic of interest. Management of documentation, quality outcome, and health outcomes were identified as the most common themes, which were identified in 60% of all papers reviewed.

There were both positive and negatives outcomes reported in this systematic review; however, the former was found in the literature more than the latter. Some reported a boost in productivity only after 23 months; others did not put a time frame on it—they just reported slower processes.

The LTC market has been slow to adopt health information technology, in general, and EHRs specifically. The paucity of data on the adoption of the EHR in LTC is similar to the private health care market in the United States before the major legislation in 2009. The adoption rates for EHRs in the United States greatly increased with incentives that helped to offset the steep adoption costs of the technology. Future research could determine the level at which the cost of investing in the EHR is equal or better than the cost of abstaining.

Limitations

The researchers reviewed only those papers that were published between the years 2007 to 2017 and did not include the papers outside the period of study. We thought 10 years' time was adequate, commensurate with other reviews. There is unavailability of data owing to the slow adoption of EHR in LTC settings. The systematic search process in the three primary databases yielded studies that predominantly focused on the United States' LTC scenario rather than an international focus. Although selection bias and face validity are concerns, we mitigated these risks by following the AMSTAR standard and using more than one reviewer to opine on the inclusion or exclusion of papers used for analysis [17].

Conclusions

Overall, implementation of EHRs has been found to be more effective than not in LTC facilities. Implementation of EHRs in LTC facilities caused improved management of clinical documentation that enabled better decision making. Negative experiences were observed in workflow and productivity, but it is unclear whether this was because of change management and the general disruption that a major information technology (IT) implementation can exert on the organization. The authors recommend improving the design of EHRs that address issues such as time spent on documentation and enhancing the usability for physicians and nurses. These improvements would address most of the negative experiences and may promote widespread adoption of this essential technology in LTC.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE by Web of Science (Thompson Reuters).

[PDF File (Adobe PDF File), 25KB - medinform_v5i3e35_app1.pdf]

Multimedia Appendix 2

Summary of analysis.

[PDF File (Adobe PDF File), 60KB - medinform_v5i3e35_app2.pdf]

References

1.   WHO. 2012. World Health Day - Ageing and Health - Toolkit for event organizers URL: http://apps.who.int/iris/bitstream/10665/70840/1/WHO_DCO_WHD_2012.1_eng.pdf [accessed 2017-09-01] [WebCite Cache ID 6t93QRzA6]

2.   Denton FT, Spencer BG. Chronic health conditions: changing prevalence in an aging population and some implications for the delivery of health care services. Can J Aging 2010 Mar;29(1):11-21. [doi: 10.1017/S0714980809990390] [Medline: 20202262]

3.   OECD. 2005. Long-term Care for Older People: OECD study (2001-2004) URL: http://www.oecd.org/els/health-systems/long-termcareforolderpeopleoecdstudy2001-2004.htm [accessed 2017-09-01] [WebCite Cache ID 6t93hpChp]

4.   The Scan Foundation. 2012 Jun. Growing Demand for Long-Term Care in the U.S. (Updated) URL: http://www.thescanfoundation.org/sites/default/files/us_growing_demand_for_ltc_june_2012_fs.pdf [accessed 2017-05-01] [WebCite Cache ID 6q8sjG5gS]

5.   Van Eenoo L, Declercq A, Onder G, Finne-Soveri H, Garms-Homolová V, Jónsson PV, et al. Substantial between-country differences in organising community care for older people in Europe-a review. Eur J Public Health 2016 Apr;26(2):213-219. [doi: 10.1093/eurpub/ckv152] [Medline: 26338722]

6.   Senior Answers - Colorado Gerontological Society. Long Term Care URL: https://www.senioranswers.org/insurance/long-term-care/ [accessed 2017-05-01] [WebCite Cache ID 6q8t7PvNv]

7.   Pavolini E, Ranci C. Restructuring the welfare state: reforms in long-term care in Western European countries. J Eur Soc Policy 2008;18(3):246-259 [FREE Full text]

8.   AHIMA. Electronic Health Record Adoption in Long Term Care (2014 update) URL: http://library.ahima.org/doc?oid=107519 [accessed 2017-06-22] [WebCite Cache ID 6rQ8ciVfk]

9.   Clinical Innovation. Research and Markets: Long-term Care Software (EHR, eMAR, Payroll Management) Market - Global Forecast to 2019 URL: http://www.clinical-innovation.com/press-releases/research-and-markets-long-term-care-software-ehr-emar-payroll-management-market-global-forecast-2019 [accessed 2017-05-02] [WebCite Cache ID 6q9F5fQtv]

10. CMS. Electronic Health Records URL: https://www.cms.gov/Medicare/E-Health/EHealthRecords/index.html?redirect=/EhealthRecords/ [accessed 2017-05-02] [WebCite Cache ID 6q9FTv3XP]

11. Health IT. What is an electronic health record (EHR)? URL: https://www.healthit.gov/providers-professionals/faqs/what-electronic-health-record-ehr [accessed 2017-06-26] [WebCite Cache ID 6rVqIUEFE]

12. McCall N. Long term care: Definition, demand, cost, and financing. In: McCall N, editor. Who Will Pay for Long Term Care?: Insights from the Partnership Programs. Chicago: Health Administration Press; 2001:3-31.

13. Health IT. Benefits of EHRs: Improved Diagnostics & Patient Outcomes URL: https://www.healthit.gov/providers-professionals/improved-diagnostics-patient-outcomes [accessed 2017-05-02] [WebCite Cache ID 6q9GRMp6W]

14. Ford EW, Menachemi N, Peterson LT, Huerta TR. Resistance is futile: but it is slowing the pace of EHR adoption nonetheless. J Am Med Inform Assoc 2009;16(3):274-281 [FREE Full text] [doi: 10.1197/jamia.M3042] [Medline: 19261931]

15. Broughton W, Lashlee H, Marcum C, Wilson GM. Health information technology: a new world of nursing homes. J Gerontol Geriatric Res 2013;2:122. [doi: 10.4172/2167-7182.1000122]

16. Ajami S, Bagheri-Tadi T. Barriers for adopting Electronic Health Records (EHRs) by physicians. Acta Inform Med 2013;21(2):129-134 [FREE Full text] [doi: 10.5455/aim.2013.21.129-134] [Medline: 24058254]

17. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007 Feb 15;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]

18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Reprint--preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Phys Ther 2009 Sep;89(9):873-880. [Medline: 19723669]

19. Health IT. Measure Effects and Refine CDS Interventions URL: https://www.healthit.gov/sites/default/files/3-4-5-measure-effects-and-refine-cds-interv.pdf [accessed 2017-06-26] [WebCite Cache ID 6rVsG8dwi]

20. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [Medline: 22008217]

21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101.

22. Jiang T, Yu P, Hailey D, Ma J, Yang J. The impact of electronic health records on risk management of information systems in Australian residential aged care homes. J Med Syst 2016 Sep;40(9):204. [doi: 10.1007/s10916-016-0553-y] [Medline: 27501930]

XSL•FO
RenderX

23. Walker D, Mora A, Demosthenidy MM, Menachemi N, Diana ML. Meaningful use Of EHRs among hospitals ineligible for incentives lags behind that of other hospitals, 2009-13. Health Aff (Millwood) 2016 Mar;35(3):495-501. [doi: 10.1377/hlthaff.2015.0924] [Medline: 26953305]

24. Qian S, Yu P, Hailey DM. The impact of electronic medication administration records in a residential aged care home. Int J Med Inform 2015 Nov;84(11):966-973. [doi: 10.1016/j.ijmedinf.2015.08.002] [Medline: 26358850]

25. Meehan R. Electronic health records in long-term care. J Appl Gerontol 2015 Oct 01:733464815608493. [doi: 10.1177/0733464815608493] [Medline: 26464335]

26. Aguilar KM, Campbell RS, Fiester A, Simpson RL, Hertel C. Bringing care home: how telemonitoring can expand population health management beyond the hospital. Nurs Adm Q 2014;38(2):166-172. [doi: 10.1097/NAQ.0000000000000029] [Medline: 24569764]

27. Munyisia E, Yu P, Hailey D. The effect of an electronic health record system on nursing staff time in a nursing home: a longitudinal cohort study. Australas Med J 2014 Jul 31;7(7):285-293 [FREE Full text] [doi: 10.4066/AMJ.2014.2072] [Medline: 25157268]

28. Yu P, Zhang Y, Gong Y, Zhang J. Unintended adverse consequences of introducing electronic health records in residential aged care homes. Int J Med Inform 2013 Sep;82(9):772-788. [doi: 10.1016/j.ijmedinf.2013.05.008] [Medline: 23770027]

29. Wang N, Yu P, Hailey D. Description and comparison of quality of electronic versus paper-based resident admission forms in Australian aged care facilities. Int J Med Inform 2013 May;82(5):313-324. [doi: 10.1016/j.ijmedinf.2012.11.011] [Medline: 23254294]

30. Munyisia EN, Yu P, Hailey D. Caregivers' time utilization before and after the introduction of an electronic nursing documentation system in a residential aged care facility. Methods Inf Med 2013;52(5):403-410. [doi: 10.3414/ME12-01-0024] [Medline: 23703463]

31. Krüger K, Strand L, Geitung JT, Eide GE, Grimsmo A. Can electronic tools help improve nursing home quality? ISRN Nurs 2011;2011:208142 [FREE Full text] [doi: 10.5402/2011/208142] [Medline: 22013540]

32. Yu P. Aged care IT in Australia-the past, present and future. EJHI 2012;7(2):e12.

33. Munyisia EN, Yu P, Hailey D. Does the introduction of an electronic nursing documentation system in a nursing home reduce time on documentation for the nursing staff? Int J Med Inform 2011 Nov;80(11):782-792. [doi: 10.1016/j.ijmedinf.2011.08.009] [Medline: 21956002]

34. Zhang Y, Yu P, Shen J. The benefits of introducing electronic health records in residential aged care facilities: a multiple case study. Int J Med Inform 2012 Oct;81(10):690-704. [doi: 10.1016/j.ijmedinf.2012.05.013] [Medline: 22749424]

35. Cherry BJ, Ford EW, Peterson LT. Experiences with electronic health records: early adopters in long-term care facilities. Health Care Manage Rev 2011;36(3):265-274. [doi: 10.1097/HMR.0b013e31820e110f] [Medline: 21646885]

36. Rantz MJ, Alexander G, Galambos C, Flesner MK, Vogelsmeier A, Hicks L, et al. The use of bedside electronic medical record to improve quality of care in nursing facilities: a qualitative analysis. Comput Inform Nurs 2011 Mar;29(3):149-156. [doi: 10.1097/NCN.0b013e3181f9db79] [Medline: 20975545]

37. Munyisia EN, Yu P, Hailey D. The changes in caregivers' perceptions about the quality of information and benefits of nursing documentation associated with the introduction of an electronic documentation system in a nursing home. Int J Med Inform 2011 Feb;80(2):116-126. [doi: 10.1016/j.ijmedinf.2010.10.011] [Medline: 21242104]

38. Sockolow PS, Weiner JP, Bowles KH, Abbott P, Lehmann HP. Advice for decision makers based on an electronic health record evaluation at a program for all-inclusive care for elders site. Appl Clin Inform 2011 Jan 26;2(1):18-38 [FREE Full text] [doi: 10.4338/ACI-2010-09-RA-0055] [Medline: 23616858]

39. Okamoto E, Miyamoto M, Hara K, Yoshida J, Muto M, Hirai A, et al. Integrated care through disease-oriented clinical care pathways: experience from Japan's regional health planning initiatives. Int J Integr Care 2011 Jan;11(Spec 10th Anniversary Ed):e125 [FREE Full text] [Medline: 22128281]

40. Holroyd-Leduc JM, Abelseth GA, Khandwala F, Silvius JL, Hogan DB, Schmaltz HN, et al. A pragmatic study exploring the prevention of delirium among hospitalized older hip fracture patients: Applying evidence to routine clinical practice using clinical decision support. Implement Sci 2010 Oct 22;5:81 [FREE Full text] [doi: 10.1186/1748-5908-5-81] [Medline: 20969770]

41. Rantz MJ, Hicks L, Petroski GF, Madsen RW, Alexander G, Galambos C, et al. Cost, staffing and quality impact of bedside electronic medical record (EMR) in nursing homes. J Am Med Dir Assoc 2010 Sep;11(7):485-493. [doi: 10.1016/j.jamda.2009.11.010] [Medline: 20816336]

42. Brankline AL, Coyle CM, Jencks KA, Mullegama A, O'Brien MW. Practical innovations: technology-assisted referrals. Soc Work Health Care 2009;48(8):768-776. [doi: 10.1080/00981380902958213] [Medline: 20182988]

43. Milne CT, Trigilia D, Houle TL, Delong S, Rosenblum D. Reducing pressure ulcer prevalence rates in the long-term acute care setting. Ostomy Wound Manage 2009 Apr;55(4):50-59 [FREE Full text] [Medline: 19387096]

44. Scott-Cawiezell J, Madsen RW, Pepper GA, Vogelsmeier A, Petroski G, Zellmer D. Medication safety teams' guided implementation of electronic medication administration records in five nursing homes. Jt Comm J Qual Patient Saf 2009 Jan;35(1):29-35. [Medline: 19213298]

45. Hakes B, Whittington J. Assessing the impact of an electronic medical record on nurse documentation time. Comput Inform Nurs 2008;26(4):234-241. [doi: 10.1097/01.NCN.0000304801.00628.ab] [Medline: 18600132]

46.   Yu P, Hailey D, Li H. Caregivers' acceptance of electronic documentation in nursing homes. J Telemed Telecare 2008;14(5):261-265. [doi: 10.1258/jtt.2008.080310] [Medline: 18633002]

47.   Cherry B, Carter M, Owen D, Lockhart C. Factors affecting electronic health record adoption in long-term care facilities. J Healthc Qual 2008;30(2):37-47. [Medline: 18411891]

48.   Burns P, Perkins DA, Larsen K, Dalley A. The introduction of electronic medication charts and prescribing in aged care facilities: an evaluation. Australas J Ageing 2007 Sep;26(3):131-134. [doi: 10.1111/j.1741-6612.2007.00237.x]

49.   Lindner SA, Davoren JB, Vollmer A, Williams B, Landefeld CS. An electronic medical record intervention increased nursing home advance directive orders and documentation. J Am Geriatr Soc 2007 Jul;55(7):1001-1006. [doi: 10.1111/j.1532-5415.2007.01214.x] [Medline: 17608871]

Abbreviations

ADL: activities of daily living
AMSTAR: assessment of multiple systematic reviews
CINAHL: Cumulative Index of Nursing and Allied Health Literature
CMS: Centers for Medicare and Medicaid Services
HHS: US Department of Health and Human Services
EHR: electronic health records
EMR: electronic medical records
HIT: health information technology
LTC: long-term care
MeSH: medical subject headings
OECD: Organization for Economic Cooperation and Development
PRISMA: preferred reporting items for systematic reviews and meta-analyses
RAC: residential aged care
RACF: residential aged care facility

Original Paper

An Ontology to Improve Transparency in Case Definition and Increase Case Finding of Infectious Intestinal Disease: Database Study in English General Practice

Simon de Lusignan[1,2*], BSc, MBBS, MSc, MD(Res), DRCOG, FHEA, FBCS CITP, FRCGP; Stacy Shinneman[1*], BSc, MSc; Ivelina Yonova[1,2*], MA; Jeremy van Vlymen[1*], MSc; Alex J Elliot[3*], PhD; Frederick Bolton[4*], PhD; Gillian E Smith[3*], PhD; Sarah O'Brien[5*], MBBS, FRCP, FFPHM, DTM&H

[1]Section of Clinical Medicine and Ageing, Department of Clinical and Experimental Medicine, University of Surrey, Guildford, United Kingdom

[2]Royal College of General Practitioners, Research and Surveillance Centre, London, United Kingdom

[3]Real-time Syndromic Surveillance Team, National Infection Service, Public Health England, Birmingham, United Kingdom

[4]Epidemiology and Population Health, University of Liverpool, Liverpool, United Kingdom

[5]Institute of Psychology Health and Society, University of Liverpool, Liverpool, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Simon de Lusignan, BSc, MBBS, MSc, MD(Res), DRCOG, FHEA, FBCS CITP, FRCGP
Section of Clinical Medicine and Ageing
Department of Clinical and Experimental Medicine
University of Surrey
Leggett Building, 2nd Floor
Guildford, GU2 7XH
United Kingdom
Phone: 44 (0)1483 683089
Email: s.lusignan@surrey.ac.uk

Abstract

Background: Infectious intestinal disease (IID) has considerable health impact; there are 2 billion cases worldwide resulting in 1 million deaths and 78.7 million disability-adjusted life years lost. Reported IID incidence rates vary and this is partly because terms such as "diarrheal disease" and "acute infectious gastroenteritis" are used interchangeably. Ontologies provide a method of transparently comparing case definitions and disease incidence rates.

Objective: This study sought to show how differences in case definition in part account for variation in incidence estimates for IID and how an ontological approach provides greater transparency to IID case finding.

Methods: We compared three IID case definitions: (1) Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) definition based on mapping to the Ninth International Classification of Disease (ICD-9), (2) newer ICD-10 definition, and (3) ontological case definition. We calculated incidence rates and examined the contribution of four supporting concepts related to IID: symptoms, investigations, process of care (eg, notification to public health authorities), and therapies. We created a formal ontology using ontology Web language.

Results: The ontological approach identified 5712 more cases of IID than the ICD-10 definition and 4482 more than the RCGP RSC definition from an initial cohort of 1,120,490. Weekly incidence using the ontological definition was 17.93/100,000 (95% CI 15.63-20.41), whereas for the ICD-10 definition the rate was 8.13/100,000 (95% CI 6.70-9.87), and for the RSC definition the rate was 10.24/100,000 (95% CI 8.55-12.12). Codes from the four supporting concepts were generally consistent across our three IID case definitions: 37.38% (3905/10,448) (95% CI 36.16-38.5) for the ontological definition, 38.33% (2287/5966) (95% CI 36.79-39.93) for the RSC definition, and 40.82% (1933/4736) (95% CI 39.03-42.66) for the ICD-10 definition. The proportion of laboratory results associated with a positive test result was 19.68% (546/2775).

Conclusions: The standard RCGP RSC definition of IID, and its mapping to ICD-10, underestimates disease incidence. The ontological approach identified a larger proportion of new IID cases; the ontology divides contributory elements and enables transparency and comparison of rates. Results illustrate how improved diagnostic coding of IID combined with an ontological approach to case definition would provide a clearer picture of IID in the community, better inform GPs and public health services

XSL•FO
RenderX

about circulating disease, and empower them to respond. We need to improve the Pathology Bounded Code List (PBCL) currently used by laboratories to electronically report results. Given advances in stool microbiology testing with a move to nonculture, PCR-based methods, the way microbiology results are reported and coded via PBCL needs to be reviewed and modernized.

Introduction

Background

The burden of infectious intestinal disease (IID) is considerable. The World Health Organization (WHO) estimated that foodborne disease from 22 pathogens accounted for 22 diseases resulted in 2 billion cases, over 1 million deaths, and 78.7 million disability-adjusted life years in 2010 [1]. The IID in the United Kingdom (IID2 study) [2] reported 274 cases per 1000 person-years, with 17.7 (95% CI 14.4-21.8) presenting to primary care. However, this may be an underestimate. Less restrictive – more representative (of coding practice) diagnostic criteria would greatly increase, for example, their estimate of norovirus by 26% to 59/1000 (95% CI 52.32-64.98) person-years equating to 3.7 (3.3-4.1) million infections annually [3].

Reported incidence rates for IID vary between 0.5% and 20% annually in the developed world [4-9]. Variation can be greatly attributed to underreporting and data types used to calculate rates [10]. Data used to report IID rates include: primary care records, hospital and other secondary care settings, prospective and retrospective surveys or questionnaires, notifications of disease to authorities, and reports of laboratory detection of pathogens [7,11,12]. Studies have concluded that approximately 1 in 20 IID patients in the community consult a general practitioner (GP) [7,13,14], hence incidence rates calculated based on primary care data are 0.5-3.3%—much lower than rates calculated with other methods [4,13,14].

Published variations may also be caused by imprecise or interchangeable use of the terms such as "diarrheal disease," "acute infectious gastroenteritis" and "IID" and differing methods for describing cases, underscoring the importance of transparency when defining the disease [6,15]. The more general term "diarrheal disease" is used by the WHO and others in international public health as a symptom-based definition: infectious diarrhea and/or vomiting [6,11,16,17]. The terms "IID" and "acute gastroenteritis" tend to be more limited terms used to define patients with loose stools and/or vomiting for specific time periods and excluding chronic infections. Generally IID is defined as lasting less than 2 weeks, in the absence of known noninfectious causes, preceded by 2-3 symptom-free weeks [14,15]. Many studies list pathogens in their definition of IID or acute gastroenteritis; chronic or systemic conditions such as typhoid/paratyphoid and *Helicobacter* infections are often excluded [2,13].

Ontologies provide a method of systematically and transparently defining concepts and their relationships. They are used to clarify case finding and more accurately calculate disease incidence based on disease definitions that balance sensitivity and specificity [18,19]. In this study, we used a three-layer approach developed previously by the University of Surrey to develop an IID ontology [18]; we then used the ontological definition to calculate the incidence rate. The three-layer approach, an iterative process, includes development of disease concepts into an ontology, code collection, and logical data extraction [20].

UK general practice is highly computerized. Electronic registration–based systems ensure accurate denominators, and data from general practice provide opportunities for health research [21,22]. Most consultations are recorded on computers with key data—diagnosis, symptoms, investigative tests, and treatments—using a system called the Read codes [23]. The majority of UK practices are electronically connected to pathology laboratories, with generalized pathology results coded back into clinical records. Any laboratory results indicating pathogen detection should be coded directly by the clinician.

Objectives

We aimed to test new technologies that provide general practitioners near real-time test results for a wide range of pathogens associated with IID [24]. We carried out this analysis to determine IID incidence from routine data using an ontological approach to make case finding more transparent and allow comparisons to other studies and data. We compared rates calculated using standard Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) and ICD-10 definitions with an ontological approach and reported impact on incidence rate.

Methods

IID Case Definition

We reviewed common IID case definitions published in the literature and standard coding systems used to record IID diagnoses in primary care settings and chose three IID definitions (Textbox 1).

XSL•FO

RenderX

Textbox 1. Description of IID case definitions chosen for this study.

RSC definition

- Based on WHO's International Classification of Diseases, ICD-8/9 versions, infectious intestinal diseases chapter

- Used for RCGP RSC weekly returns report

- Includes all codes falling into the infectious intestinal disease group of infectious and parasitic diseases within the concept hierarchies of the 5-byte Version 2 read Code system (A00-A09 codes)

ICD-10 definition

- Based on WHO's International Classification of Diseases, ICD-10 version, infectious intestinal diseases chapter, all of which fall within A00-A09 chapter. More limited than the RSC definition

- Subset of ICD-8/9 and RSC definition due to exclusion of codes such as Helicobacter, nonintestinal Salmonella infections, Astrovirus, Calicivirus, and redundant codes

Ontological definition

- Based on IID case definition used during the Second Study of Infectious Intestinal Disease in the Community (IID2 Study)

- Includes all codes within the restricted ICD-10 definition, plus additional diagnostic codes that directly or partially map to IID2 case definition even though they fall outside of A00-A09 infectious intestinal disease group. Investigation and process of care codes that directly map to the case definition are also included

- The codes do not all necessarily fall into the A00-A09 infectious intestinal disease hierarchy used by the RSC and ICD-10 to define IID. This definition was based on the established case definition and was developed using an ontological approach designed to include all definite and possible IID cases recorded by clinicians in the RSC network

---

In the United Kingdom, the RCGP RSC case definition used for calculating weekly incidence of IID for the RSC's weekly communicable and respiratory disease report is the established "gold standard" for surveillance [14,25]. IID incidence rates for the RSC weekly report are generated using codes from the IID chapter of Read codes version 2 (5-byte set), the GP coding system most commonly used in primary care since 1985 to enter data into electronic health records. The RSC definition includes Read codes for conditions in WHO's International Classification of Diseases ICD-8/9 infectious intestinal diseases chapter (A00-A09 codes) [25]. To maintain consistency while monitoring long-term year-over-year trends in infections and outbreaks, RSC has conducted IID surveillance following the ICD-9 infectious intestinal diseases chapter, and as a result, many conditions not currently included in the newer WHO ICD-10 definition of IID continued to be included in the weekly returns report after ICD-10 was released [25]. To examine coding differences and relationships, we mapped IID codes between three ICD classifications and back to RSC weekly report codes.

For the ontological case definition of IID, we selected the more restrictive, well-documented case definition used during the Second Study of Infectious Intestinal Disease in the Community (IID2 Study), an extensively published, longitudinal study of IID incidence carried out in UK primary care [2,14,26]. The study defines IID as an infectious intestinal condition always causing diarrhea and sometimes other symptoms such as vomiting or nausea lasting 2 weeks or less [26].

IID Ontology Development and Code Mapping

We used a three-level approach previously developed by the University of Surrey to establish an ontology based on IID case definition [20]. We formalized this ontology using Protégé, which is supported by grant GM10331601 from the National

Institute of General Medical Sciences of the US National Institutes of Health [27].

The design of the ontology followed the structure used in problem-orientated records (POMR) and their associated coding system. This has its roots in the work of Lawrence Weed who created the idea of separating subjective (history) from objective (findings) and analysis (often diagnosis or problem) from plan (prescription or treatment). This is known internationally as Weed's SOAP [28-30]. The classes in our ontology (Multimedia Appendix 1) broadly followed the components of SOAP: subjective (S), clinical features; objective (O), findings from laboratory tests, but could include objective clinical features such as fever if measured; analysis (A), the problem title or diagnoses; and plan (P), which includes the process of care code (which are often nonspecific) and an prescription or referral for further care. The computerized medical record (CMR) systems in the United Kingdom were historically strictly problem orientated, though those that are now in ascendency are more episode orientated [31]. The coding systems used within these systems have historically been hierarchical and used "chapters" that fit with the POMR structure [23].

We applied the ontology to the Read Code list by searching for codes indicative of IID diagnosis and mapped each into one of the following three classes [18,32]. Complete ontology and code lists are presented in supplementary tables (Multimedia Appendices 1 and 2).

1. **Direct mapping class:** All codes included in the direct mapping class indicate a clinician's intention to record a definite IID diagnosis. Diagnostic codes fall into WHO's ICD-10 infectious intestinal diseases chapter (A00-A09 codes) [25] and the infectious intestinal disease group of infectious and parasitic diseases within concept hierarchies of the 5-byte version 2 Read code system. Additional codes

relate to investigative tests indicating laboratory detection of IID pathogens and processes of care indicating notification of IID.

2. **Partial mapping class:** All codes classified as partially mapping indicate a probable case of IID. These codes fall into the infectious intestinal disease group or other groups including gastrointestinal symptoms and other bacterial/infectious/ parasitic/digestive diseases. Additional codes relate to general IID investigations, therapies, symptoms, or process of care codes.

3. **No clear mapping class:** All codes included in this class indicate possible IID cases but do not clearly map to IID diagnosis, investigation, or symptom (eg, other viral enteritis).

Codes that refer to chronic conditions or non-intestinal conditions were defined as not mapping to IID and were excluded (eg, *Helicobacter*, *Salmonella* arthritis). We found that case finding was barely affected by the inclusion of codes in the least restrictive "no clear mapping" class and therefore did not use these codes in any analyses.

Cohort Identification

This study used primary care data recorded during a 52-week period spanning July 2014-July 2015 from the RCGP RSC, a sentinel network representative of the English population [33]. The cohort included patients with a recorded event, registered for the entire period. These data were used to determine the denominator. Data were extracted using SQL (Structured Query Language) software [34].

Case Finding and Rate Calculations

We calculated case numbers and incidence rates for the three IID definitions. When clinicians record a diagnosis, they assign episode type, which differentiates incident (first, new) cases from prevalent (follow-up, ongoing) cases. Records with "first" or "new" episode types were counted when counting cases and calculating incidence rates using diagnostic codes. When cases were found using directly mapping investigation and process care codes, all episode types were included because it is not standard clinical practice to code these events as "first" or "new." Patients with excessive IID diagnostic records (>4 per year) were excluded from case counts as they likely had chronic gastrointestinal conditions, although this represented fewer than 10 people over the one-year study period.

Concepts Supporting Case Finding

We further investigated differences between case definitions and the validity of using an ontological case definition by searching patients who had been already counted as a case for codes relating to four supporting concepts: (1) symptoms (diarrhea, vomiting, and fever), (2) pathology investigations (stool sample sent to laboratory, to test for specific pathogens), (3) process of care (notification of dysentery or food poisoning), and (4) therapies (loperamide or oral rehydration therapy). We used a 2-week sliding window due to IID's acute nature: all events for supporting concepts recorded with any episode type had to occur 2 weeks before or after the patient's diagnosis event to be included. In addition, multiple events coded for any

one factor within the 2-week window (eg, three investigation codes in one week) were counted as one event. Complete code lists for supporting factors are presented in Multimedia Appendix 2. We counted occurrences of each of the four supporting concepts and created Venn diagrams using R software [34].

The "Integrate" study received a favorable ethical opinion from the NHS NRES Committee North West-Greater Manchester East (Ref: 15/NW/0233). Patient-level data were automatically extracted and pseudonymized at the point of extraction. Data were stored at the University of Surrey Clinical Informatics and Health Outcomes Research Group data and analysis hub such that patients could not be identified from records used during the study.

Results

We identified an initial cohort (N=1,120,490) used to count cases and calculate incidence rates from the RCGP RSC population among all registered patients with at least one recorded event during a 52-week period spanning ISO 2014-W30 to ISO 2015-W29.

The results of the ontology can be found online (http://webprotege.stanford.edu) under the title "IID infectious intestinal disease ontology."

Use of the ICD-10 case definition identified 4736 cases of IID within the cohort, compared with 5966 cases found with the RSC definition (Figure 1).

Application of the ICD-10 definition when selecting Read codes resulted in a more limited code list (90 codes in ICD-9 reduced to 70 codes in ICD-10). This reduction is due to the removal of codes for *Helicobacter* and specific nonintestinal *Salmonella* infections; codes for other specific bacterial and viral infections (Arizona paracolon bacilli, Astrovirus, Calicivirus); and general infection codes that appeared redundant. Until recently, these codes were included in the RSC weekly report which, for consistency in surveillance of disease trends, continued following the ICD-9 system.

A key difference between ICD-10 and RSC weekly report code lists was the inclusion of *Helicobacter pylori* in the RSC definition , with 25% (306/1230) of cases captured within the RSC definition being recorded as *Helicobacter* codes. Although this condition is not included in the IID chapter of ICD systems, *H. pylori* infection is included in the IID chapter of the Read code system and therefore has been historically monitored in the RSC weekly report as IID. As *H. pylori* prevalence rates in Europe are at least as high as IID rates [35], its inclusion in IID surveillance could affect disease trend monitoring.

Using the ontological approach, we identified 5712 more cases than the ICD-10 definition and 4482 more cases than the RSC definition within the same cohort (Figure 1). Of the additional ontological cases, 77% (4399/5712) were recorded using specific gastroenteritis codes; 10.2% (582/5712) were coded as diarrhea and vomiting, first or new episodes; and 9.6% (546/5712) were recorded with direct pathology investigation codes (Table 1).

Table 1. Counts of additional ontological cases by code type (number of additional ontological cases not included in other case definitions=5712, data for period ISO 2014-W30 to ISO 2015-W29).

Code type	Code	Count of cases	Additional ontological cases (percentage)
Gastroenteritis, toxic gastroenteritis	J43-1 J43..11	4399	77.0
Diarrhea and vomiting	19G%	582	10.2
Clostridium difficile infection	A3Ay2%	145	2.5
Direct pathology investigation	Multiple; see Multimedia Appendix 2	546	9.6
Direct process of care	65V1%, 65V2%	29	0.5

Table 2. IID incidence and case counts (Data for period ISO 2014-W30 to ISO 2015-W29, weekly denominator N=1,120,490).

Definition	Count of cases	Annual person-time rates (per 1000 person-time units)
Standard RSC	5966	5.32 (95% CI 5.19-5.46)
ICD-10	4736	4.23 (95% CI 4.11-4.35)
Ontological	10,448	9.32 (95% CI 9.15-9.50)

Table 3. Mean weekly incidence rates and case counts (Data for period ISO 2014-W30 to ISO 2015-W29, weekly denominator N=1,120,490).

Definition	Mean weekly count of cases	Incidence rate (per 100,000/week)
Standard RSC	114.73	10.24 (95% CI 8.55-12.12)
ICD-10	91.08	8.13 (95% CI 6.70-9.87)
Ontological	200.92	17.93 (95% CI 15.63-20.41)

Using the ontological definition for case finding resulted in an annual percentage incidence rate of 0.93% (10,448/1,120,490) compared with 0.42% (4736/1,120,490) under the ICD-10 definition and 0.53% (5966/1,120,490) under the RSC definition. Annual person-time rate per 1000 person-time units for the standard RSC definition was 5.32 (95% CI 5.19-5.46), for the ICD-10 definition was 4.23 (95% CI 4.11-4.35), and for the ontological definition was 9.32 (95% CI 9.15-9.50; Table 2).

Mean weekly incidence rate was 10.24 per 100,000 (95% CI 8.55-12.12) for the RSC definition, 8.13 per 100,000 (95% CI 6.70-9.87) for the ICD-10 definition, and 17.93 per 100,000 (95% CI 15.63-20.41) for the ontological definition (Table 3).

Event counts of four supporting concepts within the 2-week period preceding or following case finding were consistent across IID definitions (Figures 2-4, Tables 4-6), with categories differing by ±1-2%.

Consistency of results supports the use of the ontological definition, as supporting concept codes are specific to acute IID. For the three definitions, majority of cases (61.67% [3679/5966], 59.18% [2803/4736], and 62.62% [6543/10,448]) had no supporting concepts recorded within the 2-week sliding window. In addition, proportion of laboratory results associated with positive test results (ie, directly mapping to IID case definition) was 19.7% (546/2775).

Table 4. Counts of supporting factors for RSC defined cases (N=5966).

Code category	Number of events coded	Percentage of RSC cases
Symptoms	295	4.94
Investigations	708	11.87
Therapies	705	11.82
Process of care	12	0.20
Symptoms and investigations	207	3.47
Symptoms and therapies	70	1.17
Symptoms and process of care	2	0.03
Investigations and therapies	147	2.46
Investigations and process of care	32	0.54
Therapies and process of care	1	0.02
Symptoms, investigations, and therapies	83	1.39
Symptoms, investigations, and process of care	14	0.23
Symptoms, therapies, and process of care	0	0.00
Investigations, therapies, and process of care	8	0.13
All supporting concepts	3	0.05
Number of cases with any of the above	2287	38.33
Number of cases with none of the above	3679	61.67

Table 5. Counts of supporting factors for ICD-10 defined cases (N=4736).

Code category	Number of events coded	Percentage of ICD-10 cases
Symptoms	235	4.96
Investigations	588	12.42
Therapies	587	12.39
Process of care	8	0.17
Symptoms and investigations	195	4.12
Symptoms and therapies	58	1.22
Symptoms and process of care	2	0.04
Investigations and therapies	129	2.72
Investigations and process of care	30	0.63
Therapies and process of care	0	0.00
Symptoms, investigations, and therapies	76	1.60
Symptoms, investigations, and process of care	14	0.30
Symptoms, therapies, and process of care	0	0.00
Investigations, therapies, and process of care	8	0.17
All supporting concepts	3	0.06
Number of cases with any of the above	1933	40.82
Number of cases with none of the above	2803	59.18

Table 6. Counts of supporting factors for cases defined ontologically (N=10,448).

Code category	Number of events coded	Percentage of ontological cases
Symptoms	632	6.05
Investigations	1050	10.05
Therapies	1337	12.80
Process of care	10	0.10
Symptoms and investigations	269	2.57
Symptoms and therapies	188	1.80
Symptoms and process of care	2	0.02
Investigations and therapies	238	2.28
Investigations and process of care	34	0.33
Therapies and process of care	2	0.02
Symptoms, investigations, and therapies	112	1.07
Symptoms, investigations, and process of care	16	0.15
Symptoms, therapies, and process of care	0	0.00
Investigations, therapies, and process of care	12	0.11
All supporting concepts	3	0.03
Number of cases with any of the above	3905	37.38
Number of cases with none of the above	6543	62.62

Figure 1. Total number of cases identified using three differing definitions of IID (RSC, ICD-10 and ontological). Cohort includes all registered patients in the RCGP RSC primary care database with at least one recorded event during a 52-week period spanning ISO 2014-W30 to ISO 2015-W29 (initial cohort, N=1120490).
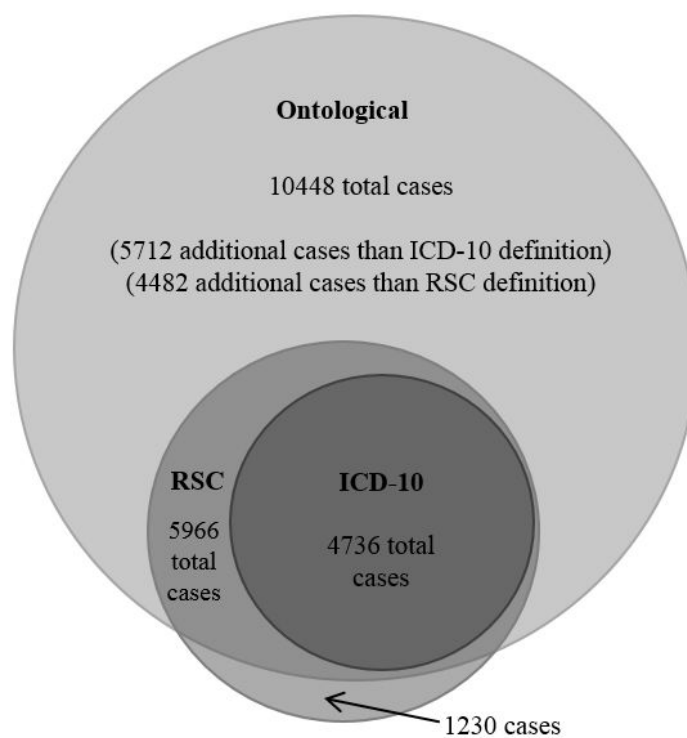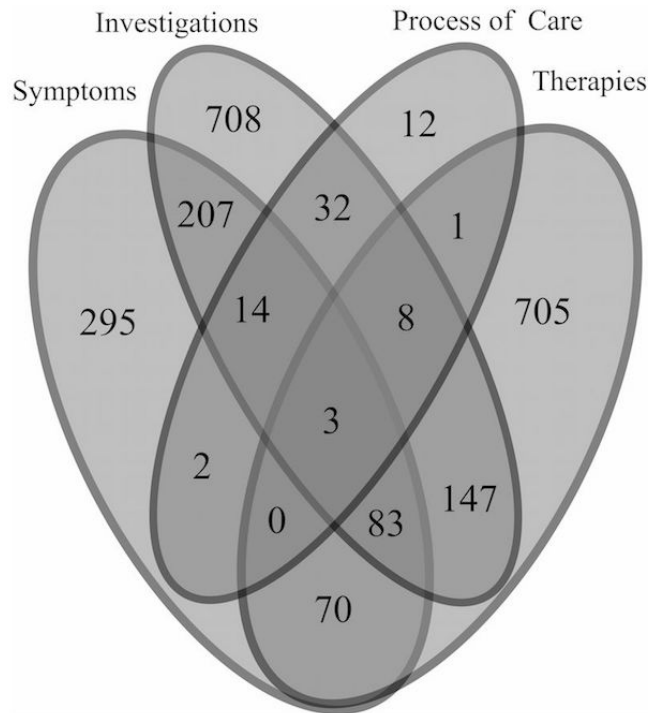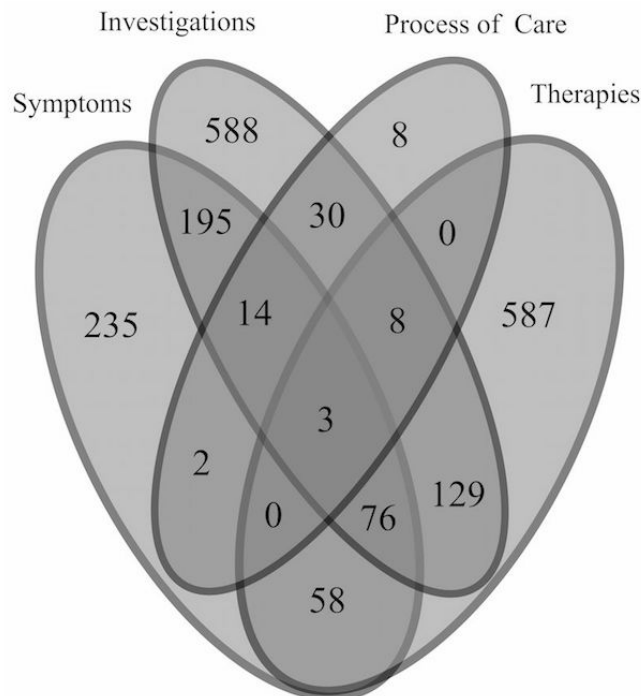
Figure 2. Count of events found using codes for supporting factors (symptom, investigation, process of care, and/or therapy) for cases identified using the standard RCGP RSC IID definition (ISO 2014-W30 to ISO 2015-W29). For Figures 2-4, events found using two-week sliding window: all recorded events for supporting concepts recorded with any episode type had to occur two weeks before or after the patient's diagnosis event to be included. Multiple events coded for any one factor within the two-week window of the case finding were counted as one event.

Figure 3. Count of events found using codes for supporting factors (symptom, investigation, process of care, and/or therapy) for cases identified using the ICD-10 IID definition (ISO 2014-W30 to ISO 2015-W29).

Figure 4. Count of events found using codes for supporting factors (symptom, investigation, process of care, and/or therapy) for cases identified using the ontological IID definition (ISO 2014-W30 to ISO 2015-W29).
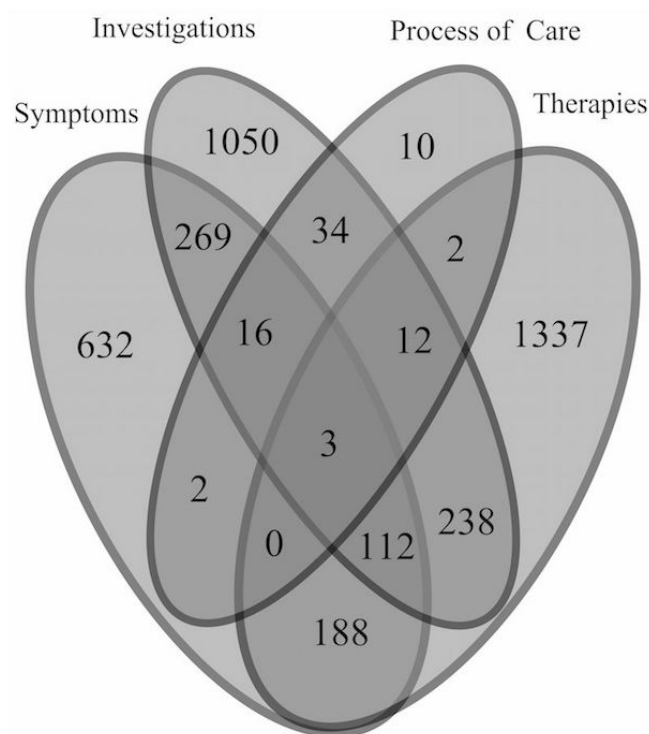
Discussion

Principal Findings

An ontological approach to IID case finding changed IID incidence rate, increasing case detection. The ontological approach is also more transparent and independent of coding systems.

The ontological approach may address elements of IID underestimation due to low rates of case finding using electronic data alone [36], depending upon the case definition used [8,15,37]. However, the major limitation to accurate case finding remains that many community cases of IID do not seek health care [38].

GPs appear more likely to enter symptom codes, which from the ontological perspective are less helpful as they overlap with other conditions rather than being specific to IID, unless the symptoms are supported by another code indicating pathogen detection [39]. Results of the ontological approach have highlighted how use of symptom codes contributes to underreporting IID patients who do not have appropriate diagnostic or surveillance codes entered into the patient record.

Implications of Findings for Clinical Practice

An ontological approach provides insights into what types of data are available for case ascertainment. Although this approach offers benefits, and has limitations, our recommendation is to start by making the laboratory results recorded much more specific.

The mechanism for transferring results from stool sampling to GPs needs to be updated. Currently UK laboratories electronically report stool sample results to GPs using the Pathology Bounded Code List (PBCL), a subset of Read codes. However, there is no standardized algorithm for reporting results, and the PBCL code list for Microscopy, Culture & Sensitivity (MC&S) results has not kept pace with developments in pathology services. For example, typical laboratory protocol is to report one generic stool sampling code per test request, regardless of the range of pathogens being screened or detected, or of the sensitivity or specificity of the testing method. When a GP receives electronic results of a stool sample, the electronic report only contains generic MC&S Read codes, indicating that a stool sample was analyzed. This is followed by a "free-text" message (ie, not coded) indicating any detected pathogens. If pathogens are detected, the clinician must then code this information manually into the computerized medical record (CMR) system. This means that, inevitably, laboratory findings are under-coded. Furthermore, for some pathogens there is only one PBCL code specifically for test requests, not for recording results. Many IID pathogens have no designated PBCL code at all, and where appropriate pathogen codes are available, they are often not used. Given likely advances in stool microbiology testing in the future, with a move away from MC&S to nonculture, PCR-based methods, the way microbiology results are reported and coded via PBCL needs to be reviewed and modernized. There might be scope to draw lessons from biochemistry and hematology where, with the exception of glucose provenance and use of nonnumeric keys [39] and the use of nonnumeric characters, results with coded data are generally readily filed into the CMR system.

Limitations

The principal limitation of this study is the lack of a gold standard; we do not know the "true" incidence of IID. There

has been no back-to-case records review to validate this approach, though the authors have gone back to records to demonstrate the reliability of case finding from clinical records in other domains, for example, chronic kidney disease [40] and diabetes [41,42]. We have also reported where we consider conclusions to be unsafe because the wrong codes were selected [43].

In addition, ontologies are developed as an iterative process; therefore, we recommend testing by running data extracts to improve sensitivity and specificity. Our ontology is online and may be superseded by better laboratory coding, advances in near-patient testing, or other unforeseen advances. For example, there was no attempt to include social media data in this exercise. Techniques are emerging to do this and should be considered as part of future investigations and for inclusion in the subjective elements of the ontology [44,45].

Bias of many types can affect the quality of data recording in CMR systems. This can be around financial incentives to adopt CMR systems which then may not get used [46]; and around pressures within systems to either investigate, refer, or prescribe more (or less) depending on the constraints within the individual health care system at the time. These effects are probably best reported for drug safety studies where the availability of a large number of CMR records or administrative datasets had not obviated the need for other mechanisms of drug safety recording [47,48].

Finally, use of a new ontological approach to measuring disease incidence might result in further discrepancies between different surveillance systems that monitor the IID incidence. Harmonization of coding systems across different systems and countries is important from an epidemiological perspective to ensure that estimates of disease burden are comparable.

Conclusions

Our study indicates that use of the standard definition of IID to identify cases in primary care results in the underestimation of disease incidence. To capture a larger proportion of new IID cases in primary care, an ontological approach should be adopted to expand the case definition to include those patients with codes falling outside more restrictive standard definitions, as well as improving the PBCL coding list used by laboratories returning pathology results. Given the high burden of IID in the community, identifying what specific organisms are circulating within a community would help GPs and public health services. For GPs this would reinforce the importance of stressing simple and important control measures, such as hand washing, and trigger the implementation-specific interventions for specific infections. Local and regional public health services would more accurately know the disease burden and be able to intervene; nationally and internationally more accurate data would enable better policy evaluation and development around hygiene and food chain management.

Using these approaches will provide a better picture for clinicians, epidemiologists, and public health officials of the burden of IID in the community and the impact of seasonal infectious disease outbreaks.

Conflicts of Interest

SdeL has a pending application for feasibility study funding with Takeda to look at the potential to identify household transmission from routine data and to assess gastroenteritis economic impact; SOB will advise this project. All other authors report no conflicts.

Multimedia Appendix 1

Formal ontology developed based on infectious intestinal disease case definition.

[XLSX File (Microsoft Excel File), 12KB - medinform_v5i3e34_app1.xlsx]

Multimedia Appendix 2

Diagnostic and supporting factor codes and code mapping classes selected based on infectious intestinal disease ontology.

[XLSX File (Microsoft Excel File), 22KB - medinform_v5i3e34_app2.xlsx]

XSL•FO

RenderX

References

1. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, et al. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. PLoS Med 2015 Dec 03;12(12):e1001921 [FREE Full text] [doi: 10.1371/journal.pmed.1001921] [Medline: 26633831]

2. O'Brien SJ, Rait G, Hunter PR, Gray JJ, Bolton FJ, Tompkins DS, et al. Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). BMC Med Res Methodol 2010 May 05;10:39 [FREE Full text] [doi: 10.1186/1471-2288-10-39] [Medline: 20444246]

3. Harris JP, Iturriza-Gomara M, O'Brien SJ. Re-assessing the total burden of norovirus circulating in the United Kingdom population. Vaccine 2017 Feb 07;35(6):853-855 [FREE Full text] [doi: 10.1016/j.vaccine.2017.01.009] [Medline: 28094075]

4. de Lusignan S, Correa A, Pathirannehelage S, Byford R, Yonova I, Elliot AJ, et al. RCGP Research and Surveillance Centre Annual Report 2014-2015: disparities in presentations to primary care. Br J Gen Pract 2017 Jan;67(654):e29-e40. [doi: 10.3399/bjgp16X688573] [Medline: 27993900]

5. de Wit MA, Koopmans MP, Kortbeek LM, Wannet WJ, Vinjé J, van Leusden F, et al. Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. Am J Epidemiol 2001 Oct 01;154(7):666-674. [Medline: 11581101]

6. Flint JA, Van Duynhoven YT, Angulo FJ, DeLong SM, Braun P, Kirk M, et al. Estimating the burden of acute gastroenteritis, foodborne disease, and pathogens commonly transmitted by food: an international review. Clin Infect Dis 2005 Sep 01;41(5):698-704. [doi: 10.1086/432064] [Medline: 16080093]

7. Palmer S, Houston H, Lervy B, Ribeiro D, Thomas P. Problems in the diagnosis of foodborne infection in general practice. Epidemiol Infect 1996 Dec;117(3):479-484 [FREE Full text] [Medline: 8972672]

8. Scallan E, Majowicz SE, Hall G, Banerjee A, Bowman CL, Daly L, et al. Prevalence of diarrhoea in the community in Australia, Canada, Ireland, and the United States. Int J Epidemiol 2005 Apr;34(2):454-460. [doi: 10.1093/ije/dyh413] [Medline: 15659464]

9. Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, et al. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. BMJ 1999 Apr 17;318(7190):1046-1050 [FREE Full text] [Medline: 10205103]

10. Haagsma JA, Geenen PL, Ethelberg S, Fetsch A, Hansdotter F, Jansen A, Med-Vet-Net Working Group. Community incidence of pathogen-specific gastroenteritis: reconstructing the surveillance pyramid for seven pathogens in seven European Union member states. Epidemiol Infect 2013 Aug;141(8):1625-1639. [doi: 10.1017/S0950268812002166] [Medline: 23013659]

11. Kosek M, Bern C, Guerrant RL. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. Bull World Health Organ 2003;81(3):197-204 [FREE Full text] [Medline: 12764516]

12. Viviani L, van der Es M, Irvine L, Tam CC, Rodrigues LC, Jackson KA, IID2 Study Executive Committee. Estimating the incidence of acute infectious intestinal disease in the community in the UK: a retrospective telephone survey. PLoS One 2016 Jan 25;11(1):e0146171 [FREE Full text] [doi: 10.1371/journal.pone.0146171] [Medline: 26807916]

13. de Wit MA, Kortbeek LM, Koopmans MP, de Jager CJ, Wannet WJ, Bartelds AI, et al. A comparison of gastroenteritis in a general practice-based study and a community-based study. Epidemiol Infect 2001 Dec;127(3):389-397. [Medline: 11811870]

14. Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, Hunter PR, IID2 Study Executive Committee. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut 2012 Jan;61(1):69-77 [FREE Full text] [doi: 10.1136/gut.2011.238386] [Medline: 21708822]

15. Majowicz SE, Hall G, Scallan E, Adak GK, Gauci C, Jones TF, et al. A common, symptom-based case definition for gastroenteritis. Epidemiol Infect 2008 Jul;136(7):886-894. [doi: 10.1017/S0950268807009375] [Medline: 17686196]

16. World Health Organization. 2017 May. Diarrhoel disease URL: http://www.who.int/mediacentre/factsheets/fs330/en/ [accessed 2016-12-01] [WebCite Cache ID 6HmxnvD0j]

17. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]

18. de Lusignan S, Liaw ST, Michalakidis G, Jones S. Defining datasets and creating data dictionaries for quality improvement and research in chronic disease using routinely collected data: an ontology-driven approach. Inform Prim Care 2011;19(3):127-134 [FREE Full text] [Medline: 22688221]

19. Rollason W, Khunti K, de Lusignan S. Variation in the recording of diabetes diagnostic data in primary care computer systems: implications for the quality of care. Inform Prim Care 2009;17(2):113-119 [FREE Full text] [Medline: 19807953]

20. de Lusignan S. In this issue: Ontologies a key concept in informatics and key for open definitions of cases, exposures, and outcome measures. J Innov Health Inform 2015 Jul 10;22(2):170 [FREE Full text] [Medline: 26245238]

21. de Lusignan S, Metsemakers JF, Houwink P, Gunnarsdottir V, van der Lei J. Routinely collected general practice data: goldmines for research? A report of the European Federation for Medical Informatics Primary Care Informatics Working

Group (EFMI PCIWG) from MIE2006, Maastricht, The Netherlands. Inform Prim Care 2006;14(3):203-209 [FREE Full text] [Medline: 17288707]

22. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. Fam Pract 2006 Apr;23(2):253-263 [FREE Full text] [doi: 10.1093/fampra/cmi106] [Medline: 16368704]

23. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. Inform Prim Care 2005;13(1):65-70 [FREE Full text] [Medline: 15949178]

24. Integrate Project. Integrate: Fully-integrated, real-time detection, diagnosis and control of community diarrhoeal disease clusters and outbreaks URL: http://www.integrateproject.org.uk [WebCite Cache ID 6sdJgGUW5]

25. Royal College of General Practitioners. 2011. Research & Surveillance Centre: Weekly Returns Service Annual Report 2011 URL: http://www.rcgp.org.uk/rsc [accessed 2017-08-11] [WebCite Cache ID 6sdMeH7AM]

26. Tam C, Viviani L, Adak B, Bolton E, Dodds J, Cowden J, IID2 Study Executive Committee. UK Food Standards Agency. 2012. The Second Study of Infectious Intestinal Disease in the Community (IID2 Study): Final report URL: https://www.food.gov.uk/sites/default/files/711-1-1393_IID2_FINAL_REPORT.pdf [accessed 2017-08-16] [WebCite Cache ID 6sl2LaUmT]

27. WHO. 2016. ICD-10 Version:2016: International Statistical Classification of Diseases and Related Health Problems 10th Revision URL: http://apps.who.int/classifications/icd10/browse/2016/en [accessed 2015-05-01] [WebCite Cache ID 5EmVp119m]

28. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 21;278(12):652-7 concl. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]

29. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 14;278(11):593-600. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]

30. Wright A, Sittig DF, McGowan J, Ash JS, Weed LL. Bringing science to medicine: an interview with Larry Weed, inventor of the problem-oriented medical record. J Am Med Inform Assoc 2014;21(6):964-968 [FREE Full text] [doi: 10.1136/amiajnl-2014-002776] [Medline: 24872343]

31. de Lusignan S, Liaw ST, Dedman D, Khunti K, Sadek K, Jones S. An algorithm to improve diagnostic accuracy in diabetes in computerised problem orientated medical records (POMR) compared with an established algorithm developed in episode orientated records (EOMR). J Innov Health Inform 2015 Jun 05;22(2):255-264 [FREE Full text] [Medline: 26245239]

32. Liyanage H, Krause P, De Lusignan S. Using ontologies to improve semantic interoperability in health data. J Innov Health Inform 2015 Jul 10;22(2):309-315 [FREE Full text] [Medline: 26245245]

33. Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, et al. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. BMJ Open 2016 Apr 20;6(4):e011092 [FREE Full text] [doi: 10.1136/bmjopen-2016-011092] [Medline: 27098827]

34. R-Project. The R Project for Statistical Computing URL: https://www.r-project.org/ [accessed 2016-05-01] [WebCite Cache ID 6sdTpeW32]

35. Roberts SE, Morrison-Rees S, Samuel DG, Thorne K, Akbari A, Williams JG. Review article: the prevalence of Helicobacter pylori and the incidence of gastric cancer across Europe. Aliment Pharmacol Ther 2016 Feb;43(3):334-345. [doi: 10.1111/apt.13474] [Medline: 26592801]

36. Gu Y, Kennelly J, Warren J, Nathani P, Boyce T. Automatic detection of skin and subcutaneous tissue infections from primary care electronic medical records. Stud Health Technol Inform 2015;214:74-80. [Medline: 26210421]

37. Wilson SE, Deeks SL, Rosella LC. Importance of ICD-10 coding directive change for acute gastroenteritis (unspecified) for rotavirus vaccine impact studies: illustration from a population-based cohort study from Ontario, Canada. BMC Res Notes 2015 Sep 15;8:439 [FREE Full text] [doi: 10.1186/s13104-015-1412-5] [Medline: 26374397]

38. Gibbons CL, Mangen MJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, Burden of Communicable diseases in Europe (BCoDE) consortium. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health 2014 Feb 11;14:147 [FREE Full text] [doi: 10.1186/1471-2458-14-147] [Medline: 24517715]

39. Harkness EF, Grant L, O'Brien SJ, Chew-Graham CA, Thompson DG. Using read codes to identify patients with irritable bowel syndrome in general practice: a database study. BMC Fam Pract 2013 Dec 02;14:183 [FREE Full text] [doi: 10.1186/1471-2296-14-183] [Medline: 24295337]

40. Anandarajah S, Tai T, de Lusignan S, Stevens P, O'Donoghue D, Walker M, et al. The validity of searching routinely collected general practice computer data to identify patients with chronic kidney disease (CKD): a manual review of 500 medical records. Nephrol Dial Transplant 2005 Oct;20(10):2089-2096. [doi: 10.1093/ndt/gfi006] [Medline: 16030033]

41. de Lusignan S, Khunti K, Belsey J, Hattersley A, van Vlymen J, Gallagher H, et al. A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. Diabet Med 2010 Feb;27(2):203-209. [doi: 10.1111/j.1464-5491.2009.02917.x] [Medline: 20546265]

42. Hassan Sadek N, Sadek AR, Tahir A, Khunti K, Desombre T, de Lusignan S. Evaluating tools to support a new practical classification of diabetes: excellent control may represent misdiagnosis and omission from disease registers is associated with worse control. Int J Clin Pract 2012 Sep;66(9):874-882 [FREE Full text] [doi: 10.1111/j.1742-1241.2012.02979.x] [Medline: 22784308]

43. de Lusignan S, Sun B, Pearce C, Farmer C, Steven P, Jones S. Coding errors in an analysis of the impact of pay-for-performance on the care for long-term cardiovascular disease: a case study. Inform Prim Care 2014;21(2):92-101 [FREE Full text] [doi: 10.14236/jhi.v21i2.62] [Medline: 24841410]

44. Konovalov S, Scotch M, Post L, Brandt C. Biomedical informatics techniques for processing and analyzing web blogs of military service members. J Med Internet Res 2010 Oct 05;12(4):e45 [FREE Full text] [doi: 10.2196/jmir.1538] [Medline: 20923755]

45. Yom-Tov E, Borsa D, Cox IJ, McKendry RA. Detecting disease outbreaks in mass gatherings using Internet data. J Med Internet Res 2014 Jun 18;16(6):e154 [FREE Full text] [doi: 10.2196/jmir.3156] [Medline: 24943128]

46. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 7;160(1):48-54. [doi: 10.7326/M13-1531] [Medline: 24573664]

47. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. J Epidemiol Community Health 2014 Mar;68(3):283-287. [doi: 10.1136/jech-2013-202744] [Medline: 24248997]

48. Patadia VK, Schuemie MJ, Coloma P, Herings R, van der Lei J, Straus S, et al. Evaluating performance of electronic healthcare records and spontaneous reporting data in drug safety signal detection. Int J Clin Pharm 2015 Feb;37(1):94-104. [doi: 10.1007/s11096-014-0044-5] [Medline: 25488315]

49. Apollo Medical Software Solutions. Apollo Data Extraction: Turning data into information URL: http://www.apollo-medical.com/apollo-sql-suite/ [accessed 2017-07-01] [WebCite Cache ID 6sdR6CADj]

50. EMIShealth. EMIS Health Website URL: https://www.emishealth.com/ [accessed 2017-06-21] [WebCite Cache ID 6sdRxJoxq]

51. Vision Health. Vision URL: https://www.visionhealth.co.uk/ [accessed 2017-06-21] [WebCite Cache ID 6sdUWMrg5]

52. TPP-UK. SystmOne URL: http://www.tpp-uk.com/products/systmone [accessed 2017-08-11] [WebCite Cache ID 6sdaC9VHJ]

Abbreviations

CMR: computerized medical record
GP: general practitioner
ICD: International Classification of Diseases
IID: infectious intestinal disease
MC&S: Microscopy, Culture & Sensitivity
PBCL: Pathology Bounded Code List
RCGP: Royal College of General Practitioners
RSC: Research and Surveillance Centre
WHO: World Health Organisation

XSL•FO
RenderX

<u>Original Paper</u>

The Rules of Engagement: Perspectives on Secure Messaging From Experienced Ambulatory Patient Portal Users

Cynthia J Sieck[1*], MPH, PhD; Jennifer L Hefner[1*], MPH, PhD; Jeanette Schnierle[1], BS; Hannah Florian[1], BS; Aradhna Agarwal[1], BS; Kristen Rundell[1], MD; Ann Scheck McAlearney[1,2], ScD, MS

[1]The Ohio State University College of Medicine, Columbus, OH, United States
[2]The Ohio State University College of Public Health, Division of Health Services Management and Policy, Columbus, OH, United States
[*]these authors contributed equally

Corresponding Author:
Cynthia J Sieck, MPH, PhD
The Ohio State University College of Medicine
2231 North High St
277 Northwood-High Bldg
Columbus, OH, 43201
United States
Phone: 1 614 366 0218
Fax: 1 614 293 2715
Email: cynthia.sieck@osumc.edu

Abstract

Background: Patient portals have shown promise in engaging individuals in self-management of chronic conditions by allowing patients to input and track health information and exchange secure electronic messages with their providers. Past studies have identified patient barriers to portal use including usability issues, low health literacy, and concerns about loss of personal contact as well as provider concerns such as increased time spent responding to messages. However, to date, studies of both patient and provider perspectives on portal use have focused on the pre-implementation or initial implementation phases and do not consider how these issues may change as patients and providers gain greater experience with portals.

Objective: Our study examined the following research question: Within primary care offices with high rates of patient-portal use, what do experienced physician and patient users of the ambulatory portal perceive as the benefits and challenges of portal use in general and secure messaging in particular?

Methods: This qualitative study involved 42 interviews with experienced physician and patient users of an ambulatory patient portal, Epic's MyChart. Participants were recruited from the Department of Family Medicine at a large Academic Medical Center (AMC) and included providers and their patients, who had been diagnosed with at least one chronic condition. A total of 29 patients and 13 primary care physicians participated in the interviews. All interviews were conducted by telephone and followed a semistructured interview guide. Interviews were transcribed verbatim to permit rigorous qualitative analysis. Both inductive and deductive methods were used to code and analyze the data iteratively, paying particular attention to themes involving secure messaging.

Results: Experienced portal users discussed several emergent themes related to a need for greater clarity on when and how to use the secure messaging feature. Patient concerns included worry about imposing on their physician's time, the lack of provider compensation for responding to secure messages, and uncertainty about when to use secure messaging to communicate with their providers. Similarly, providers articulated a lack of clarity as to the appropriate way to communicate via MyChart and suggested that additional training for both patients and providers might be important. Patient training could include orienting patients to the "rules of engagement" at portal sign-up, either in the office or through an online tutorial.

Conclusions: As secure messaging through patient portals is increasingly being used as a method of physician-patient communication, both patients and providers are looking for guidance on how to appropriately engage with each other using this tool. Patients worry about whether their use is appropriate, and providers are concerned about the content of messages, which allow them to effectively manage patient questions. Our findings suggest that additional training may help address the concerns of both patients and providers, by providing "rules of engagement" for communication via patient portals.

Introduction

Patient portals provide access to information in the patient's electronic health record, in addition to serving as a platform to view and schedule appointments and engage in secure communication with providers [1]. These types of portals, typically accessed by patients through a website, are increasingly positioned as a central component of patient engagement in healthcare [2-5]. Specifically, portals have shown promise in engaging individuals in self-management of chronic conditions by allowing patients to input and track health information, facilitating communication between patients and providers, and providing access to consumer-friendly information about diseases [6-10].

One particular feature of patient portals, secure messaging, has seen a significant increase in use over time [11,12]. This popular feature allows patients and providers to communicate asynchronously, without waiting for the other to be available on the telephone. For example, through secure messaging, patients can provide updates on symptoms discussed during a visit or efficiently monitor the initiation of some types of medications. The literature suggests that secure messaging can facilitate access to care, improve patient satisfaction, and improve health outcomes [13,14]. Unlike the other features of a patient portal, such as viewing lab and test results or requesting appointments, secure messaging allows for the exchange of direct communication between patients and providers. As a result, studies suggest that a patient's relationship with a provider is a key predictor of the patient's intention to use secure messaging [15].

While secure messaging is a function generally desired by patients, both patients and providers share concerns about its use. Some worry about the loss of interpersonal contact [16-18] as well as about the privacy and security of information exchanged through a patient portal [17,19-22]. Additionally, providers have expressed concerns about the impact of secure messaging on their workload [20,23-26], noting that they are typically not reimbursed for this type of work [26-29].

One important limitation of this literature on patient portals is that studies of patient and provider perspectives on portal use focus on the pre-implementation or initial implementation phases and do not reflect how use and perspectives may change as users gain more experience. For example, surveys or interviews of providers are typically conducted before the implementation of the portal to gauge providers' willingness to accept the portal and inform decisions about portal design [30-33] or immediately after portal implementation [34-38]. A 2016 study of primary care providers' views on patient portals published in the *Journal of Medical Internet Research* included only 7 current portal users among the 20 interviewees [23]. The same is true for patients, with most qualitative studies involving only early or recent patient-portal adopters rather than experienced users

[6,21,39,40]; as a result, these studies are frequently focused on barriers to adoption [17,41-46].

While the perspectives of patients and providers are critical in the early phases of patient-portal implementation and use, there is a gap in the literature regarding how experienced users, both patients and providers, engage with portals and use secure messaging features. Our study aimed to address this gap by exploring the following research question: Within primary care offices with high rates of patient-portal use, what do experienced physician and patient users of the ambulatory portal perceive as the benefits and challenges of portal use in general and secure messaging in particular? Interviewing both physicians and patients with use experience allowed us to consider questions such as whether privacy and security are still prominent patient concerns among active, long-term users, and whether provider's work flow concerns persist once portal use is established within the office.

Methods

Study Design

We designed an exploratory qualitative study to improve our understanding of patients' and providers' perspectives on patient portals and the use of secure messaging within those portals. Our data were collected through telephone interviews with participants recruited for the study. Data were then iteratively analyzed, using both deductive and inductive methods, to characterize the themes we present in this paper. This study was approved by the study site's Institutional Review Board.

Study Setting

Our study took place at a large Midwestern Academic Medical Center (AMC) that uses Epic's MyChart, an interactive tethered patient portal that allows patients to view test and lab results, schedule appointments, request refills, and send secure messages to providers. Patients using the portal are presented with a notice on the secure messaging screen (1) telling them to use this feature for non-urgent messages only, (2) telling them to expect a response within 24-48 hours, (3) reminding them that their message becomes part of their medical record, and (4) telling them to call 911 if they feel their concern represents an emergency.

Since implementing MyChart across the entire AMC in 2012, over 35,000 patients have created a MyChart account, with the majority having logged on at least once. The demographics of portal users are skewed toward greater representation by females, whites, and patients between the ages of 36 and 54. Of the MyChart features available, messaging and viewing results are the most commonly used, followed by appointment scheduling. Across all departments in the AMC, Family Medicine providers have the highest percentage of active MyChart users (65% of their patients), followed by

Obstetrics/Gynecology (55%); other departments average between 35%-50%.

Study Sample

We recruited a purposive sample of patients and primary care physicians in the summer and fall of 2015. Interviewees were all experienced users of MyChart and included 13 Family Medicine providers in the Department of Family Medicine (DFM) and 29 of their patients who had at least one chronic condition. Patients were identified by their physician using the reporting function of the electronic health record (EHR). Inclusion criteria were having at least one cardiopulmonary condition and being among the most frequent users of MyChart when patients were rank ordered by frequency of message. Providers forwarded a recruitment e-mail from the study principal investigator (PI) to the top 25 frequent users identified in their query. The recruitment email explained the purpose of the study and provided a contact number for patients to call to schedule telephone interviews. Providers were recruited to participate in interviews through a similar e-mail sent directly from the study PI. Interviews lasted approximately 30 minutes, and all interviews were conducted by telephone and recorded.

Data Collection

We used two versions of a semistructured interview guide to conduct the interviews, drawing upon concerns about using portals identified in our literature review [16-29], as well as our own research questions related to the portal user experience. Interview questions for patients asked about motivations for using MyChart, how patients use MyChart, and perceptions about how MyChart impacts patient-provider communication. Providers were asked about the primary activities they completed on MyChart and their experiences with these activities, including releasing lab results and fielding patient questions via the portal. Providers were also asked about perceived impacts on the patient-provider relationship and challenges to engaging with patients through MyChart. Interviews were transcribed verbatim to permit rigorous analysis.

Analysis

Our analytic approach used both inductive and deductive methods iteratively, using a constant comparative analytic approach throughout the study [47]. First, a three-person coding team identified broad themes from the data and developed a preliminary non-mutually exclusive coding dictionary. This team also proposed new codes as patterns emerged from the data and as subsequent interviews were conducted, following the methods described by Constas [48]. While the three-person team made initial coding decisions, frequent meetings with the entire study team were held to discuss discrepancies, reach consensus, and ensure that saturation of concepts was reached. We used the Atlas.ti (version 6.0) qualitative data analysis software to support our analysis.

Results

We conducted 42 interviews of 29 patients and 13 primary care physicians. Our qualitative analysis of interview transcripts revealed five major themes related to the use of secure messaging within the patient portal, as well as a theme involving providers' perspectives about the need for training on portal use. Below we describe these themes related to benefits and concerns about secure messaging, including sub-themes about concerns from the perspectives of patients and providers, respectively. We conclude with an exploration of sub-themes around the need for "rules of engagement" to support portal use.

Perceived Benefits of Secure Messaging

Asynchronous Communication

Both patients and providers appreciated the ability to use secure messaging for communication. Most commonly, both groups felt that the ability for each party to respond according to their own schedule increased the efficiency of communicating. Several patients specifically mentioned the benefit of conversations that could occur asynchronously, according to the patient's and provider's individual schedules, without reliance on telephone calls to the office. For example, one patient described communication via MyChart in comparison with how he had to call the office before using MyChart:

> If I had a question for them, I would call in and deal with what seems to be a number of [people]. First you talk to the receptionist, and then you get to the nurse, and then you try to do the medication option. And call back when you get lost in the line of communication there some way.

Providers also described this benefit and noted increased efficiency in communications. A provider described it thus:

> Because sometimes, when it's a phone call, I'm not necessarily making the call. I let my staff do it. So it goes from me, to the staff, to the patient. So this way [using messaging in MyChart] I get straight to the patient. So it's a lot quicker.

Electronic Record of Communication

In addition to facilitating communication, patients also discussed the benefits of having an electronic record of exchanges with their provider. A patient told us this:

> It's just I can go in and access the message. I have a written copy, too, of what was said which, again to me getting older, is enormously important for me to have something I can go back to and go, 'Now, what did he say about that?'

Another patient described having this electronic record in a similar manner, as MyChart was perceived to help focus the office visit:

> I think it helps us more to focus on things. I can come in and say, 'Oh hey doc, I saw your note.' So when I am in the office, we already kind of got an idea of what is going on most of the time. And when I am out of the office, through MyChart, I can actually keep up on things. I just feel like the doctor knows better what is going on with me, and is able to respond to my situation quicker.

Provider Preferences in the Use of Secure Messaging

At the same time, we noted variations in the expectations and attitudes of providers toward using secure messaging and in the ways they address this topic with patients. For instance, when asked how they determined whether a patient needed an office visit or not, some providers offered clear guidance while others were more equivocal. A provider described how having the portal was a helpful way to remotely manage a patient's chronic conditions via secure messages:

> I will say, 'I want you to check your blood pressure, once a week for the next month, email me the results and then we can decide what we need to do from there.' Whereas before, I would have had to have them come back and show me their results on paper.

However, another noted that this benefit would depend on a variety of factors:

> Oh geez...it depends probably on the complexity of the problem. There are some problems that I would say, even though it's a new problem, a problem that I've never seen, sending a MyChart message to me seems totally appropriate.

While providers reportedly appreciated this possibility, they noted that more information was needed about portal use. As one summarized it:

> I think, that it would be good to have a little more education.

Perceived Concerns About Secure Messaging

Patients' Concerns About Secure Messaging

Three subthemes emerged involving patients' concerns about the secure messaging feature in the patient portal: (1) concern about imposing on the physician, (2) concern about lack of compensation for the provider, and (3) confusion about when to use the feature. Each of these sub-themes is further explained below, with additional evidence supporting these findings presented in Table 1.

Imposing on Providers' Time

Some patients were concerned that they would be taking up too much of their provider's time if they sent messages via the portal instead of going to the office to meet in person. A patient explained it as follows:

> I try to make sure that I only use it for important things. Or things that I know they want to know about. Well, like when I contact the doctor about getting labs before I come in, that is a useful thing. But, I am not going to contact one of my specialists in the middle, or 6 months away from an appointment just to say, hey I have this little itch or something.

Patients were also reportedly uncertain about how much messaging was too much, noting that they did not want to be a nuisance or a bother. A patient remarked:

> ...my biggest fear is that I don't want to get to a point where I am annoying the doctor and sending him three messages every day or something.

Another patient had similar thoughts:

> But I try not to interrupt. She's got a life...and this is a new thing for me. I don't want to be a nuisance.

Uncompensated Provider Time

Patients also reported concern about the fact that messaging a provider via the portal could result in uncompensated time for the provider. For example, one patient stated:

> So yeah, there have been times when I might have gone up for an appointment and I got enough answers through MyChart that I did not. So yeah, in one sense that's good for me that it prevented a trip, you know. For the business of medicine, I don't know.

Another patient similarly acknowledged the lack of provider reimbursement for interactions on MyChart:

> ...otherwise I would've had to go in and this is a business after all.

Lack of Clarity About When to Send a Secure Message

Patients in our study also noted that they were often uncertain about when it is appropriate to use the messaging feature to communicate with their physician. While most recognized that emergency situations were inappropriate, there was considerable lack of clarity as to what to do in non-emergent situations. As one patient described their thoughts:

> If everything is stable, I could probably go three months without using it. It's more when something is stirred up, which is, as I get older, that happens more frequently. And, you know, it's just a transitional time of life when, 'I don't even know if that's normal or not. Should I come in for that or am I wasting your time?'

Another patient echoed this sentiment, noting:

> That is the hard part.

Providers' Concerns About Secure Messaging

Three subthemes also emerged involving providers' concerns about the secure messaging feature: (1) concern about unfocused or insufficient information in the messages, (2) concern about inappropriate message topics, and (3) concern about incorrect use of the secure messaging feature. Here, we describe these sub-themes in greater detail, with additional supporting evidence provided in Table 1.

Unfocused or Insufficient Information in Messages

Most frequently, providers noted that patient messages did not contain sufficient information upon which they could make a recommendation, despite the messages sometimes being quite lengthy. A provider gave us an example of this lack of clarity:

> I may get 10 to 15 messages constantly in 2-3 hours from the same patient. 'Okay...I am feeling fatigued for 2 weeks.' So you know, that is not enough information for me. So I ask, 'Okay, do you have any other symptoms or do you want to see me?' And in the end you are lost, because you need to see the patient.

Table 1. Patient and provider concerns about secure messaging.

Concerns	Representative verbatim comment
Patient concerns	
Imposing on provider's time	"Try to keep it to the important stuff and if I need to be seen, then make an appointment, at least that is what I am trying to do."
	"I mean, I try to use...leave my physicians alone because, you know, I know that they have, you know, their number one priority is to take care of patients that are in the office."
	"I didn't want to be a pain in the arse to all the doctors by, you know, trying to ask them so many questions."
Uncompensated provider time	"And you know, sometimes I think, well I feel bad that I don't go in and give him his due for his time. But you know, this only took a second or two."
	"It was just that he would take the time to read it and respond without like coming in and paying for an appointment just increased my trust, I guess, that when a lot of things these days seem to be for the money, he had my well-being in mind."
Uncertainty about when to use the portal	"Yeah and it's like I say it's at his convenience for that. So he's not rushed, and I'm not taking away from anything."
	"Yeah, I don't know if I should be using it for that purpose, I don't know how much of his time I should take up."
Provider concerns	
Unfocused and/or insufficient information	"I mean I have had people, I can think of one in particular. A guy sent in about a 4-paragraph message, detailing numerous complaints, I'm not sure what he expected, but my answer was like, 'This is much too complicated, you have to come to the office.'"
	"So, to get valid information from patients, over the Internet probably requires a little bit more education than a lot of our patients have. Because if you can't accurately describe symptoms, then you can't accurately describe what you are doing, then it is going to be really hard to manage this appropriately. It is really hard to manage things appropriate regardless, but over MyChart, the degree of difficulty just increases."
	"Well that again, some of my people they'll go on and on. I have another colleague whose patient will go on and on even more than mine. And when it gets to a point you probably need to have a conversation back and forth, you probably need a face to face conversation, I try to set up an appointment."
Inappropriate topic	"Yeah one of the big pitfalls of MyChart messaging is the chest pain message. So, I have had people message, 'I have been having left side chest pain radiating to my arm, I get short of breath, what should I do?' So, these messages, we are not sitting by the computer waiting for the message to come in. I saw her message 4 hours later, I just happened to be going on, because I was on-call on a Saturday. And then I had to call first thing, didn't answer, so it created a big crisis really. But it ended up that she was okay. And I had to get her son to go to her house, and he ended up taking her to the ER, and everything turned out fine. But at the time we didn't know."
	"They want to give you this litany follow up, of what has been happening at home, you know, like you are email buddies. I don't like being any patient's email buddy."
Incorrect use of message feature	"I guess I don't like when it is used incorrectly, emergencies, for clerical issues, things that should be handled by another staff member that doesn't need to go directly to me. More and more the message comes to me and no one deals with it or answers the question. The patient just feels empowered to say, 'Hey I need to schedule an appointment,' It took me like two minutes to open it up figure it out and send a message, close it give it to someone else."
	"'Can you check on my prescription for something,' and normally a nurse would be able to do that without me even knowing about it or getting involved. But now I have to get involved. I have to do it all."
	"The patients can make appointments but, they often don't click on the right button so those come to us."

Another provider reflected, *"they will write paragraphs."* Even with long messages, however, providers were concerned about the quality of the information provided. As one provider noted, long descriptions without a clear question were of concern:

> *...writing pages and paragraphs, to give you the history of their problem. The history should come in a visit, not a question. That is not a question.*

Inappropriate Message Topic

Providers were also concerned that patients would send them messages via MyChart that were inappropriate for that mode. For instance, one provider explained how a patient would add detail that was not about the patient himself or herself:

> *I don't like it when patients, like a family member will send, for example a mother will say, 'Johnny got a fever today,' and she sends it on her chart. And that happens a lot. And it sort of contaminates her chart.*

And now we have information, confidential information, cause it can get like 'Well, you know my husband, you know his diabetes is worse now, and blah blah blah,' and now it is on the wife's chart. So, now Johnny Smith's diabetes information is on Susie Smith's chart. And for me that is like a confidentiality breach.

Similarly, several providers we interviewed felt patients treated messages as informal, friendly communications. A provider explained this with an example:

...like my patients, they send me a picture from India. Like 'Hi, we are having fun from India, just wanted to say hello…' This is not a public email. It's nice to chat, but that is not the purpose of MyChart.

Incorrect Use of Message Feature

Another area of concern raised by providers was incorrect use of the MyChart secure messaging feature. For instance, several providers complained that patients would use the secure messaging feature directly to request an appointment, rather than the "schedule an appointment" button. A provider explained this:

A patient says, 'I want to see you for an appointment. Please schedule me,' and stuff like that. I don't do scheduling.

Another incorrect use of the secure messaging feature emerged in the context of requesting refills. As one provider explained:

...people send refills on MyChart, and I don't mean the refill mechanism, but they message me with a refill.

Providers' Suggestions to Improve Patient-Portal Use

From providers' suggestions on how to improve use of the secure messaging feature in the patient portal, an important theme emerged. Taken together, these comments suggested an important opportunity to clarify the "rules of engagement" for a patient portal. We identified three sub-themes in this area, related to how patient-portal use could be improved by providing guidance on these "rules" as well as how the feature could be enhanced to reinforce the "rules": (1) offer patient training on appropriate portal use, (2) make patients accountable for learning how to use the portal, and (3) enhance the secure messaging feature to reinforce the "rules." We describe these sub-themes below and present additional supportive quotes in Table 2.

Offer Patient Training on Appropriate Portal Use

To address provider concerns about how patients use the portal, some providers suggested developing instructions or training for patients focused on how to use MyChart appropriately to communicate efficiently and effectively with their providers. Providers noted that this training would need to address issues beyond the technical aspects of how to navigate within MyChart and suggested the opportunity to emphasize the "rules of engagement" with a patient portal. For instance, this content would need to provide directions on how to communicate via the portal, including when to use secure messages versus when to call or schedule an appointment. A provider summarized it:

So, to get valid information from patients over the Internet probably requires a little bit more education than a lot of our patients have. Because if you can't accurately describe symptoms, then you can't accurately describe what you are doing, then it is going to be really hard to manage this appropriately.

Similarly, another provider suggested:

...to make sure the communication is more effective and more productive is something that probably could be trained.

Table 2. Opportunities to clarify "rules of engagement" and improve patient-portal use.

Providers' suggestions	Representative verbatim comment
Offer patient training on appropriate patient-portal use	"One thing that I think might be helpful is to have like almost guidelines for the patient, of what kinds of things are appropriate for MyChart and what kind of things aren't. So, you know, this is not to discuss new problems or symptoms you are having. That needs to be an office visit. It is to follow up, for quick questions. That kind of thing."
	"...with training patients and probably providers to some extent too, on how to use it appropriately and transmit the appropriate information."
Make patients accountable for learning how to use the patient portal	"When they sign up, if we have it written on paper or something like that, that we can hand them and say, 'Please review these guidelines.' Maybe have them initial off that they have read them."
	"Electronically, like have a course. They can take a course, like very brief course. And sign an agreement. And after they sign the agreement, and they understand the application of MyChart, then they would be allowed to sign in for MyChart."
Enhance secure messaging feature to reinforce "rules"	"I think that when people send refills on MyChart, and I don't mean the refill mechanism, but they message me with a refill. So it might be good if there was a pop up saying, 'There is another way to do refills,' 'There is another way for emergencies, which is to call on-call,' 'There is another way to if it is not about you you go to that person's MyChart.' So it might be good to have some kind of pop-up, just so they stop and read. It could probably save a lot of nonsense messages."
	"I think it would be great if it could be filtered, through some system or people. Or some messages need to go to the desk, scheduling person, somewhere. It should go directly to them rather than coming to me and I have to answer and route it to them."

XSL•FO

RenderX

Make Patients Accountable for Learning How to Use the Patient Portal

Providers also noted the importance of making patients responsible for learning how to properly use the portal. These providers suggested that there might be different opportunities to provide the training, including at portal sign up or during a visit, but emphasized that patients should be held accountable for this learning. More than one provider suggested the drafting of a document that patients would be asked to sign, acknowledging receipt of this education, and noted they would then be able to refer to the document later when discussing appropriate messaging during future visits. A provider proposed this:

> *The patient can read the agreement, and you know click on it. And then, you know, you can go to the patient and they can sign up for MyChart. And we have a document saying, listen you have read this and you cannot use it like that.*

Enhance Secure Messaging Feature

Providers proposed several opportunities to enhance the MyChart application functionality in ways that could automatically provide guidance to patients within the secure messaging feature. Of these, one opportunity was around providing information about the urgency of the message. A provider suggested the following solution:

> *I think that when they open it up to send a message, it should say like hang on a minute, are you complaining about an emergency situation? It is like when they call our office and the message says if it is an emergency, call 9-1-1. And maybe there needs to be something, a pop-up, saying, 'Are you sure that this is the appropriate medium?'*

Another enhancement proposed was that secure messages could be limited to a certain number of characters. A provider told us this:

> *The university has a policy that, for any message, you need to limit it to so many characters. And when they get too much characters, the university says—sends them a little note saying, 'Sorry, but with the use of this, we need to limit the amount of information in this due to your physician's need to address all his patient's concerns.'*

Providers similarly commented about opportunities to provide direction to patients about the appropriateness of message content around refills and appointment scheduling, suggesting that pop-up messages or other portal enhancements might work.

Discussion

Overall Findings

Our study suggests that initial concerns about overuse and security of information expressed by patients and providers in pre-implementation studies [19,23-26] may no longer apply as users gain experience. Instead, experienced users identified concerns beyond the technical aspects of using a portal. Patients worried about imposing on a provider's time, uncompensated

provider time, and a lack of clarity about when to send a secure message. Providers did not discuss an increased workload as has been noted in pre-implementation studies [20,23-26]; instead, they were concerned about unfocused and/or insufficient information in messages, inappropriate message topics, and incorrect use of the message feature. In discussing these concerns, providers suggested a need for further training focused on these issues.

The portal used in this setting provides patients with instructions, described in the Study Setting above, about when to use a secure message to set patient expectations about response times and provides some guidance on whether to send a message or call 911. However, patients we interviewed expressed confusion about how to define non-urgent concerns, and providers noted that some patients still included information in their messages that was inappropriate for their medical record.

Unlike other portal features such as scheduling appointments or requesting prescription refills, secure messaging requires interaction with another individual and therefore users need to understand more than simply the technical aspects of how to access a feature. Appropriate use requires an understanding of the type of information that should be conveyed via the portal and the etiquette rules of electronic communication. Yet, little guidance is provided to patients or providers related to the "rules of engagement" in secure messaging.

Practice Recommendations

Our findings suggest that information and training on the "rules of engagement" is needed on several levels. For patients, print materials and instructional videos can be presented as they begin to use a portal. Such materials can provide patients with information about creating an account and navigating through the portal's features. However, additional training and information related to how to engage and communicate via a portal may be required to improve communication for both patients and providers, particularly for experienced users such as those we interviewed.

Patient-focused information could be developed to set the tone for the "rules of engagement" and address issues such as when secure messaging is appropriate, question topics that can be addressed via secure messaging, what type of information to include in the messages, and how to understand information sent by the provider. Additionally, such material represents another opportunity to address patient safety by reminding patients that their provider may not see the message immediately, and, therefore, secure messaging should not be used for emergency situations. This information would thus provide patients with guidance on how to engage with, and not just how to navigate, the portal, thereby potentially alleviating patient concerns related to perceived burden as well as facilitating more efficient communication within the portal.

Providers could also benefit from clarifying the "rules of engagement" from their perspective. Currently, providers may receive training on the aspects of the patient portal that face them as providers, such as how to view and send a secure message. Additional training that exposes providers to the patient view of the portal may provide a more complete

understanding of the patient experience and help them to better interact with their patients. In addition, guidance could be provided on how to communicate in secure messages or alongside lab and test results. Past studies of patient-provider communication have focused mainly on in-person communication, with electronic communication studied primarily to document trends in use [11-14]. Therefore, providers, like patients, typically have little guidance on the language they could use in portal communication or how to structure such communications. In addition, unlike in a face-to-face encounter, electronic communications make it difficult for providers to assess patient comprehension. Training providers to send better messages may increase the quality of patient-provider communication and reduce the need for additional clarifying messages. Topics this training could address include communicating positive and negative results, communicating at the appropriate level of health literacy, and providing educational materials to facilitate patient understanding.

At the same time, providers also need to establish clear and consistent guidelines of the expectations they have for patients in communicating via a patient portal. Before the patient portal was implemented, patients would call their provider's office with questions. While this process had its own inefficiencies, such as waiting time on the telephone or leaving and returning phone messages, information was most often filtered through office staff who had general knowledge about the information a physician would need to respond to that particular question. Communication via a patient portal, however, lacks such a filter to focus patient questions and the information they convey. In addition, secure messaging is asynchronous and therefore may lack the conversational nature of an in-person visit in which information can be exchanged and clarified quickly. Further, our study demonstrates that even patients experienced in patient portal use lack clarity on when to use a secure message and what information to include. Similarly, while some providers in our study mentioned preparing patients to receive lab or test results via the patient portal, none discussed communication expectations with patients. In our study, we note that these expectations may vary by individual provider, suggesting that discussions about portal use may help to improve the efficiency of patient-provider communication and alleviate patient concerns about being a burden to their providers.

In practice, portal technology could leverage electronic communication capabilities by incorporating features such as built-in guidance. For example, as physician interviewees suggested, including a link on the secure messaging screen to guide patients in determining whether their concern meets the criteria for being "non-urgent" could be helpful. Furthermore, developing structured message boxes to guide patients to complete the information providers need to address patient concerns may not only help ensure that necessary information is conveyed, but also help patients focus their messages and more clearly describe their concerns.

As patients, providers, and health care systems gain greater experience with patient portals, new needs emerge to define the "rules of engagement" through a portal. While there are a range of technical solutions that could be implemented to improve patient and provider communication via secure messaging, it is important to elicit input from all stakeholders in designing these modifications. The patients in our study, who were experienced users, had clear thoughts on what they liked about the secure messaging and identified specific areas in which they were uncertain about how to use this tool. Discussions with patients can help to further refine their concerns and develop new ways to address them. As noted above, for the most part, providers in our study did not express the concerns noted in the literature in pre-implementation studies, specifically related to the increased workload of secure messaging. However, they identified areas in which the process of secure messaging could be improved. Further work is needed to develop stakeholder-driven solutions to these issues. While our study did not include healthcare system administrators, they play a significant role in encouraging the use of patient portals in general and secure messaging in particular. Their goals for secure messaging could also be important in shaping the next round of education and training to clarify the "rules of engagement."

Limitations

We note the inclusion of only one health system as a limitation of our study. Although the features of the patient portal used by this health system are common to those used across the country, the experiences of interviewees in our study are limited to how the portal has been implemented and used in this health system. While we reached saturation on the topics covered in our interviews, patients and providers in other health systems or using other patient portals may have different perspectives. Additionally, as is typical in qualitative studies, we did not collect demographic data from the interviewees. Differing perspectives by demographic characteristics may be explored in future studies.

Conclusions

As patients and providers gain more experience with patient portals, the needs and perspectives of both groups regarding portals are evolving. Many patients are now beyond the "new user" phase and are realizing the benefits of more comprehensive portal use. Communication through portals is increasingly viewed as an extension of care between visits. While we can expect that this will result in better management of patient conditions, our study demonstrates new concerns that arise with greater use. Patients struggle to balance their desire to respect their provider's time with their need for answers to health-related questions. Providers are still figuring out how best to communicate with patients via portals in a way that addresses patient needs without overstepping boundaries. These findings suggest that additional information and training on the "rules of engagement" may help address the concerns of both patients and providers and improve the efficiency of communication via patient portals.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Wald JS, Middleton B, Bloom A, Walmsley D, Gleason M, Nelson E, et al. A patient-controlled journal for an electronic medical record: issues and challenges. Stud Health Technol Inform 2004;107(Pt 2):1166-1170. [Medline: 15360996]

2. Nelson JA. Personal health records. Home Health Care Manag Pract 2009;21(2):141-142.

3. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 05;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]

4. Murphy J. Engaging patients and families in ehealth. Nurs Econ 2011;29(6):339-341. [Medline: 22360112]

5. Wilson ML, Murphy LS, Newhouse RP. Patients' access to their health information: a meaningful-use mandate. J Nurs Adm 2012 Nov;42(11):493-496. [doi: 10.1097/NNA.0b013e3182714538] [Medline: 23099997]

6. Hess R, Bryce CL, Paone S, Fischer G, McTigue KM, Olshansky E, et al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. Telemed J E Health 2007 Oct;13(5):509-517. [doi: 10.1089/tmj.2006.0089] [Medline: 17999613]

7. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]

8. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008;15(1):1-7 [FREE Full text] [doi: 10.1197/jamia.M2562] [Medline: 17947615]

9. Kaelber D, Pan EC. The value of personal health record (PHR) systems. In: AMIA Annual Symposium proceedings. 2008 Presented at: AMIA Symposium; 2008; Washington, DC p. 343-347.

10. Emont S. CHCF. 2011 May. Measuring the impact of patient portals: what the literature tells us URL: http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/PDF%20M/PDF%20MeasuringImpactPatientPortals.pdf [accessed 2017-06-09] [WebCite Cache ID 6r5XnUnoD]

11. Cronin RM, Davis SE, Shenson JA, Chen Q, Rosenbloom ST, Jackson GP. Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. Appl Clin Inform 2015 Apr 29;6(2):288-304 [FREE Full text] [doi: 10.4338/ACI-2014-12-RA-0117] [Medline: 26171076]

12. Shenson JA, Cronin RM, Davis SE, Chen Q, Jackson GP. Rapid growth in surgeons' use of secure messaging in a patient portal. Surg Endosc 2016 Apr;30(4):1432-1440 [FREE Full text] [doi: 10.1007/s00464-015-4347-y] [Medline: 26123340]

13. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 01;20(3):519-525. [Medline: 23242764]

14. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014 Mar 06;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]

15. 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 26;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]

16. Britto MT, Hesse EA, Kamdar OJ, Munafo JK. Parents' perceptions of a patient portal for managing their child's chronic illness. J Pediatr 2013 Jul;163(1):e1-e2. [Medline: 23541773]

17. 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. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Apr Presented at: CHI '15; 2015; Seoul, Republic of Korea p. 3859-3868.

18. 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 Dec 03;17(12):e275. [Medline: 26681155]

19. Clark SJ, Costello LE, Gebremariam A, Dombkowski KJ. A national survey of parent perspectives on use of patient portals for their children's health care. Appl Clin Inform 2015 Feb 25;6(1):110-119 [FREE Full text] [doi: 10.4338/ACI-2014-10-RA-0098] [Medline: 25848417]

20. 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 02;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]

21. Haggstrom DA, Saleem JJ, Russ AL, Jones J, Russell SA, Chumbler NR. Lessons learned from usability testing of the VA's personal health record. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i13-i17 [FREE Full text] [doi: 10.1136/amiajnl-2010-000082] [Medline: 21984604]

22. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]

23. 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]

24. Garrido T, Meng D, Wang JJ, Palen TE, Kanter MH. Secure e-mailing between physicians and patients: transformational change in ambulatory care. J Ambul Care Manage 2014;37(3):211-218 [FREE Full text] [doi: 10.1097/JAC.0000000000000043] [Medline: 24887522]

25. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013 Jul 03;15(7):e133. [Medline: 23823974]

26. Wynia MK, Torres GW, Lemieux J. Many physicians are willing to use patients' electronic personal health records, but doctors differ by location, gender, and practice. Health Aff (Millwood) 2011 Feb;30(2):266-273 [FREE Full text] [doi: 10.1377/hlthaff.2010.0342] [Medline: 21289348]

27. 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]

28. 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]

29. Vydra TP, Cuaresma E, Kretovics M, Bose-Brill S. Diffusion and use of tethered personal health records in primary care. Perspect Health Inf Manag 2015 Apr 01;12:1c [FREE Full text] [Medline: 26755897]

30. Fiks AG, Mayne S, Karavite DJ, DeBartolo E, Grundmeier RW. A shared e-decision support portal for pediatric asthma. J Ambul Care Manage 2014;37(2):120-126 [FREE Full text] [doi: 10.1097/JAC.0000000000000025] [Medline: 24594560]

31. Gaskin GL, Longhurst CA, Slayton R, Das AK. Sociotechnical challenges of developing an interoperable personal health record: lessons learned. Appl Clin Inform 2011;2(4):406-419 [FREE Full text] [doi: 10.4338/ACI-2011-06-RA-0035] [Medline: 22003373]

32. Savoy M, Hammond JA, Castellano L. Implementing a patient portal at the family medicine center. Del Med J 2015 Mar;87(3):81-84. [Medline: 25985541]

33. Siteman E, Businger A, Gandhi T, Grant R, Poon E, Schnipper J, et al. Clinicians recognize value of patient review of their electronic health record data. In: AMIA Annu Symp Proc. 2006 Presented at: AMIA Annual Symposium proceedings; 2006; Washington, DC p. 1101 URL: http://europepmc.org/abstract/MED/17238720

34. Brennan PF, Burke L, Casper G, Sebern M, Krause C, Kossman S, et al. Creating technology-enhanced practice: a university-home care-corporate alliance. Stud Health Technol Inform 2006;122:644-648. [Medline: 17102341]

35. 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 05;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]

36. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]

37. Kummerow Broman K, Oyefule OO, Phillips SE, Baucom RB, Holzman MD, Sharp KW, et al. Postoperative care using a secure online patient portal: changing the (inter)face of general surgery. J Am Coll Surg 2015 Dec;221(6):1057-1066 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2015.08.429] [Medline: 26453260]

38. Reti SR, Feldman HJ, Ross SE, Safran C. Improving personal health records for patient-centered care. J Am Med Inform Assoc 2010;17(2):192-195 [FREE Full text] [doi: 10.1136/jamia.2009.000927] [Medline: 20190063]

39. Chunchu K, Mauksch L, Charles C, Ross V, Pauwels J. A patient centered care plan in the EHR: improving collaboration and engagement. Fam Syst Health 2012 Sep;30(3):199-209. [doi: 10.1037/a0029100] [Medline: 22866953]

40. Kerns JW, Krist AH, Longo DR, Kuzel AJ, Woolf SH. How patients want to engage with their personal health record: a qualitative study. BMJ Open 2013;3(7):e002931.

41. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013 Aug 26;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]

42. Black H, Gonzalez R, Priolo C, Schapira MM, Sonnad SS, Hanson 3rd CW, et al. True "meaningful use": technology meets both patient and provider needs. Am J Manag Care 2015 May 01;21(5):e329-e337 [FREE Full text] [Medline: 26167781]

43. Ross SE, Nowels CT, Haverhals LM, Bull SS, Lin CT, Main DS. Qualitative assessment of Diabetes-STAR: a patient portal with disease management functions. In: AMIA Annual Symposium proceedings. 2007 Oct Presented at: AMIA Symposium; 2007; Chicago, IL.

44.   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. [Medline: 22071530]

45.   Atreja A, Mehta N, Miller D, Moore S, Nichols K, Miller H, et al. One size does not fit all: using qualitative methods to inform the development of an Internet portal for multiple sclerosis patients. In: AMIA Annual Symposium proceedings. 2005 Presented at: AMIA Symposium; 2005; Washington, DC p. 16-20 URL: http://europepmc.org/abstract/MED/16778993

46.   Varsi C, Gammon D, Wibe T, Ruland CM. Patients' reported reasons for non-use of an internet-based patient-provider communication service: qualitative interview study. J Med Internet Res 2013 Nov 11;15(11):e246 [FREE Full text] [doi: 10.2196/jmir.2683] [Medline: 24220233]

47.   Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. London: Weidenfield & Nicolson; 1968.

48.   Constas MA. Qualitative analysis as a public event: The documentation of category development procedures. Am Educ Res J 1992;29(2):253-266.

Abbreviations

AMC: Academic Medical Center
HER: electronic health record
PI: principal investigator

Original Paper

Impact of At-Home Telemonitoring on Health Services Expenditure and Hospital Admissions in Patients With Chronic Conditions: Before and After Control Intervention Analysis

Branko Celler[1], PhD; Marlien Varnfield[2], PhD; Surya Nepal[3], PhD; Ross Sparks[4], PhD; Jane Li[5], MD, PhD; Rajiv Jayasena[6], PhD

[1]Biomedical Systems Research Laboratory, University of New South Wales, Sydney, NSW, Australia

[2]Health and Biosecurity Business Unit, eHealth Research Program, Commonwealth Scientific and Industrial Research Organisation, Herston QLD, Australia

[3]Data 61, Software and Computational Systems Program, Commonwealth Scientific and Industrial Research Organisation, Marsfield, NSW, Australia

[4]Health and Biosecurity Business Unit, eHealth Research Program, Commonwealth Scientific and Industrial Research Organisation, North Ryde, NSW, Australia

[5]Health and Biosecurity Business Unit, eHealth Research Unit, Commonwealth Scientific and Industrial Research Organisation, Marsfield, NSW, Australia

[6]Health and Biosecurity Business Unit, eHealth Research Program, Commonwealth Scientific and Industrial Research Organisation, Parkville, VIC, Australia

Corresponding Author:
Branko Celler, PhD
Biomedical Systems Research Laboratory
University of New South Wales
Building G7
Sydney, NSW, 2052
Australia
Phone: 61 2 9385 0746
Fax: 61 2 9385 5993
Email: b.celler@unsw.edu.au

Abstract

Background: Telemonitoring is becoming increasingly important for the management of patients with chronic conditions, especially in countries with large distances such as Australia. However, despite large national investments in health information technology, little policy work has been undertaken in Australia in deploying telehealth in the home as a solution to the increasing demands and costs of managing chronic disease.

Objective: The objective of this trial was to evaluate the impact of introducing at-home telemonitoring to patients living with chronic conditions on health care expenditure, number of admissions to hospital, and length of stay (LOS).

Methods: A before and after control intervention analysis model was adopted whereby at each location patients were selected from a list of eligible patients living with a range of chronic conditions. Each test patient was case matched with at least one control patient. Test patients were supplied with a telehealth vital signs monitor and were remotely managed by a trained clinical care coordinator, while control patients continued to receive usual care. A total of 100 test patients and 137 control patients were analyzed. Primary health care benefits provided to Australian patients were investigated for the trial cohort. Time series data were analyzed using linear regression and analysis of covariance for a period of 3 years before the intervention and 1 year after.

Results: There were no significant differences between test and control patients at baseline. Test patients were monitored for an average of 276 days with 75% of patients monitored for more than 6 months. Test patients 1 year after the start of their intervention showed a 46.3% reduction in rate of predicted medical expenditure, a 25.5% reduction in the rate of predicted pharmaceutical expenditure, a 53.2% reduction in the rate of predicted unscheduled admission to hospital, a 67.9% reduction in the predicted rate of LOS when admitted to hospital, and a reduction in mortality of between 41.3% and 44.5% relative to control patients. Control patients did not demonstrate any significant change in their predicted trajectory for any of the above variables.

Conclusions: At-home telemonitoring of chronically ill patients showed a statistically robust positive impact increasing over time on health care expenditure, number of admissions to hospital, and LOS as well as a reduction in mortality.

XSL•FO
RenderX

Introduction

In industrialized nations, approximately 70% to 78% of health care budgets are spent on the management of chronic disease or its exacerbation [1]. As the population ages the burden of chronic disease will increase and place health care budgets under increasing strain [2,3]. Telehealth services, with at-home telemonitoring of vital signs, have been demonstrated to deliver cost effective, timely, and improved access to quality care [4-7]. These services also reduce social dislocation and enhance the quality of life by allowing chronically ill and aged members to stay in their homes and communities longer [6-7].

One of the largest trials for evaluating telehealth outcomes was the Whole System Demonstrator (WSD) trial in the United Kingdom [8-10]. Results from this trial have been extensively reported and showed reductions in hospital admissions, bed days, costs, and mortality. However, experience in Australia with the deployment of at-home telemonitoring services is very limited [11-12]. Most trials are small scale and lack detailed analysis of key success factors such as health care outcomes, health economic benefits, impact on clinical work force availability, and acceptability by patients, carers, nurses, primary care physicians (PCP), and health care managers as well as the effect of workplace culture and capacity for organizational change management [13].

Despite large national investments in health information technology, very little policy work has been undertaken in Australia in deploying at-home telemonitoring as a solution to the increasing demands and costs of managing chronic disease.

This trial was designed to develop a robust evidence base for a number of key factors and demonstrate an effective and scalable model for Internet-enabled at-home telemonitoring services in Australia. Armed with the insights provided by this evidence base, policy makers may have much of the data they require to implement funding models and create a sustainable telehealth services sector in Australia.

Methods

Research Ethics Committee Approval

The clinical trial protocol for this study was approved by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Human Research Ethics Committee (HREC) (Approval Number 13/04, March 25, 2013) as well as 5 other local HRECs. A journal article on the clinical trial design has been previously published [13].

Patient Selection and Recruitment

A before and after control intervention (BACI) design was used where control patients were matched to each test patient. This design [13-16] is well known in environmental intervention studies but is less known for health interventions. However, it has theoretical justifications for studies involving heterogeneous populations and has been successfully applied in many environmental intervention studies [14-16].

Candidates were eligible to participate in the study if they met inclusion criteria which were comprehensively described in an earlier publication [13] but are mentioned here briefly for convenience: age 50 years and older; 2 or more unplanned acute admissions during the last 12 months or 4 or more unplanned acute admissions during the previous 5 years, with a principal diagnosis of chronic obstructive pulmonary disease, coronary artery disease, hypertensive diseases, congestive heart failure, diabetes, or asthma. Eligible patients could be under the care of a community nurse or PCP or participants in a government care program other than special targeted programs to support individuals with high-care needs. Patients were also excluded if they were diagnosed with compromised cognitive function [17], a neuromuscular disease, or a psychiatric condition.

For each test participant, as many as 4 control candidates were automatically case matched [13] on gender, age, chronic condition, and socioeconomic indexes for areas (SEIFA) [18]. On their consent, the 2 closest matching control candidates commenced as participants in the study. We noted that in many cases only 1 acceptable match was available. When a test patient had more than 1 control, the data for the matched control patients were averaged to obtain a single matched pair. Both test patients and control patients continued to receive normal care under the management of their PCP.

Figure 1 shows the recruitment process and flow of participants through the study. A total of 1429 eligible patients were identified from hospital lists provided by local health districts and patients known to clinical staff. From these, 479 were still deemed eligible following individual screening and were contactable.

Following exclusions, a master list of 114 test patients and 173 control patients, all with pharmaceutical benefit scheme (PBS) and medical benefit scheme (MBS) data, was formed. On careful analysis of these data made available from the Australian Government Department of Human Services (DHS), it was observed that some patients had missing data. As a result, data from a number of test patients and control patients were rejected from further analysis. This led to a final matched cohort of 100 test patients and 137 control patients.

Hospital data were intended to be sourced for all test and control patients selected from hospital lists at each of the 5 test sites. However, as some test and control patients were not selected from hospital lists, their hospital data were thus not available for analysis. From the 100 test and 137 control patients matched for analysis of medical and pharmaceutical benefits data, 86 and 107, respectively, were matched for analysis of hospital admission and length of stay (LOS).

On detailed analysis of available patient hospital data, it was found that some patients had attended the emergency department

of their local hospital, in some cases more than once on the same day, without being admitted. As a result, we decided to count an admission as involving at least 1 overnight stay, and this led to the further rejection of 33 test patients and 43 control patients, who based on these criteria, had no admissions to hospital.

This resulted in a final cohort of 53 test and 64 control patients for which full historical hospital data were available.

Figure 1. Recruitment flow chart.
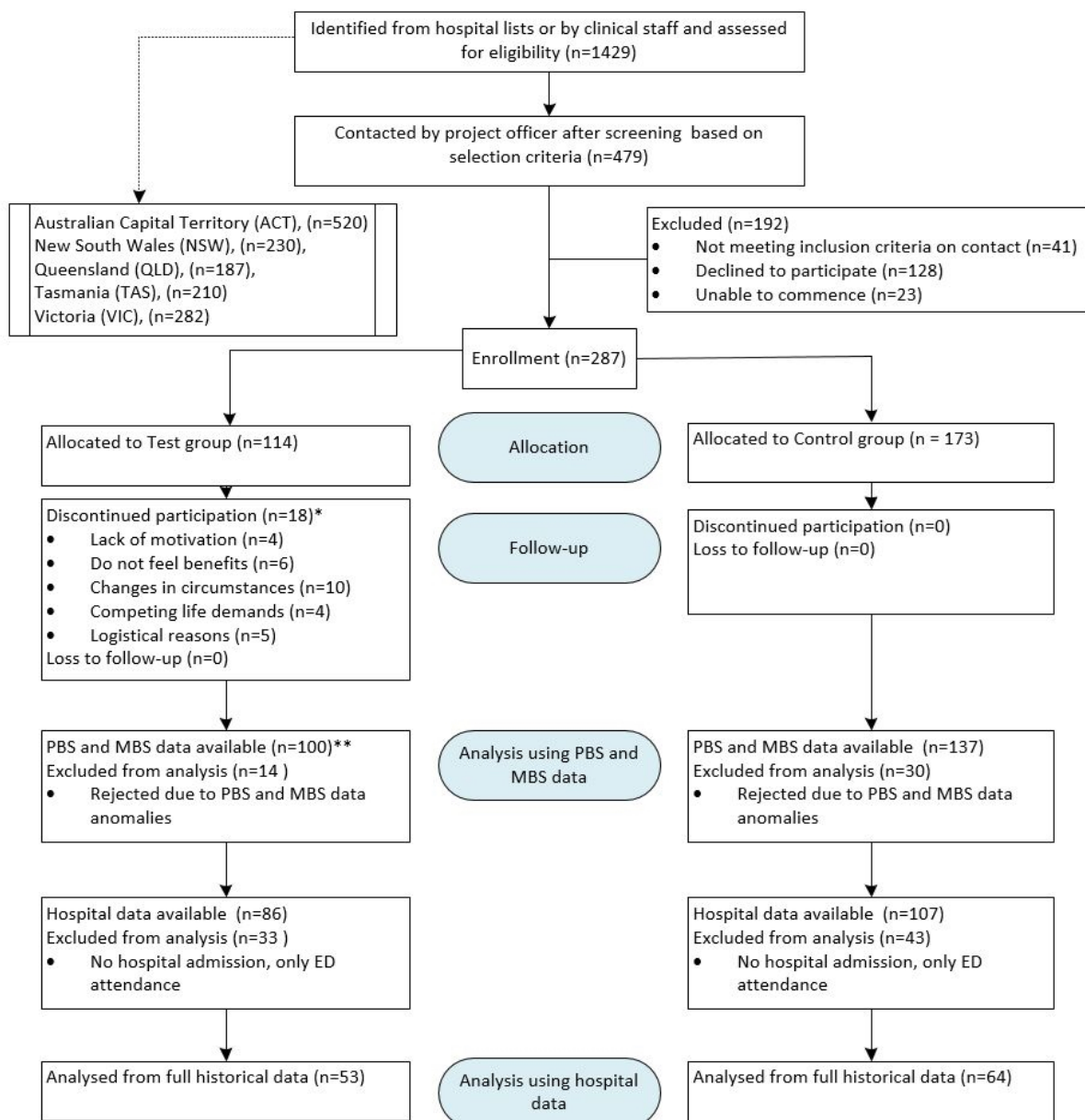
* The numbers reported are counts per reason as some participants stated more than one reason.
** Even though 18 discontinued, 4 patients had Pharmaceutical Benefits Scheme (PBS) and Medical Benefits Scheme (MBS) data available with consent.

XSL•FO
RenderX

Organization of Care

A project officer (PO) at each test site was responsible for managing operational and research activities for the study, thereby separating patient care from study operations. Test patients were supplied with the Telemedcare Telemonitoring Unit (TMU) by the PO who also trained them on its use [13]. The PO was also responsible for consenting patients, onsite visits, equipment maintenance, and technical support.

The clinical care coordinator (CCC) monitored patient vital signs and clinical questionnaire responses recorded via the Telemedcare TMU daily during business hours. The CCCs were experienced nursing staff, seconded part-time from each trial site health service provider. Their role was to coordinate the delivery of care when the telemonitoring of vital signs data and follow-up contact with patients indicated that they were experiencing an exacerbation of their condition. Normal care for the majority of test patients was by their local PCP.

Participants in the test group were provided with the Telemedcare TMU which was configured by the site PO or CCCs to reflect clinical best practice for the patient's clinical condition. Patients would be reminded to record their vital signs measurements (such as blood pressure, oxygen saturation, electrocardiogram, spirometry, temperature, weight, and blood glucose), scheduled at a convenient time, typically in the morning before taking their medications.

Control participants received care as usual according to the service model of the respective trial site. They had no further contact with the PO after the consent process.

Comparison of Medical and Pharmaceutical Expenditure Before the Start and Close to the End of the Trial

In order to compare the statistical match of test and control patients with respect to medical and pharmaceutical expenditure at the onset of telemonitoring, individual costs were summed over a period of 100 days just prior to the beginning of the intervention and in the last 100 days prior to the end of the intervention. The paired t test was then used to identify significant changes between test and control patients in both time periods.

Regression Modeling

Medical, pharmaceutical, and hospital data were all synchronized to the date when the telemonitoring commenced to average out seasonal effects. Medical and pharmaceutical cost data for every patient were summed over 36 30-day periods before the start of the trial and 12 30-day periods after the start of the trial. This approximates analyzing data over 3 years before and 1 year after the start of telemonitoring.

Hospital admissions and LOS data were similarly treated, except that the time interval chosen was 100 days. This was a preferred interval as hospital admissions were much less frequent and would otherwise generate data with a large number of zero entries.

All the outcome variables were expected to increase over time because all patients were chronically ill and aging. We fitted a linear model including the explanatory variables 30-day or 100-day time period number, before-after indicator variable, and the interaction between these two variables.

To carry this out, the outcome variables of all test and control patients were averaged within each time period number. Normality of data was tested in each outcome and where necessary, square root (sqrt) or LogNormal transforms were applied.

Before and after data were analyzed, both as separate time period numbers with different slopes or as 1 time period having the same slope. This analysis was applied to (1) test patient data, (2) control patient data, and (3) difference (control-test) data.

These time series analyses permitted the determination of how well test patients and control patients were indeed matched, controlled for possible effects of the intervention on control patients by also analyzing differences (control-test) and reduced possible seasonal and other time varying influences.

Sqrt transformation was applied to medical and pharmaceutical benefits data before linear regression analysis was carried out. This was repeated both for test patient data and control patient data. Difference data calculated from control-test values for each data point were found to be normally distributed and did not need the application of any transform.

The time course of before and after data was modeled using linear regression and analysis of covariance (ANCOVA) analysis of slopes to identify statistically significant differences in before and after slopes, using the differences (control-test) to test and validate the results.

To estimate savings in expenditure over the year following the start of intervention, sqrt (30-day costs) were converted to annual costs by multiplying each 30-day data point by 365/30 and each 100 day data point by 365/100.

As a result of sqrt normalization, the functions for medical and pharmaceutical costs before and after intervention become quadratic, and estimates of savings require the calculation of predicted costs 1 year after the start of intervention based on the projection of the 3-year historical trajectory, 1 year past the start of intervention. The total predicted medical benefits expenditure for the year following intervention was estimated from the area under the annual expenditure curve projected 1 year from the start of intervention.

Following intervention, we would expect the slope of the regression line to change, and the area of the curve beneath the actual expenditure curve then provides an estimate of the actual expenditure for that year. The difference in the 2 areas is an estimate of savings over the year.

Linear regression was carried out using the fit command in the MATLAB (The MathWorks Inc) statistics toolbox. Outliers were excluded from the linear regression. The command predObs was used to plot 95% prediction intervals. Prediction intervals indicate a 95% probability that a future observation at x will fall within its boundaries.

Standard goodness of fit measures, including the sum of squares due to error, the coefficient of determination (R^2), the R^2 value

adjusted for degree of freedom, and the standard error or root mean square error were also available. The control-test difference data were similarly analyzed.

Estimating Mortality

A master register (MR) file of 1429 patients was formed by combining the hospital records from each local health district in each state and territory. Deaths of patients in this master file were subsequently cross-checked against the records of the Births, Deaths, and Marriages Register (BDMR) in each state and territory.

To more accurately compare mortality between test and control groups, the effect of the population's age distribution must be taken into account. We thus use age-specific death rates (ASDRs), defined as the ratio of the number of deaths in a given age group to the population of that age group. For both methods, we compare actual mortality data against ASDRs calculated from the master register of eligible patients.

Statistical Analysis

Comparisons, using the cases available, were made between the 2 groups at baseline using the chi-square test or Fisher exact test for categorical variables, the 2-sample t test for continuous variables, and the Wilcoxon rank-sum test for skewed variables. Baseline characteristics for both test and control patients are described using mean and standard deviations (SD) for continuous symmetrical variables and medians and 95% confidence intervals (CI) for skewed data.

Categorical variables are presented as counts and percentages. Within matched group differences (matched control minus test data) from baseline to last point were examined using the paired t test for symmetrical data and the Wilcoxon signed-rank test for skewed data. All statistical tests were 2-tailed, and a P value of <.05 was used to indicate statistical significance. Statistical analysis was performed using Stata verison 12 (StataCorp LLC), SPSS version 17 (IBM Corp), Matlab R2015b (The MathWorks Inc), and Excel (Microsoft Corp).

A data integration engine [18] facilitated the generation of various graphs using structured query language queries and either built-in Excel graph functions or Visual Basic programming for more complex graphs.

Results

There were no significant differences in age between test (71.1 [SD 8.7] years; n=100) and control (71.7 [SD 9.0] years; n=137) patients or between male and female patients. There were also no statistical differences observed between test and control patients with respect to their SEIFA status or their primary disease diagnosis.

A total of 67% (67/100) of the test patients were male and 33% (33/100) were female. For control patients, 43.8% (60/137) were female and 56.2% (77/137) were male. Most patients had more than 1 condition listed as a primary diagnosis. For simplicity, primary disease conditions were grouped in the broad categories of cardiovascular disease (N_{Test}=50), respiratory disease (N_{Test}=30) and diabetes (N_{Test}=20) although some patients had multiple comorbidities.

Test patients were monitored on average for 276 days, with no significant difference between average monitoring durations for female patients (266 days) and male patients (281 days). A total of 75% (75/100) of all test patients were monitored for periods exceeding 6 months.

Table 1 shows that there were no significant differences between test and control patients in terms of baseline total cost of medical and pharmaceutical benefits items for 100 days immediately preceding the start of intervention. However, for the last 100 days prior to the end of intervention, there was a significant difference in medical and pharmaceutical expenditure, with control patients spending on average $3298 more per year than test patients.

Table 1. Baseline comparison between test patients and control patients over 100 days prior to intervention and last 100 days prior to end of intervention.

Variable	Control patients $ (95% CI)	Test patients $ (95% CI)	P value
Expenditure in last 100 days prior to start of intervention (N $_{Test}$ =100, N $_{Control}$ =137)			
Total cost of medications prescribed	975.8 (755-1205)	919.4 (748-1080)	.42
Total expenditure on medical and pharmaceutical items	1931.7 (1525-2339)	2044 (1648-2423)	.12
Expenditure in last 100 days prior to end of intervention (N $_{Test}$ =100, N $_{Control}$ =137)			
Total cost of medications prescribed	859.7 (615-1149)	505.9 (318-770)	<.001
Total expenditure on medical and pharmaceutical items	1941.7 (1366-2637)	1038.2 (656-1570)	<.001

Linear Regression Analysis

Figure 2 shows the medical costs averaged over 36 intervals of 30 days before and 12 intervals of 30 days after intervention. Figure 2 A and 2B show the sqrt($ medical expenditure) for test and control patients, Figure 2 C shows the linear difference in $ medical expenditure. Additional ANCOVA analysis comparing slopes for control patients of the combined before and after data as a single line to the before data alone, as shown in Figure 2 D, shows that there was no significant before and after difference (*P*=.929).

Figure 2. sqrt(MBS medical $ expenditure) plotted for (A) test patients and (B) control patients. Linear differences (control-test) are plotted in (C). Panel (D) shows change in regression line when before and after data are combined for control patients.
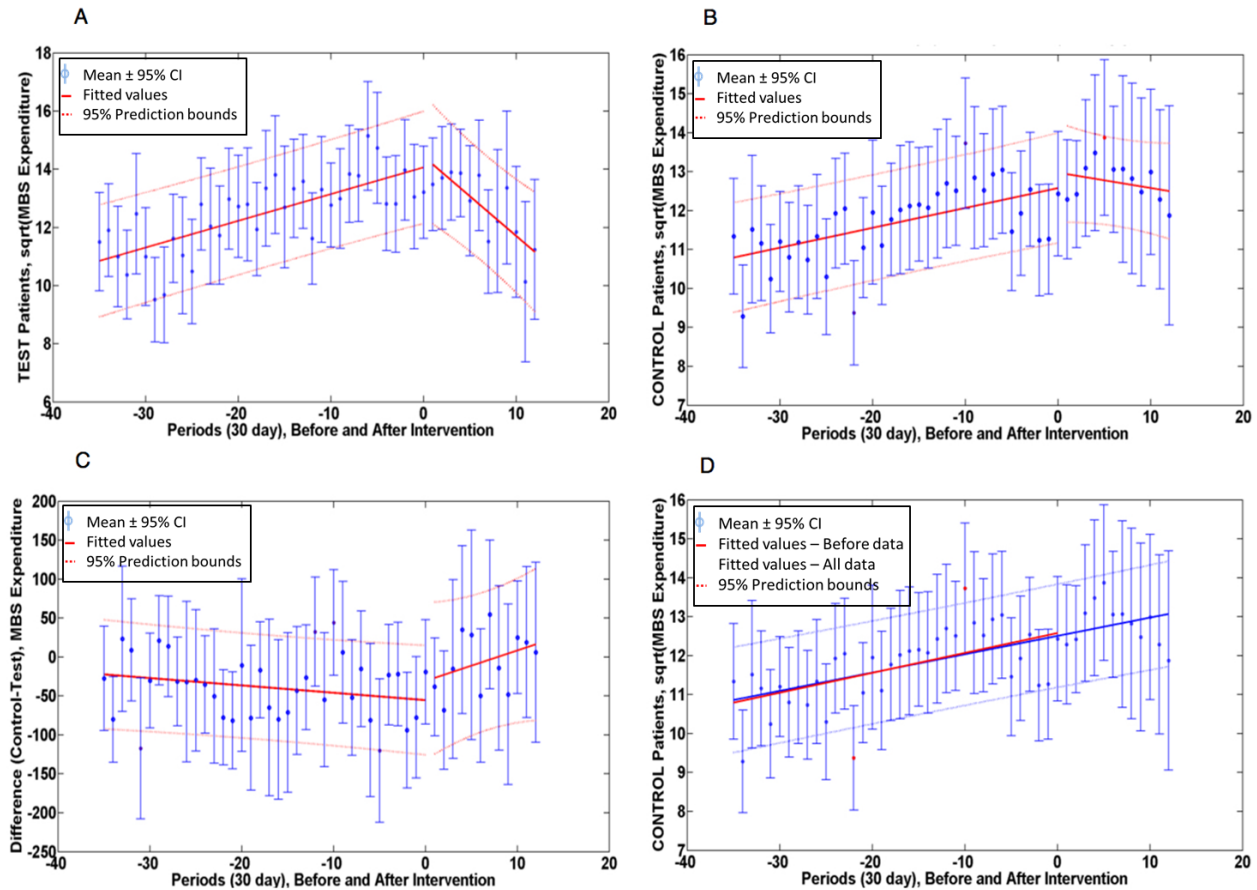
Comparing Slopes Before Intervention

The plots shown above the linear regression fits and the results of the ANCOVA analysis for Figure 2 are given in tabular form in Multimedia Appendix 1, which also gives linear regression data for pharmaceutical expenditure, number of admissions to hospital, and hospital LOS.

Before the intervention, test patients and control patients had no significant difference in the rate of admission to hospital (*P*=.443), but test patients had a significantly greater rate of LOS (*P*=.013) in hospital.

Comparing Before and After Slopes

For test patient medical expenditure, the slope before the intervention was significantly reduced (*P*<.001) following the intervention, indicating a significant reduction in the rate of medical expenditure. In contrast, for control patients the change in slope was not significant (*P*=.10).

For pharmaceutical expenditure, the fall in slope for test patients was highly significant (*P*<.001), while control patients showed a marginally significant (*P*=.046) increase in slope. As a result,

the change in slope of the (control-test) difference was also significantly different (*P*=.008).

Control patients had no significant difference (*P*=.458) in their rate of admission to hospital before and after the intervention, while test patients had a significant fall (*P*=.009). Similarly, control patients had no significant change in LOS (*P*=.869), while test patients showed a significant (*P*=.006) fall in their LOS after the intervention, and the differences in (control-test) slopes before and after intervention were also significantly different (*P*<.001).

Estimating Changes in the Rate of Expenditure and Savings After One Year

Both the change in the rate of expenditure and savings in the year following the start of intervention for medical benefits or pharmaceutical benefits expenditure as well as number of admissions and LOS were estimated from the linear regression equations given in Multimedia Appendix 1. The method for estimating changes in rates and savings over the year is demonstrated in Figure 3 using medical benefits expenditure as an example.

The linear regressions for sqrt(30 day medical costs) developed for test patients, control patients, and differences (control-test) provide a best fit estimate of expenditure before and after intervention. The regression equations for data for 3 years prior to the intervention are projected forward by 1 year to estimate the predicted costs 1 year after the start of intervention.

This is shown in Figure 3 with some simplification for medical costs for all test patients. In Figure 3, the average age of test patients was approximately 71 years old at the start of intervention and was used as the reference point. The difference between the projected curve and the actual expenditure curve, representing the estimated saving over 1 year, was $720 or 28% of the projected expenditure.

However, the assumption that the 2 curves meet exactly at the onset of intervention is a simplification that may overestimate the savings. If indeed the impact of intervention needs some time to take effect, we would expect the point of intersection to fall sometime after the start of telemonitoring, subject to the variability of the expenditure data. This is in fact what was observed in the majority of cases as shown in Figure 4 for medical expenditure for all test patients.

Figure 4 shows that the curvilinear function for control patients before intervention was extended to the after period. For test patients, the intercept of the 2 curvilinear plots before and after intervention occurred at approximately 31 days after the start of intervention, leading to a reduced estimate of the savings in medical benefits expenditure from $720 to $611 per annum. Applying a similar analysis to the regression equation for differences (control-test) in Multimedia Appendix 1 (panel A) results in a similar estimate of savings of $657 per annum.

Estimates for the reduction in rates of expenditure for medical and pharmaceutical costs, number of admissions, LOS, and average savings over 1 year is given in Multimedia Appendix 2.

Figure 3. Model-based method of estimating impact of telemonitoring on medical expenditure.
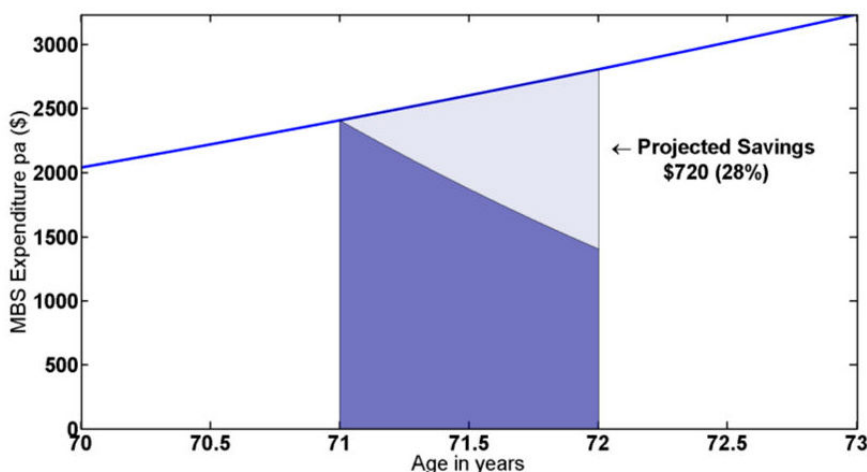

Figure 4. Regression-based estimates of time course of annual medical benefits expenditure for test patients and control patients, before and after intervention. Based on data presented in Multimedia Appendix 1.
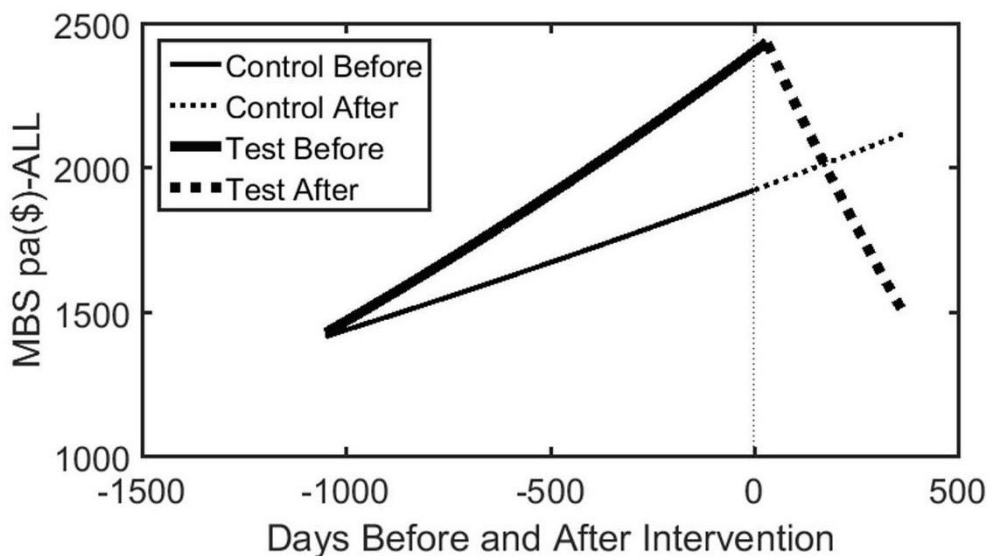
Table 2. Age-specific death rates of test patients.

| | Age distribution | | | | | |
	50-60	60-70	70-80	80-90	90-100	Total
Age distribution in MR[a]	180	310	441	414	84	1429
Age distribution, %	12.60	21.69	30.86	28.97	5.88	100.0
Deaths in MR[a] , n	17	46	60	91	37	251
ASDR[b] , %	9.44	14.84	13.61	21.98	44.05	17.56[c]
Test patients by age, n	41	31	14	13	1	100
Age specific deaths, n	1	2	4	1	0	8
Expected deaths, n	3.87	4.60	1.91	2.86	0.44	13.68
Deaths saved, n	2.87	2.60	–2.09	1.86	0.44	5.68

[a]MR: master register.

[b]ASDR: age-specific death rate.

[c]Crude death rate.

Effect of Intervention on Mortality

A total of 57 test patients and 76 control patients in the study were from the MR of 1429 patients. The crude death rate was 8.8% (5/57) for test patients and 17.1% (13/76) for their matched controls, giving a reduction in mortality of 48.5%.

For the 100 test patients for whom survival data was accurately available through the BDMR in each state, the ASDR of the test patients relative to those from the MR file are given in Table 2.

Using ADSRs in Table 2 calculated from the MR of eligible patients, 13.68 deaths were expected but only 8.0 were recorded. This represents a saving of 5.68 lives, a reduction of 41.5%.

Discussion

Principal Findings

The results of this study demonstrate a statistically robust positive impact, increasing over time, of at-home telemonitoring on health care expenditure, number of admissions to the hospital, and LOS as well as a reduction in mortality.

Table 1 demonstrates that test patients and their controls were generally well matched with respect to expenditure on pharmaceutical and medical items at the start of intervention. However, in the last 100 days prior to the end of the intervention, test patients were spending on average $3298 less on medical and pharmaceutical items than control patients.

Multimedia Appendix 1 shows that for the 3 years before the intervention, there was a significant difference in slope between test patients and control patients for medical and pharmaceutical expenditure and hospital LOS but not for the number of hospital admissions.

Interestingly, the slope for medical expenditure prior to intervention was larger for test patients than control patients, but the slope for pharmaceutical expenditure was smaller for test patients than control patients, thus indicating that test patients were more likely than control patients to use medical services but were less likely to spend money on pharmaceutical prescription medications.

Impact of Telemonitoring on Medical and Pharmaceutical Expenditure

The predicted rate of medical and pharmaceutical expenditure one year after the start of intervention was estimated as $2803 per annum and $3176 per annum, respectively (Multimedia Appendix 2). As a result of the telemonitoring intervention, these rates of expenditure fell to $1504 per annum and $2365 per annum, a reduction of 46.3% and 25.5%, respectively. Over the year of the intervention, average savings in medical and pharmaceutical costs were estimated as $611 and $354, or 23.5% and 11.5%, respectively. However, differences (control-test) data suggest that savings in pharmaceutical costs may be marginal.

Impact of Telemonitoring on Hospital Admissions and Length of Stay

Test patients at the start of telemonitoring had a rate of LOS averaging 19.8 days, which after 1 year were projected to increase to 24.6 days per annum (Multimedia Appendix 2). Telemonitoring reduced the projected yearly rate of LOS after 1 year from 24.6 days per annum to 7.9 days per annum, a reduction of 67.9%. Over the year following the telemonitoring intervention, this leads to an average saving of 7.5 days or 33.8% in hospital stays relative to the 22.2 days predicted over that year without the intervention.

Effect of Telemonitoring on Mortality

The crude death rate was 8.8% for test patients and 15.8% for their matched controls, giving a reduction in mortality of 48.5% (Table 2). A more accurate method based on comparison of ASDRs of 100 test patients (8 deaths) for whom survival data was accurately available against the expected ASDRs generated from a master registry of 1429 patients (13.64 deaths) indicated a reduction in mortality of 41.5%.

XSL•FO
RenderX

Generalization of Trial Results

The project design was a multistate, multisite trial along the eastern seaboard of Australia, and because health service provision across the country in urban settings is relatively uniform because of Medicare, the government-funded universal health care system, we believe that results can be generalized to the broader urban Australian population but not necessarily to rural and remote locations or other countries with different health systems.

These results are broadly in agreement but more favorable than those reported for the UK WSD trial [8-10] or the US Veterans Health Administration (VHA) Care Coordination/Home Telehealth (CCHT) program [19].

The headline findings for the WSD [8-10] included a 15% reduction in accident and emergency visits, a 20% reduction in emergency admissions, a 14% reduction in elective admissions, a 14% reduction in bed days, an 8% reduction in tariff costs, and a 45% reduction in mortality rates.

The differences in the results reported in this study can be attributed to different protocols for patient selection (general practitioner selection vs selection of matched test and controls patients from hospital lists) as well as differences in the quality and mode of analysis of the available data.

Between July 2003 and December 2007, the VHA implemented a national home telehealth program, CCHT, that supports the care for veterans with chronic conditions in their homes as they age.

The technology adopted in VHA service was considerably different to the telemonitoring technology used in this study and included videophones, messaging devices, biometric devices, digital cameras, and telemonitoring devices. More importantly, the age distribution was considerably different and included participants as young as 20 and older than 80 years with a wider range of conditions including posttraumatic stress disorder, depression, and other mental health condition.

Routine outcomes analysis for performance measurement of health care resource utilization by CCHT patients involved comparing hospital admission data for patients during the year prior to enrollment into CCHT with the data from 6 months postenrollment. This cohort of patients had a 19.74% reduction in hospital admissions and 25.31% reduction in bed days of care (BDOC) following enrollment into the CCHT program. During the same time period, there was a decline of 4.6% in BDOC for all patients enrolled within VHA, which needs to be taken into account when interpreting this change. Given the size, complexity, and resourcing of this program and the comprehensive and systematic approach to the clinical, educational, technology, and business processes that constitute VHA's CCHT model of care, it is impossible to make a formal comparison of results. However given the small size, tight control on eligibility, and the greater homogeneity of the study cohort in our study, it is not surprising that we report considerably better results.

Limitations

Like all complex clinical trials this project suffered numerous setbacks. Some of the major issues that impacted execution of the trial and subsequent data analysis are as follows:

- A significant number of eligible patients, including 93 test and 33 control patients, declined to participate, while 27 of those who agreed to participate were not able to commence and 18 test patients withdrew after they had begun monitoring.
- We recruited and consented 114 test patients and 173 control patients, but of these, only 71 test patients and 110 control patients were from the hospital lists provided. This caused considerable difficulty in the reliable assessment of mortality and the analysis of hospital admissions and LOS.
- Of the 114 test patients consented, 14 had missing data in their DHS records and had to be removed from further analysis. Similarly, of the 173 test patients consented, only 137 patients had complete DHS data. No explanation was available from DHS as to why some patients had missing data in their records.
- Test patients were recruited and initiated telemonitoring over a long period of time so that while the average number of days that patients were monitored was 276 days, there was a considerable spread, from <100 days to >500 days. The period for analysis of the effect of telemonitoring was thus limited to 12 months as patient numbers rapidly fall and the data spread increases for periods >12 months.
- For some patients consented early in the trial, signed consent was provided only through June 2014. When the trial duration was extended to the end of December 2014, new consent forms for the extended period were not signed and as a result, DHS data for these patients were only available through June 2014.
- CCCs were typically registered nurses employed by the service providers participating in the study. Most did not have any previous experience with telemonitoring but all graduated from a 2-day intensive training program on how to use the telemonitoring equipment and how to interpret the clinical data recorded. On average, CCCs spent a little over 33 minutes per week reviewing individual patient data, suggesting that as their average patient load was 20 patients, they were employed in this role <30% full-time over the week. This is less than optimal for this critical role.
- Although test and control patients were generally well matched by primary diagnosis, number of hospital admissions, and SEIFA index across sites, it was later observed that historical rates of medical and pharmaceutical expenditure were not well matched at the start of telemonitoring, as shown in Figure 4. Since the historical rate of growth of medical expenditure may be a good proxy for the present level of severity of a patient's chronic condition, future studies should consider using this variable to match test and control patients in addition to the matching criteria used in this study.

Conclusion

At-home telemonitoring leads to a significant time-dependent reduction in expenditure on medical services, a reduction in the

number of hospital admissions, and a reduction in LOS averaging 7.5 days per annum. Mortality of test patients relative to control patients was also reduced by between 41.5% and 48.5% over the period of the trial.

It is not possible from this study to separate the effect of care coordination and coaching by the CCC from the direct and exclusive impact of at-home telemonitoring and patient self-management. However, the data presented shows clearly that the impact of at-home telemonitoring increases almost linearly over the first year following the intervention.

One would envisage that the impact would inevitably plateau and possibly begin to rise with increasing age and morbidity; however, longer term studies are required to elucidate the impact of telemonitoring over longer time frames.

Further research is also required to understand why hospital admissions that were recorded could not be avoided. Did the available vital signs not provide a sufficient warning of an exacerbation or were these warning signs ignored or not acted upon in a timely fashion by the nurse coordinator or patient's PCP?

A detailed cost-benefit analysis of at-home telemonitoring as well as organizational change management requirements and workplace cultural issues that need to be considered in delivering the services reported in this study will be reported separately.

However, a preliminary cost-benefit analysis based on an estimate of the cost of delivering the telemonitoring service—approximately Aud $2760 per annum (Aud $7.40 per day)—against potential savings of more than Aud $19,000 per annum based on average cost of one bed day of Aud $2051 provides a return of investment of approximately 6 times.

Conflicts of Interest

There was no conflict of interest during the planning and execution of the project. Six months after its completion, Professor Branko Celler, Chief Investigator and Project Director, was appointed to a part-time position at Telemedcare Pty Ltd as Director of Research.

Multimedia Appendix 1

Linear regression and analysis of covariance for (1) sqrt(medical expenditure), (2) sqrt(pharmaceutical expenditure), (3) number of hospital admissions, and (4) hospital length of stay.

[PDF File (Adobe PDF File), 70KB - medinform_v5i3e29_app1.pdf]

Multimedia Appendix 2

Estimated changes in medical and pharmaceutical expenditure, hospital admissions, and length of stay for test patients with and without intervention.

[PDF File (Adobe PDF File), 10KB - medinform_v5i3e29_app2.pdf]

References

1. National Center for Chronic Disease Prevention and Health Promotion. The Power of Prevention. 2009. URL: https://www.cdc.gov/chronicdisease/pdf/2009-Power-of-Prevention.pdf [accessed 2017-07-09] [WebCite Cache ID 6rp05KjQZ]
2. Anderson G, Horvath J. The growing burden of chronic disease in America. Public Health Rep 2004 May;119(3):263-270 [FREE Full text] [doi: 10.1016/j.phr.2004.04.005] [Medline: 15158105]
3. Australian Institute of Health and Welfare. Health expenditure Australia 2011–12. URL: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129544656 [accessed 2017-07-09] [WebCite Cache ID 6rp0wGkGv]
4. Totten A, Womack D, Eden K, McDonagh M, Griffin J, Grusing S, et al. Telehealth: mapping the evidence for patient outcomes from systematic reviews. URL: https://www.effectivehealthcare.ahrq.gov/ehc/products/624/2317/telehealth-disposition-160929.pdf [accessed 2017-03-13] [WebCite Cache ID 6ovRiHSN0]
5. McBain H, Shipley M, Newman S. The impact of self-monitoring in chronic illness on healthcare utilisation: a systematic review of reviews. BMC Health Serv Res 2015;15(1):1.

XSL•FO

RenderX

6.   Brettle AJ, Brown TM, Hardiker NR, Radcliffe JN, Smith CL. Telehealth: the effects on clinical outcomes, cost effectiveness and the patient experience: a systematic overview of the literature. URL: http://usir.salford.ac.uk/29392/4/Telehealth_v8_.pdf [accessed 2017-03-13] [WebCite Cache ID 6ovRoMB5m]

7.   Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assn 2007;14(3):269-277.

8.   Bower P, Cartwright M, Hirani S. A comprehensive evaluation of the impact of telemonitoring in patients with long-term conditions and social care needs: protocol for the whole systems demonstrator cluster randomised trial. BMC Health Serv Res 2011;11(1):184.

9.   Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Beynon M, et al. Effect of telecare on use of health and social care services: findings from the Whole Systems Demonstrator cluster randomised trial. Age Ageing 2013;42(4):501-508.

10. Steventon A, Ariti C, Fisher E, Bardsley M. Effect of telehealth on hospital utilisation and mortality in routine clinical practice: a matched control cohort study in an early adopter site. BMJ Open 2016;6(2).

11. Bradford N, Caffery L, Smith A. Telehealth services in rural and remote Australia: a systematic review of models of care and factors influencing success and sustainability. 2016. URL: http://www.rrh.org.au/articles/subviewnew.asp?ArticleID=3808 [accessed 2017-03-13] [WebCite Cache ID 6ovRu4MVm]

12. Medical Technology Association of Australia. A telehealth strategy for Australia: supporting patients in the community. URL: https://www.mtaa.org.au/sites/default/files/uploaded-content/website-content/supporting-a-telehealth-strategy-for-australia-release-version-may-2012.pdf [accessed 2017-07-09] [WebCite Cache ID 6rp1U4P2z]

13. Celler BG, Sparks R, Nepal S, Alem L, Varnfield M, Li J, et al. Design of a multi-site multi-state clinical trial of home monitoring of chronic disease in the community in Australia. BMC Pub Health 2014;14(1):1270.

14. Petticrew M, Cummins S, Sparks L, Findlay A. Validating health impact assessment: prediction is difficult (especially about the future). Environ Impact Asses 2007;27(1):101-107.

15. Morrison-Saunders A, Bailey J. Practitioner perspectives on the role of science in environmental impact assessment. Environ Manage 2003 Jun;31(6):683-695. [doi: 10.1007/s00267-003-2709-z] [Medline: 14565690]

16. Underwood A. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. Mar Freshwater Res 1991;42(5).

17. Jitapunkul S, Pillay I, Ebrahim S. The abbreviated mental test: its use and validity. Age Ageing 1991;20(5):332-336.

18. Information paper: an introduction to socio-economic indexes for areas (SEIFA). 2006. URL: http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/D729075E079F9FDECA2574170011B088/$File/20390_2006.pdf [accessed 2017-03-13] [WebCite Cache ID 6ovUwxhDi]

19. 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]

Abbreviations

ANCOVA: analysis of covariance
ASDR: age-specific death rate
BACI: before and after control intervention
BDMR: Births, Deaths, and Marriages Register
BDOC: bed days of care
CCC: clinical care coordinator
CCHT: Care Coordination/Home Telehealth
CSIRO: Commonwealth Scientific and Industrial Research Organisation
DHS: Department of Human Services
HREC: Human Research Ethics Committee
LOS: length of stay
MBS: medical benefits scheme (medical expenditure)
MR: master register of eligible patients
PBS: pharmaceutical benefits scheme (pharmaceutical expenditure)
PCP: primary care physician
PO: project officer
R2: coefficient of determination
SEIFA: socioeconomic indexes for areas
sqrt: square root
TMU: telemonitoring unit
VHA: Veterans Health Administration

XSL·FO

RenderX

Viewpoint

A Roadmap for Optimizing Asthma Care Management via Computational Approaches

Gang Luo[1], PhD; Katherine Sward[2], RN, PhD

[1]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States
[2]College of Nursing, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Gang Luo, PhD
Department of Biomedical Informatics and Medical Education
University of Washington
UW Medicine South Lake Union
850 Republican Street, Building C, Box 358047
Seattle, WA, 98109
United States
Phone: 1 206 221 4596
Email: gangluo@cs.wisc.edu

Abstract

Asthma affects 9% of Americans and incurs US $56 billion in cost, 439,000 hospitalizations, and 1.8 million emergency room visits annually. A small fraction of asthma patients with high vulnerabilities, severe disease, or great barriers to care consume most health care costs and resources. An effective approach is urgently needed to identify high-risk patients and intervene to improve outcomes and to reduce costs and resource use. Care management is widely used to implement tailored care plans for this purpose, but it is expensive and has limited service capacity. To maximize benefit, we should enroll only patients anticipated to have the highest costs or worst prognosis. Effective care management requires correctly identifying high-risk patients, but current patient identification approaches have major limitations. This paper pinpoints these limitations and outlines multiple machine learning techniques to address them, providing a roadmap for future research.

Introduction

Asthma affects 9% of Americans [1-3] and incurs US $56 billion in cost [4], 3630 deaths, 439,000 hospitalizations, and 1.8 million emergency room visits annually [1]. As is true for many chronic diseases, a small fraction of asthma patients with severe disease, high vulnerabilities, or great barriers to care consume most health care costs and resources [5,6]. The top 20% of patients consume 80% of costs, and the top 1% consume 25% [6,7]. An effective approach is needed to find high-risk patients and implement appropriate interventions to improve outcomes and to reduce costs and resource use.

Almost all private health plans provide, and most major employers purchase, care management services that implement tailored care plans with early interventions for high-risk patients to avoid high costs and health status degradation [8-10]. Care management is a cooperative process to assess, plan, coordinate, implement, evaluate, and monitor services and options to fulfill a patient's health and service needs [11]. It includes a care manager who regularly calls the patient, arranges for health and related services, and helps make medical appointments. Asthma exacerbations account for 63% of annual total asthma cost [12,13]. Using care management properly can reduce asthma exacerbations, cut hospital admissions and emergency room visits by up to 40% [9,14-18], trim cost by up to 15% [15-19], and enhance patient treatment adherence, quality of life, and satisfaction by 30% to 60% [14].

Although widely used, care management has costs of its own and can require more than US $5000 per patient per year [15]. Owing to resource constraints, usually only 1% to 3% of asthma patients are enrolled in care management [7]. Ideally, the ones enrolled should be those at the highest risk. Predictive modeling is the best method to find high-risk patients [20]. It uses a model for predicting individual patient cost or health outcome to automatically find high-risk patients [14,21-26]. Cost reflects use and efficiency of care and indirectly reflects outcomes such

as hospitalization and emergency room visit. For patients predicted to have the highest costs or worst outcomes, care managers examine patient records, consider various factors such as social ones, and make the ultimate allocation and intervention decisions. Correct identification of high-risk patients is key to effective care management, but current identification methods have limitations. This paper makes two contributions. First, we pinpoint these limitations. Second, we outline several machine learning techniques to address them, offering a roadmap for future research. Clinical machine learning is a promising technology for finding high-risk patients [27]. Our discussion focuses on the machine learning predictive modeling aspect of care management for identifying high-risk patients. Besides this, several other factors such as patient behavior pattern, patient motivation, trigger for patient engagement [28-30], and patient and caregiver education [31] also impact a care management program's performance. A detailed discussion of how to incorporate or change these factors to optimize asthma care management is beyond this paper's scope.

Limitations of Current Patient Identification Methods for Asthma Care Management

Limitation 1: Low Prediction Accuracy Causes Misclassification, Unnecessary Costs, and Suboptimal Care

Current predictive models for individual patient costs and health outcomes exhibit poor accuracy causing misclassification and need improvement. When projecting individual patient cost, the R^2 accuracy measure of models reported in the literature is less than 20% [32], and the average error is typically comparable to the average cost [33]. When projecting individual patient health outcome, the area under the receiver operating characteristic curve accuracy measure is usually much smaller than 0.8 [6,34-40]. Those large errors make enrollment miss more than half of patients a care management program can help most [14,26]. Weir et al [26] showed that the top 10% risk group identified by a predictive model missed more than 60% of the top 10% and about 50% of the top 1% of patients who had the largest costs. If we could find 10% more of the top 1% patients who had the largest costs and enroll them, we could save up to US $210 million in asthma care each year and also improve health outcomes [6,36,37]. In general, because of the large patient base, a small improvement in accuracy will benefit many patients, having a large positive impact. A 5% absolute improvement in accuracy already makes a health care system willing to deploy a new model [41].

Existing predictive models have low accuracy for multiple reasons, which include the following:

1. Although several dozen risk factors for adverse outcomes in asthma are known [6,18,36,39,40,42-46], an existing model typically uses only a few of them (eg, less than 10) [6,36-39]. Existing models were often constructed using data obtained from clinical trials or old-fashioned electronic medical records (EMRs) that collected only a limited set of variables [47]. No published model explores all known

risk factors available in modern EMRs, which collect an extensive set of variables [47].

2. As with many diseases, many features (also known as independent variables that include both raw and transformed variables) predictive of adverse outcomes in asthma have likely not been identified. For instance, using a data-driven approach to find new predictive features from many variables in EMRs, Sun et al [48] improved prediction accuracy of heart failure onset by more than 20%. Existing predictive models for health outcomes of individual asthma patients were developed mainly using a small number of patients (eg, <1000) or variables (eg, <10) [6,36-39], creating difficulty in finding many predictive features and their interactions.

3. Existing models mainly use patient features only, presuming that each patient's cost and health outcomes relate only to the patient's characteristics and are unassociated with characteristics of the health care system (eg, the treating physician and facility). However, system features are known to be influential, have larger impacts on patients with the worst outcomes, and can account for up to half of the variance in their outcomes in certain cases [49-52]. The use of physician characteristics has been examined in predictive modeling only minimally [35], creating a knowledge gap for system features in general.

4. Applying care management to a patient tends to improve the patient's health outcomes and reduce the patient's cost, excluding the cost of care management. Yet, existing models omit the factor of care management enrollment.

5. A health care system often has limited training data, whereas a model's accuracy generally increases with more training data. Different systems have differing data distributions [53] and collected attributes, impacting the performance and applicability of a model trained using one system's data for another system [54-57]. To address these two issues, one can perform transfer learning and use other source systems' information to improve model accuracy for the target system [54,58,59]. Transfer learning typically requires using other source systems' raw data [60,61]. However, systems are rarely willing to share their raw data because of confidentiality concerns with regard to patient data. Research networks [62-64] mitigate, but do not solve, the problem. Many systems are outside a network, whereas systems in it share raw data of limited attributes. Alternatively, one can conduct model updating, model averaging, or ensemble averaging that requires only the trained models, but not the raw data, from other source systems. Model updating applies to only one source system and cannot combine information from multiple source systems, limiting the improvement in model accuracy. Many model updating methods work for only certain kinds of models [65]. Model averaging usually employs the same averaging approach such as weights in all regions of the feature space [66]. Yet, to boost model accuracy, different averaging approaches are often needed in differing regions [67]. Also, if the target system does not have enough data to train a reasonably accurate model as a starting point, further averaging with the trained models from other source

systems may not improve the final model's accuracy to a satisfactory level.

Limitation 2: No Explanation of the Reasons for a Prediction Causes Poor Adoption of the Prediction and Busy Care Managers to Spend Extra Time and Miss Suitable Interventions

Unlike physicians who see patients regularly, care managers often have no prior interaction with a patient and are unfamiliar with the patient's medical history when they need to make enrollment decisions. They need to understand why a patient is forecasted to be at high risk before allocating to care management and creating a tailored care plan, but have limited time to review extensive patient records with many variables, possibly accumulated over a long time and often including hundreds of pages [68]. Patients are at high risk for various and often multiple reasons, each linking to one feature or a combination of several features. Each combination represents a risk pattern rather than a risk factor (a single variable) and cannot be found by regular risk factor finding methods. An example risk pattern is that the patient had two or more urgent care visits for asthma last year AND lives 15 miles or more away from the patient's physician. Complicated predictive models, covering the majority of machine learning models such as random forest, provide no justification for predictions of high risk. This causes poor adoption of the prediction and forces care managers to spend great effort finding root causes, which often involves manual temporal aggregation of clinical variables such as counting urgent care visits. This is time consuming, likely to miss more patients who would gain most from care management, and difficult to do when appropriate cut-off thresholds for numerical variables (eg, 15 miles in distance) are unknown.

Existing predictive models provide limited help in creating tailored care plans. An intervention targeting the reason underlying the high risk is likely to have better effect than nonspecific ones. A patient can have high risk for several reasons. A care manager may develop a tailored care plan for a patient using variable and subjective clinical judgment, but may miss certain suitable interventions because of the following reasons. First, many features exist. As true for any human, a typical care manager can process no more than 9 information items at once [69], making it hard to find all reasons from many possible feature combinations. Second, considerable practice variation such as by 1.6 to 5.6 times appears across differing care managers, facilities, and regions [5,34,70-78]. Third, care managers usually include in the care plan interventions addressing patient factors only. For the health care system, some useful interventions such as extending physician office hours are not identified as possible interventions. Interventions at the system level can be more efficient and effective than those for patients [50,79]. Some system levels, such as treating physicians, are more accessible than patients. An intervention at the system level can affect many patients, whereas an intervention for a patient affects only that patient. Missing suitable interventions degrades outcomes.

Limitation 3: For Patients on Care Management, a Lack of Causal Inference Capability Causes the

Predictive Model to Give No Clear Guidance on Which Patients Could Be Moved off Care Management

An asthma patient's risk changes over time, whereas a care management program can enroll only limited patients. To best use the program, all patients remaining in the health plan are reevaluated for their risk periodically, for example, on an annual basis. The patients who are in the program and now predicted to be at low risk are moved off the program to make room for those previously at low risk but now at high risk. Doing this properly requires answering intervention queries via causal inference [80,81], which is beyond most existing predictive models' capability. Some patients in the program are in a stable status and ready to safely leave the program. For some others, using the program is essential for keeping them at low risk. Moving them off the program can lead to undesirable outcomes. An existing model can predict a patient in the program to be at low risk, but often does not tell which of the two cases the patient falls into and does not give clear guidance on whether the patient should be moved off the program. This is particularly the case if we expect care management to have greatly varying impact across different subgroups of patients and would like to consider their differences in impact explicitly.

Machine Learning Techniques for Optimizing Asthma Care Management

New techniques are needed to identify more high-risk asthma patients and provide appropriate care. Besides those proposed in our paper [27], we can use the following machine learning techniques to optimize asthma care management.

Techniques for Improving Prediction Accuracy of Individual Patient Costs and Health Outcomes

Use All Known Risk Factors for Adverse Asthma Outcomes Available in Modern Electronic Medical Records

Many risk factors for adverse asthma outcomes are known [6,13,36,39,40,42-46] and available in modern EMRs. To fully use their predictive power, we consider all of these risk factors when building models for predicting individual patient costs and health outcomes. We perform feature selection to remove known risk factors that are not predictive for reasons such as data quality and variable redundancy. Clinical experts can suggest for consideration additional features that have clear medical meaning but have not previously been used for predicting asthma outcomes or costs. Two examples of such features are as follows: (1) exercise vital signs and (2) whether the patient has seen an asthma specialist (allergist or pulmonologist) recently. Patients who have seen asthma specialists tend to have more severe asthma, worse health outcomes, and higher costs than those who have seen primary care physicians only. Another way to consider this factor is to build separate models for patients who have seen asthma specialists and patients who have seen primary care physicians only.

Use Many Asthma Patients and Patient Features

Many features predictive of adverse outcomes in asthma have not yet been identified. To find new predictive features, we use many asthma patients and a data-driven approach to explore many standard patient features listed in the clinical predictive modeling literature [5,34,65]. Some patient features cover social, economic, and community factors. An example of such features is the average income level of the area that the patient lives in. To combine known risk factors and predictive features derived from data, during feature selection we give a higher weight to known risk factors (eg, by multiplying their scores by a factor larger than 1) so that they are more likely to be selected than the other features. This new approach can handle both categorical and numerical variables, discover new predictive features, and remove known risk factors that are not predictive for reasons such as data quality and variable redundancy. In contrast, the existing method for combining known risk factors and predictive features derived from data [48,82] can neither directly handle categorical variables nor remove known nonpredictive risk factors.

Use Health Care System Features

To consider their impact, we include health care system features in building models for predicting individual patient costs and health outcomes. For each health care system level, such as physician or facility, we construct a profile containing its own information (eg, facility hours) and aggregated historical data of its patients (omitting the index patient) extracted from the provider's administrative and EMR systems. The count of the physician's asthma patients [83] is an example of profile variables.

Some system features are computed using only system profile variables. Our paper [27] listed several physician-level features such as the average outcome of a physician's asthma patients. Examples of facility-level features are as follows: (1) whether a facility is open at night or on weekends, (2) the number of staffed beds in a hospital, (3) facility type, and (4) availability of enhanced services such as asthma hotline, foreign language translation, special primary care team for asthma, and special home care. The other system features are computed by combining system profile and patient variables, reflecting the match of physician or facility and patient. An example of such features is the distance between the patient's home and closest urgent care facility.

Use All Patients

The standard approach for predicting individual patient costs or health outcomes in asthma is to build a model using only asthma patient data. In the presence of many features, we may not have enough asthma patients to train the model and to obtain high prediction accuracy. To address this issue, we add a binary indicator feature for asthma and train the model on all patients, not just asthma patients. Asthma patients and other patients share many features in common. We can better tune these features' coefficients in the model by using all patients.

Consider the Factor of Care Management Enrollment

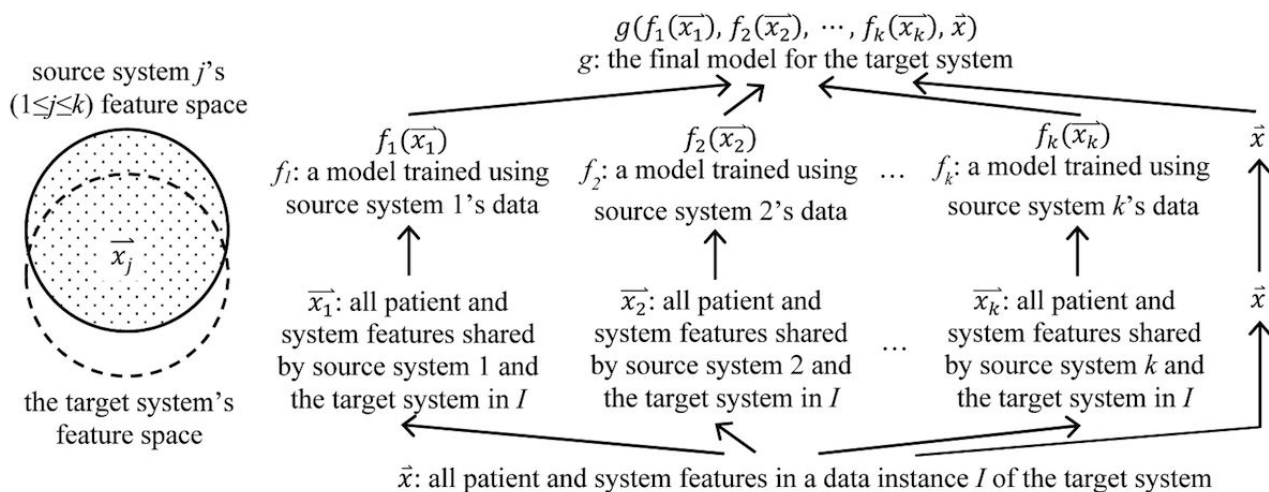To consider care management's impact on costs and health outcomes, we add a binary indicator feature for care management enrollment when building models for predicting individual patient costs and health outcomes [84].

Perform Transfer Learning Using Trained Models From Other Source Health Care Systems

To address limited training data and improve model accuracy for the target health care system, we perform transfer learning using trained models from other source systems. Organizations are usually more willing to share their trained models than their raw data. Publications often describe trained models in detail. A model trained using a source system's data includes much information useful for the prediction task on the target system, particularly if the source system has lots of training data. Our transfer learning approach can handle all kinds of features, prediction targets, and models used in the source and target systems. Our approach can potentially improve model accuracy regardless of the amount of training data available at the target system. Even if the target system has enough training data in general, it may not have enough training data for a particular pattern. A trained model from a source system can help make this up if the source system contains enough training data for the pattern.

Different health care systems use differing schemas, medical coding systems, and medical terminologies. To enable the application of a model trained using a source system's data to the target system's data, we convert the datasets of every source system and the target system into the same common data model (eg, OMOP [85]) format and its linked standardized terminologies [86]. For each available source system, we use the method described in our paper [27] to form a table listing various combinations of attributes. For each combination of attributes, the table includes the model trained using it and the source system's data. For the combination of attributes collected by both the source and target systems, we find the corresponding model trained for the source system. For every data instance of the target system, we apply the model to the data instance, obtain the prediction result, and append it as a new feature to the data instance. In this way, the expanded data instance includes two types of features: (1) the new features obtained using the models trained for the source systems, with one feature per source system and (2) the patient and system features in the target system. For the target system, we use both types of features and its data to build the final model (Figure 1). As correlation exists among features of the first type constructed for the same prediction target, regularization is likely needed to make the final model stable. Features of the second type can either serve as inputs to the final model directly or be used to build a model whose output serves as an input to the final model. If the target system has limited training data, we perform aggressive feature selection on the second type of features to let the number of remaining features match the amount of training data. This does not impact the first type of features. When a source system has enough data to train a model, the model can include many patient and system features as its inputs. The corresponding feature of the first type is computed using these inputs. In this case, the final model for the target system uses information from many patient and system features, regardless of whether the target system has a large amount of training data.

Figure 1. An illustration of our transfer learning approach.

$$g(f_1(\vec{x_1}), f_2(\vec{x_2}), \cdots, f_k(\vec{x_k}), \vec{x})$$
g: the final model for the target system

source system j's ($1 \le j \le k$) feature space

$\vec{x_j}$

the target system's feature space

$f_1(\vec{x_1})$
f_1: a model trained using source system 1's data

$f_2(\vec{x_2})$
f_2: a model trained using source system 2's data

$f_k(\vec{x_k})$
f_k: a model trained using source system k's data

$\vec{x}$

$\vec{x_1}$: all patient and system features shared by source system 1 and the target system in I

$\vec{x_2}$: all patient and system features shared by source system 2 and the target system in I

$\vec{x_k}$: all patient and system features shared by source system k and the target system in I

$\vec{x}$

$\vec{x}$: all patient and system features in a data instance I of the target system

Techniques for Creating a New Function to Automatically Explain Prediction Results for Identified High-Risk Patients

To improve prediction accuracy, it is desirable to use machine learning to construct models for predicting individual patient costs and health outcomes [27]. For patients with projected risk above a fixed threshold, such as the 95th percentile, we can use our previously developed method [27,87] to automatically explain machine learning prediction results with no accuracy loss. The explanations can help clinicians make care management enrollment decisions, identify interventions at various levels, create tailored care plans based on objective data, and inspect structured attributes in patient records more efficiently. An example of patient interventions that can be put into tailored care plans is to offer transportation for a patient living far from the primary care physician. An example of interventions at the system level is to launch a new primary care clinic in a region with no such clinic close by.

Each patient has the same set of patient and health care system features and is marked as high or not high risk. Our method mines from historical data class-based association rules related to high risk. Each rule's left hand side is the conjunction of one or more feature-value pairs. An example rule is as follows: the patient had two or more urgent care visits for asthma last year AND lives 15 miles or away from the patient's physician → high risk. Through discussion and consensus, the clinicians in the automatic explanation function's design team check mined rules and drop those making no or little clinical sense. For every rule kept, the clinicians enumerate zero or more interventions targeting the reason the rule shows. At prediction time, for each patient the predictive model identifies as high risk, we find and display all rules of which the patient fulfills the left hand side conditions. Every rule shows a reason why the patient is projected to be at high risk.

Conditional Risk Factors

Our method can find a new type of risk factor termed conditional risk factors, which increase a patient's risk only when some other variables are also present and can be used to design tailored interventions. This broadens risk factors' scope, as

ordinary risk factors are independent of other variables. Our method can automatically find appropriate cut-off thresholds for numerical variables and inform new interventions based on objective data. For instance, for the aforementioned association rule, our method would automatically find the cut-off thresholds of two in the number of urgent care visits and 15 miles in distance. Then we map all the patients who satisfy the rule's left hand side conditions and have adverse outcomes in the next year. For the intervention of opening new primary care clinics, this informs the new clinics' locations by maximizing the number of these patients living less than 15 miles away. A cost-benefit analysis can determine whether adopting this intervention is worthwhile.

Use Association Rules to Help Understand the Subtleties in the Data and Improve Model Accuracy

For each association rule related to high risk, the proportion of patients who are at high risk and satisfy the rule's left hand side is called the rule's support showing the rule's coverage. Among all patients fulfilling the rule's left hand side, the proportion of patients at high risk is called the rule's confidence showing the rule's accuracy. Our method discretizes each numerical feature into a categorical one, and mines rules exceeding some predefined minimum support s_1 and minimum confidence c_1 and containing only features that the predictive model uses to make predictions, no more than a preselected number n_1 of feature-value pairs on the left hand side and no feature-value pairs that the automatic explanation function's designers specify as unrelated to high risk.

Consider all of the association rules related to high risk and satisfying all conditions above except for the last one. If a feature-value pair is specified by the automatic explanation function's designers as unrelated to high risk but appears in many of these rules, the designers can examine the pair in detail and determine the following [84]:

1. Whether the pair is associated with a surrogate condition related to high risk. This helps us understand the subtleties in the data and how they affect machine learning. Sometimes, we can use the information to design new interventions targeting the surrogate condition. For instance,

XSL•FO
RenderX

suppose the pair is that the patient had two outpatient visits for asthma last year and the associated surrogate condition is noncompliance coupled with high vulnerability, for example, because of genetics or working environment. For each rule related to high risk and whose left hand side contains the pair and indicates the surrogate condition (eg, by mentioning that the patient had at least two hospitalizations for asthma last year), we keep the rule, inspect the patients satisfying the rule's left hand side, and arrange regular phone checks for some of them.

2. Whether the feature is uninformative. Retraining the predictive model after dropping the feature can possibly serve as a new way to improve model accuracy and make the model generalize better to other health care systems beyond the one in which it was developed. Ribeiro et al [88] showed that on nonclinical data, users of an automatic explanation function could use sparse linear model-based explanations to find uninformative features. Retraining the model after dropping these features improved model accuracy. We are unaware of any published work using rule-based explanations to do this, particularly on clinical data. As Ribeiro et al [89] stated, rule-based explanations are preferred over sparse linear model-based ones. The approach by Ribeiro et al [88] works for binary features only. In comparison, our approach can handle all kinds of features.

A health care system often has limited training data impacting model accuracy. To improve model accuracy, we can enlarge the training set by generating synthetic data instances:

1. Using historical data from the target or other source systems, we mine another set R_2 of association rules related to high risk. The clinicians in the automatic explanation function's design team check the rules in R_2 and keep only those making much clinical sense (eg, tending to generalize across different systems). If desired, we can remove additional rules from R_2 so that the remaining ones are not too similar to each other. For each remaining rule $r \in R_2$, we generate multiple synthetic data instances. Each synthetic data instance I_s satisfies the left hand side of r and is labeled high risk. For each feature not on the left hand side of r, the feature value of I_s is chosen randomly. For each numerical feature that our automatic explanation method discretizes into a categorical one, the numerical feature value of I_s is chosen randomly within the bounds of the category corresponding to the categorical feature value of I_s. Compared with those used for giving explanations, the rules in R_2 are required to exceed some predefined minimum confidence c_2 that is both larger than c_1 and close to 1 (eg, 90%), so that the synthetic data instances are likely to be correctly labeled. To help ensure R_2 contains enough rules, each rule in R_2 needs to exceed a lower predefined minimum support $s_2 < s_1$ and contains no more than a larger, preselected number $n_2 > n_1$ of feature-value pairs on the left hand side.

2. The clinicians specify some rules related to high risk based on medical knowledge. Each rule is used to generate

multiple synthetic data instances in a way similar to above. Alternatively, we can use these rules and the predictive model together at prediction time. We use these rules to identify a subset of high-risk patients and apply the predictive model to the other patients not satisfying the left hand side of any of these rules.

Using synthetic data instances to improve model accuracy has been done before, for example, for images [90] or via making interpolations among actual data instances [91]. We are unaware of any published work using association rules for this purpose. In contrast to interpolating all feature values of each synthetic data instance, our association rule-based method retains key feature values to increase the chance that the data instance is correctly labeled.

Expand Automatic Explanation's Coverage of Patients

The mined association rules R_1 used for giving explanations represent frequent patterns linked to high risk. Yet, certain patients are at high risk for uncommon reasons and not covered by any of these rules. To expand automatic explanation's coverage of patients, we improve our prior method [27,87] by generating synthetic data instances, adopting the predictive model to label them, and using them to mine additional rules to cover more patients [88]. The improved method generalizes to many clinical applications.

More specifically, during association rule mining, some rules are found and then removed because they fall below the predefined minimum support s_1 or minimum confidence c_1. Instead of removing them, we keep as backup all such rules R_3 that exceed both the minimum confidence c_1 and another predefined minimum support $s_3 < s_1$ and sort them in descending order of support. We can use techniques similar to those used in our prior method [27,87] to prune redundant rules in R_3. At prediction time, for each patient the predictive model identifies as high risk and not covered by any rule in the set R_1, we check the rules in the backup set R_3 sequentially. For each rule $r \in R_3$, we generate one or more synthetic data instances in a way similar to above to make the total number of data instances satisfying r's left hand side reach the minimum support s_1. We use the predictive model to make predictions on and label the synthetic data instances. Using both the synthetic data instances and data instances in the training set satisfying r's left hand side, we check whether r exceeds the minimum confidence c_1. If so, we stop the rule checking process and display r as the automatically generated explanation for the patient. Otherwise, we continue to check the next rule in R_3. The predictive model may make incorrect predictions on and mislabel some synthetic data instances, causing the finally chosen rule to not reflect the true reason why the patient is at high risk. By sorting the rules in R_3 in descending order of support, we minimize the number of synthetic data instances to be generated for the finally chosen rule and reduce this likelihood.

Unlike the rules in the set R_1, the rules in the backup set R_3 are not prechecked by the automatic explanation function's design team. Some rules in R_3 may make no or little clinical sense. At prediction time, users of the automatic explanation function can

provide feedback on the displayed rules chosen from R_3. This helps the automatic explanation function's design team identify unreasonable rules and remove them from R_3 so that they will not be displayed in the future. For example, if the number of times that a rule in R_3 has been displayed to users exceeds a given threshold and the proportion of times that users report the rule as unreasonable is over a fixed limit, the rule can become a candidate for removal from R_3.

Technique for Making Causal Inference for Periodically Reidentifying High-Risk Patients

To provide causal inference capability, we need to estimate the impact of care management on a patient's cost or health outcome. We use this estimate to adjust the cost or health outcome threshold for deciding whether a patient on care management should be moved off care management. Propensity score matching is one technique for doing this on observational data [80,81,92]. Using the same features adopted for predicting individual patient cost or health outcome, we build a model to predict whether a patient will be enrolled in care management. The propensity score is the predicted probability of enrollment. We match each patient on care management to a patient not on care management on propensity score. The impact of care management is estimated as the average cost or health outcome difference between the group of patients on care management and the matched group of patients not on care management. We can apply the propensity score matching technique to the entire group of patients. Alternatively, if we expect care management to have greatly varying impact across different subgroups of patients, we can apply the propensity score matching technique to each subgroup of patients separately.

Conclusions

Care management is broadly used for improving asthma outcomes and cutting costs, but current high-risk patient identification methods have major limitations degrading its effectiveness. This paper pinpoints these limitations and outlines multiple machine learning techniques to address them, offering a roadmap for future research. Besides being used for asthma, care management is also broadly adopted in managing patients with diabetes, heart diseases, or chronic obstructive pulmonary disease [5], where similar limitations in patient identification exist and techniques similar to those outlined in this paper can be used to optimize care management. The principles of many of our proposed techniques generalize to other predictive modeling tasks beyond those for care management.

Authors' Contributions

GL was mainly responsible for the paper. He performed the literature review, conceptualized the presentation approach, and drafted the manuscript. KAS gave feedback on various medical issues and revised the manuscript. Both authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1.   CDC. Asthma URL: http://www.cdc.gov/nchs/fastats/asthma.htm [accessed 2015-08-10] [WebCite Cache ID 6agaQMYxr]
2.   Akinbami LJ, Moorman JE, Liu X. Asthma prevalence, health care use, and mortality: United States, 2005-2009. Natl Health Stat Report 2011 Jan 12(32):1-14 [FREE Full text] [Medline: 21355352]
3.   Akinbami LJ, Moorman JE, Bailey C, Zahran HS, King M, Johnson CA, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. NCHS Data Brief 2012 May(94):1-8 [FREE Full text] [Medline: 22617340]
4.   CDC. Asthma in the US URL: https://www.cdc.gov/vitalsigns/asthma/ [accessed 2017-02-19] [WebCite Cache ID 6oOjKVf75]
5.   Duncan I. Healthcare Risk Adjustment and Predictive Modeling. Winsted, CT: ACTEX Publications Inc; 2011.
6.   Schatz M, Nakahiro R, Jones CH, Roth RM, Joshua A, Petitti D. Asthma population management: development and validation of a practical 3-level risk stratification scheme. Am J Manag Care 2004 Jan;10(1):25-32 [FREE Full text] [Medline: 14738184]
7.   Axelrod RC, Vogel D. Predictive modeling in health plans. Disease Management & Health Outcomes 2003;11(12):779-787. [doi: 10.2165/00115677-200311120-00003]

XSL•FO

RenderX

8.   Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med 2007 Dec;22(Suppl 3):391-395 [FREE Full text] [doi: 10.1007/s11606-007-0322-1] [Medline: 18026807]

9.   Caloyeras JP, Liu H, Exum E, Broderick M, Mattke S. Managing manifest diseases, but not health risks, saved PepsiCo money over seven years. Health Aff (Millwood) 2014 Jan;33(1):124-131. [doi: 10.1377/hlthaff.2013.0625] [Medline: 24395944]

10. Nelson L. CBO. Lessons from medicare's demonstration projects on disease management and care coordination URL: https://www.cbo.gov/sites/default/files/112th-congress-2011-2012/workingpaper/ WP2012-01_Nelson_Medicare_DMCC_Demonstrations_1.pdf [accessed 2015-08-10] [WebCite Cache ID 6agZcFxD1]

11. Ccmcertification. Definition and philosophy of case management URL: http://ccmcertification.org/about-us/ about-case-management/definition-and-philosophy-case-management [accessed 2015-08-10] [WebCite Cache ID 6agZuMc1N]

12. Kelly CS, Morrow AL, Shults J, Nakas N, Strope GL, Adelman RD. Outcomes evaluation of a comprehensive intervention program for asthmatic children enrolled in medicaid. Pediatrics 2000 May;105(5):1029-1035. [Medline: 10790458]

13. Forno E, Celedón JC. Predicting asthma exacerbations in children. Curr Opin Pulm Med 2012 Jan;18(1):63-69 [FREE Full text] [doi: 10.1097/MCP.0b013e32834db288] [Medline: 22081091]

14. Levine S, Adams J, Attaway K, Dorr D, Leung M, Popescu P, et al. CHCF. Predicting the financial risks of seriously ill patients URL: http://www.chcf.org/publications/2011/12/predictive-financial-risks [accessed 2017-02-19] [WebCite Cache ID 6oOjQCINu]

15. Rubin RJ, Dietrich KA, Hawk AD. Clinical and economic impact of implementing a comprehensive diabetes management program in managed care. J Clin Endocrinol Metab 1998 Aug;83(8):2635-2642. [doi: 10.1210/jcem.83.8.5075] [Medline: 9709924]

16. Greineder DK, Loane KC, Parks P. A randomized controlled trial of a pediatric asthma outreach program. J Allergy Clin Immunol 1999 Mar;103(3 Pt 1):436-440. [Medline: 10069877]

17. Axelrod R, Zimbro K, Chetney R, Sabol J, Ainsworth V. A disease management program utilizing life coaches for children with asthma. J Clin Outcomes Manag 2001;8(6):38-42 [FREE Full text]

18. Dorr DA, Wilcox AB, Brunker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. J Am Geriatr Soc 2008 Dec;56(12):2195-2202. [doi: 10.1111/j.1532-5415.2008.02005.x] [Medline: 19093919]

19. Beaulieu N, Cutler D, Ho K, Isham G, Lindquist T, Nelson A, et al. The business case for diabetes disease management for managed care organizations. Forum Health Econ Policy 2006;9(1):1-37.

20. Curry N, Billings J, Darin B, Dixon J, Williams M, Wennberg D. Kingsfund. Predictive risk project literature review URL: http://www.kingsfund.org.uk/sites/files/kf/field/field_document/predictive-risk-literature-review-june2005.pdf [accessed 2015-08-10] [WebCite Cache ID 6aga2XBFC]

21. Moturu S, Johnson W, Liu H. Predicting future high-cost patients: a real-world risk modeling application. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. 2007 Presented at: IEEE International Conference on Bioinformatics and Biomedicine; November 2-4, 2007; Silicon Valley, CA p. 202-208. [doi: 10.1109/BIBM.2007.54]

22. Fleishman JA, Cohen JW, Manning WG, Kosinski M. Using the SF-12 health status measure to improve predictions of medical expenditures. Med Care 2006 May;44(5 Suppl):I54-I63. [doi: 10.1097/01.mlr.0000208141.02083.86] [Medline: 16625065]

23. Bertsimas D, Bjarnadóttir MV, Kane MA, Kryder JC, Pandey R, Vempala S, et al. Algorithmic prediction of health-care costs. Operations Research 2008 Dec;56(6):1382-1392. [doi: 10.1287/opre.1080.0619]

24. Anderson RT, Balkrishnan R, Camacho F. Risk classification of Medicare HMO enrollee cost levels using a decision-tree approach. Am J Manag Care 2004 Feb;10(2 Pt 1):89-98 [FREE Full text] [Medline: 15011809]

25. Meenan RT, Goodman MJ, Fishman PA, Hornbrook MC, O'Keeffe-Rosetti MC, Bachman DJ. Using risk-adjustment models to identify high-cost risks. Med Care 2003 Nov;41(11):1301-1312. [doi: 10.1097/01.MLR.0000094480.13057.75] [Medline: 14583693]

26. Weir S, Aweh G, Clark RE. Case selection for a Medicaid chronic care management program. Health Care Financ Rev 2008;30(1):61-74 [FREE Full text] [Medline: 19040174]

27. Luo G, Stone BL, Sakaguchi F, Sheng X, Murtaugh MA. Using computational approaches to improve risk-stratified patient management: rationale and methods. JMIR Res Protoc 2015;4(4):e128 [FREE Full text] [doi: 10.2196/resprot.5039] [Medline: 26503357]

28. Hibbard JH, Greene J, Sacks RM, Overton V, Parrotta C. Improving population health management strategies: identifying patients who are more likely to be users of avoidable costly care and those more likely to develop a new chronic disease. Health Serv Res 2017 Aug;52(4):1297-1309. [doi: 10.1111/1475-6773.12545] [Medline: 27546032]

29. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar 1;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]

30. Fogg B. A behavior model for persuasive design. In: Proceedings of the International Conference on Persuasive Technology. 2009 Presented at: International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA p. 40. [doi: 10.1145/1541948.1541999]

31. Naylor MD, Bowles KH, McCauley KM, Maccoy MC, Maislin G, Pauly MV, et al. High-value transitional care: translation of research into practice. J Eval Clin Pract 2013 Oct;19(5):727-733. [doi: 10.1111/j.1365-2753.2011.01659.x] [Medline: 21410844]

32. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analyzing health care utilization and costs. Annu Rev Public Health 1999;20:125-144. [doi: 10.1146/annurev.publhealth.20.1.125] [Medline: 10352853]

33. Ash A, McCall N. RTI. Risk assessment of military populations to predict health care cost and utilization URL: http://www. rti.org/pubs/tricare_riskassessment_final_report_combined.pdf [accessed 2015-08-10] [WebCite Cache ID 6aga7wYZC]

34. Iezzoni L. Risk Adjustment for Measuring Health Care Outcomes. 4th edition. Chicago, IL: Health Administration Press; 2013.

35. Neuvirth H, Ozery-Flato M, Hu J, Laserson J, Kohn M, Ebadollahi S, et al. Toward personalized care management of patients at risk: the diabetes case study. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2011 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 21-24, 2011; San Diego, CA p. 395-403. [doi: 10.1145/2020408.2020472]

36. Schatz M, Cook EF, Joshua A, Petitti D. Risk factors for asthma hospitalizations in a managed care organization: development of a clinical prediction rule. Am J Manag Care 2003 Aug;9(8):538-547 [FREE Full text] [Medline: 12921231]

37. Lieu TA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high-risk children with asthma. Am J Respir Crit Care Med 1998 Apr;157(4 Pt 1):1173-1180. [doi: 10.1164/ajrccm.157.4.9708124] [Medline: 9563736]

38. Lieu TA, Capra AM, Quesenberry CP, Mendoza GR, Mazar M. Computer-based models to identify high-risk adults with asthma: is the glass half empty or half full? J Asthma 1999 Jun;36(4):359-370. [Medline: 10386500]

39. Forno E, Fuhlbrigge A, Soto-Quirós ME, Avila L, Raby BA, Brehm J, et al. Risk factors and predictive clinical scores for asthma exacerbations in childhood. Chest 2010 Nov;138(5):1156-1165 [FREE Full text] [doi: 10.1378/chest.09-2426] [Medline: 20472862]

40. Miller MK, Lee JH, Blanc PD, Pasta DJ, Gujrathi S, Barron H, et al. TENOR risk score predicts healthcare in adults with severe or difficult-to-treat asthma. Eur Respir J 2006 Dec;28(6):1145-1155 [FREE Full text] [doi: 10.1183/09031936.06.00145105] [Medline: 16870656]

41. Razavian N, Blecker S, Schmidt AM, Smith-McLallen A, Nigam S, Sontag D. Population-level prediction of type 2 diabetes from claims data and analysis of risk factors. Big Data 2015 Dec;3(4):277-287. [doi: 10.1089/big.2015.0020] [Medline: 27441408]

42. Schatz M. Predictors of asthma control: what can we modify? Curr Opin Allergy Clin Immunol 2012 Jun;12(3):263-268. [doi: 10.1097/ACI.0b013e32835335ac] [Medline: 22517290]

43. Stanford RH, Shah MB, D'Souza AO, Schatz M. Predicting asthma outcomes in commercially insured and Medicaid populations. Am J Manag Care 2013 Jan;19(1):60-67 [FREE Full text] [Medline: 23379745]

44. Hyland ME, Whalley B, Halpin DM, Greaves CJ, Seamark C, Blake S, et al. Frequency of non-asthma GP visits predicts asthma exacerbations: an observational study in general practice. Prim Care Respir J 2012 Dec;21(4):405-411 [FREE Full text] [doi: 10.4104/pcrj.2012.00061] [Medline: 22836742]

45. Crawford AG, Fuhr JP, Clarke J, Hubbs B. Comparative effectiveness of total population versus disease-specific neural network models in predicting medical costs. Dis Manag 2005 Oct;8(5):277-287. [doi: 10.1089/dis.2005.8.277] [Medline: 16212513]

46. Coyle YM. Predictors of acute asthma relapse: strategies for its prevention. J Asthma 2003 May;40(3):217-224. [Medline: 12807164]

47. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2016 May 20;Suppl 1:S48-S61. [doi: 10.15265/IYS-2016-s006] [Medline: 27199197]

48. Sun J, Hu J, Luo D, Markatou M, Wang F, Edabollahi S, et al. Combining knowledge and data driven insights for identifying risk factors using electronic health records. AMIA Annu Symp Proc 2012;2012:901-910 [FREE Full text] [Medline: 23304365]

49. Phillips CD, Chen M, Sherman M. To what degree does provider performance affect a quality indicator? the case of nursing homes and ADL change. Gerontologist 2008 Jun;48(3):330-337. [Medline: 18591358]

50. Fung V, Schmittdiel JA, Fireman B, Meer A, Thomas S, Smider N, et al. Meaningful variation in performance: a systematic literature review. Med Care 2010 Feb;48(2):140-148. [doi: 10.1097/MLR.0b013e3181bd4dc3] [Medline: 20057334]

51. Lipitz-Snyderman A, Sima CS, Atoria CL, Elkin EB, Anderson C, Blinder V, et al. Physician-driven variation in nonrecommended services among older adults diagnosed with cancer. JAMA Intern Med 2016 Oct 01;176(10):1541-1548. [doi: 10.1001/jamainternmed.2016.4426] [Medline: 27533635]

52. Obermeyer Z, Powers BW, Makar M, Keating NL, Cutler DM. Physician characteristics strongly predict patient enrollment in hospice. Health Aff (Millwood) 2015 Jun;34(6):993-1000 [FREE Full text] [doi: 10.1377/hlthaff.2014.1055] [Medline: 26056205]

53. Borsi JP. Hypothesis-free search for connections between birth month and disease prevalence in large, geographically varied cohorts. AMIA Annu Symp Proc 2016;2016:319-325 [FREE Full text] [Medline: 28269826]

54. Wiens J, Guttag J, Horvitz E. A study in transfer learning: leveraging data from multiple hospitals to enhance hospital-specific predictions. J Am Med Inform Assoc 2014;21(4):699-706 [FREE Full text] [doi: 10.1136/amiajnl-2013-002162] [Medline: 24481703]

55. Singleton KW, Hsu W, Bui AAT. Comparing predictive models of glioblastoma multiforme built using multi-institutional and local data sources. AMIA Annu Symp Proc 2012;2012:1385-1392 [FREE Full text] [Medline: 23304418]

56. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003 Sep;56(9):826-832. [Medline: 14505766]

57. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015 Jan;68(1):25-34. [doi: 10.1016/j.jclinepi.2014.09.007] [Medline: 25441703]

58. Gong J, Sundt T, Rawn J, Guttag J. Instance weighting for patient-specific risk stratification models. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 10-13, 2015; Sydney, NSW, Australia p. 369-378. [doi: 10.1145/2783258.2783397]

59. Lee G, Rubinfeld I, Syed Z. Adapting surgical models to individual hospitals using transfer learning. In: Proceedings of IEEE International Conference on Data Mining Workshops. 2012 Presented at: IEEE International Conference on Data Mining Workshops; December 10, 2012; Brussels, Belgium p. 57-63. [doi: 10.1109/ICDMW.2012.93]

60. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010 Oct;22(10):1345-1359. [doi: 10.1109/TKDE.2009.191]

61. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data 2016 May 28;3:9. [doi: 10.1186/s40537-016-0043-6]

62. Jayanthi A. Beckershospitalreview. Down the rabbit hole at epic: 9 key points from the users group meeting URL: http://www.beckershospitalreview.com/healthcare-information-technology/down-the-rabbit-hole-at-epic-8-key-points-from-the-users-group-meeting.html [accessed 2017-06-17] [WebCite Cache ID 6rI15hVXC]

63. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]

64. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]

65. Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York, NY: Springer; 2009.

66. Zhou Z. Ensemble Methods: Foundations and Algorithms. Boca Raton, FL: Chapman and Hall/CRC; 2012.

67. Gao J, Fan W, Jiang J, Han J. Knowledge transfer via multiple model local structure mapping. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2008 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 24-27, 2008; Las Vegas, NV p. 283-291. [doi: 10.1145/1401890.1401928]

68. Halamka JD. Early experiences with big data at an academic medical center. Health Aff (Millwood) 2014 Jul;33(7):1132-1138. [doi: 10.1377/hlthaff.2014.0031] [Medline: 25006138]

69. Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev 1956 Mar;63(2):81-97. [Medline: 13310704]

70. Luo G. A roadmap for designing a personalized search tool for individual healthcare providers. J Med Syst 2014 Feb;38(2):6. [doi: 10.1007/s10916-014-0006-4] [Medline: 24424432]

71. James BC, Savitz LA. How Intermountain trimmed health care costs through robust quality improvement efforts. Health Aff (Millwood) 2011 Jun;30(6):1185-1191 [FREE Full text] [doi: 10.1377/hlthaff.2011.0358] [Medline: 21596758]

72. Dartmouthatlas. The Dartmouth atlas of health care URL: http://www.dartmouthatlas.org/data/topic/topic.aspx?cat=21 [accessed 2015-08-10] [WebCite Cache ID 6agaBuCzH]

73. Pacific Business Group on Health. PBGH. Advancing physician performance measurement: using administrative data to assess physician quality and efficiency URL: http://www.pbgh.org/storage/documents/reports/PBGHP3Report_09-01-05final.pdf [accessed 2015-08-10] [WebCite Cache ID 6agaGjTgH]

74. Gifford E, Foster EM. Provider-level effects on psychiatric inpatient length of stay for youth with mental health and substance abuse disorders. Med Care 2008 Mar;46(3):240-246. [doi: 10.1097/MLR.0b013e318158aee7] [Medline: 18388838]

75. Kramer TL, Daniels AS, Zieman GL, Williams C, Dewan NA. Psychiatric practice variations in the diagnosis and treatment of major depression. Psychiatr Serv 2000 Mar;51(3):336-340. [Medline: 10686240]

76. Iwashyna TJ, Chang VW, Zhang JX, Christakis NA. The lack of effect of market structure on hospice use. Health Serv Res 2002 Dec;37(6):1531-1551 [FREE Full text] [Medline: 12546285]

XSL·FO
RenderX

77.   Barnett ML, Olenski AR, Jena AB. Opioid-prescribing patterns of emergency physicians and risk of long-term use. N Engl J Med 2017 Dec 16;376(7):663-673. [doi: 10.1056/NEJMsa1610524] [Medline: 28199807]

78.   Keating NL, Herrinton LJ, Zaslavsky AM, Liu L, Ayanian JZ. Variations in hospice use among cancer patients. J Natl Cancer Inst 2006 Aug 02;98(15):1053-1059. [doi: 10.1093/jnci/djj298] [Medline: 16882942]

79.   Orueta JF, García-Alvarez A, Grandes G, Nuño-Solinís R. The origin of variation in primary care process and outcome indicators: patients, professionals, centers, and health districts. Medicine (Baltimore) 2015 Aug;94(31):e1314 [FREE Full text] [doi: 10.1097/MD.0000000000001314] [Medline: 26252315]

80.   Morgan S, Winship C. Counterfactuals and Causal Inference: Methods and Principles for Social Research. 2nd edition. New York, NY: Cambridge University Press; 2014.

81.   Pearl J. Causality: Models, Reasoning, and Inference. 2nd edition. New York, NY: Cambridge University Press; 2009.

82.   Luo D, Wang F, Sun J, Markatou M, Hu J, Ebadollahi S. SOR: Scalable orthogonal regression for low-redundancy feature selection and its healthcare applications. In: Proceedings of the SIAM International Conference on Data Mining. 2012 Presented at: SIAM International Conference on Data Mining; April 26-28, 2012; Anaheim, CA p. 576-587. [doi: 10.1137/1.9781611972825.50]

83.   Hall ES, Thornton SN. Generating nurse profiles from computerized labor and delivery documentation. AMIA Annu Symp Proc 2008;2008:268-272 [FREE Full text] [Medline: 18998796]

84.   Schuit E, Groenwold RH, Harrell FE, de Kort WL, Kwee A, Mol BW, et al. Unexpected predictor-outcome associations in clinical prediction research: causes and solutions. CMAJ 2013 Jul 09;185(10):E499-E505 [FREE Full text] [doi: 10.1503/cmaj.120812] [Medline: 23339155]

85.   OMOP. Observational medical outcomes partnership (OMOP) common data model homepage URL: http://omop.org/CDM [accessed 2015-08-10] [WebCite Cache ID 6agamjByZ]

86.   OMOP. Observational medical outcomes partnership (OMOP) vocabularies URL: http://omop.org/Vocabularies [accessed 2017-02-19] [WebCite Cache ID 6oOkOnGg5]

87.   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]

88.   Ribeiro M, Singh S, Guestrin C. "Why Should I Trust You?": explaining the predictions of any classifier. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA p. 1135-1144. [doi: 10.1145/2939672.2939778]

89.   Ribeiro M, Singh S, Guestrin C. Nothing else matters: model-agnostic explanations by identifying prediction invariance. In: Proceedings of NIPS Workshop on Interpretable Machine Learning in Complex Systems. 2016 Presented at: NIPS Workshop on Interpretable Machine Learning in Complex Systems; December 9, 2016; Barcelona, Spain.

90.   Nonnemaker J, Baird H. Using synthetic data safely in classification. In: Proceedings of Document Recognition and Retrieval XVI of the IS&T-SPIE Electronic Imaging Symposium. 2009 Presented at: Document Recognition and Retrieval XVI of the IS&T-SPIE Electronic Imaging Symposium; January 20-22, 2009; San Jose, CA p. 72470G. [doi: 10.1117/12.805619]

91.   Chawla N, Bowyer K, Hall L, Kegelmeyer W. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321-357. [doi: 10.1613/jair.953]

92.   Duncan I. Managing and Evaluating Healthcare Intervention Programs. 2nd edition. Winsted, CT: ACTEX Publications; 2014.

Abbreviations

EMR: electronic medical record

XSL•FO

RenderX