JMIR Medical Informatics

Impact Factor (2022): 3.2
Volume 7 (2019), Issue 2 ISSN 2291-9694 Editor in Chief: Christian Lovis, MD, MPH, FACMI

Contents

Reviews

The Use of Artificially Intelligent Self-Diagnosing Digital Platforms by the General Public: Scoping Review (e13445)
Stephanie Aboueid, Rebecca Liu, Binyam Desta, Ashok Chaurasia, Shanil Ebrahim. . . . 4

Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review (e12239)
Seyedmostafa Sheikhalishahi, Riccardo Miotto, Joel Dudley, Alberto Lavelli, Fabio Rinaldi, Venet Osmani. . . . 14

Toward Standardized Monitoring of Patients With Chronic Diseases in Primary Care Using Electronic Medical Records: Systematic Review (e10879)
Leandra Falck, Marco Zoller, Thomas Rosemann, Nahara Martínez-González, Corinne Chmiel. . . . 136

Interoperable Electronic Health Records and Health Information Exchanges: Systematic Review (e12607)
Mark Dobrow, Jessica Bytautas, Sukirtha Tharmalingam, Simon Hagens. . . . 172

Original Papers

Physician Use of Electronic Health Records: Survey Study Assessing Factors Associated With Provider Reported Satisfaction and Perceived Patient Impact (e10949)
Daniel Williams, Robert Warren, Myla Ebeling, Annie Andrews, Ronald Teufel II. . . . 32

Clinical Requirements of Future Patient Monitoring in the Intensive Care Unit: Qualitative Study (e13064)
Akira-Sebastian Poncette, Claudia Spies, Lina Mosch, Monique Schieler, Steffen Weber-Carstens, Henning Krampe, Felix Balzer. . . . 44

Facility and Regional Factors Associated With the New Adoption of Electronic Medical Records in Japan: Nationwide Longitudinal Observational Study (e14026)
Hideaki Kawaguchi, Soichi Koike, Kazuhiko Ohe. . . . 56

Adapting State-of-the-Art Deep Language Models to Clinical Information Extraction Systems: Potentials, Challenges, and Solutions (e11499)
Liyuan Zhou, Hanna Suominen, Tom Gedeon. . . . 66

Natural Language Processing for the Identification of Silent Brain Infarcts From Neuroimaging Reports (e12109)
Sunyang Fu, Lester Leung, Yanshan Wang, Anne-Olivia Raulli, David Kallmes, Kristin Kinsman, Kristoff Nelson, Michael Clark, Patrick Luetmer, Paul Kingsbury, David Kent, Hongfang Liu. . . . 82

Structure and Content of Drug Monitoring Advices Included in Discharge Letters at Interfaces of Care: Exploratory Analysis Preceding Database Development (e10832)
Benedict Morath, Katharina Wien, Torsten Hoppe-Tichy, Walter Haefeli, Hanna Seidling. 91

A Computer Application to Predict Adverse Events in the Short-Term Evolution of Patients With Exacerbation of Chronic Obstructive Pulmonary Disease (e10773)
Inmaculada Arostegui, María Legarreta, Irantzu Barrio, Cristobal Esteban, Susana Garcia-Gutierrez, Urko Aguirre, José Quintana, IRYSS-COPD Group. 102

Exploring the Impact of the Prescription Automatic Screening System in Health Care Services: Quasi-Experiment (e11663)
Yan Li, Xitong Guo, Carol Hsu, Xiaoxiao Liu, Doug Vogel. 114

Understanding of and Barriers to Electronic Health Record Patient Portal Access in a Culturally Diverse Pediatric Population (e11570)
Daniel Miklin, Sameera Vangara, Alan Delamater, Kenneth Goodman. 127

Incorporating Social Determinants of Health in Electronic Health Records: Qualitative Study of Current Practices Among Top Vendors (e13849)
Maysoun Freij, Prashila Dullabh, Sarah Lewis, Scott Smith, Lauren Hovey, Rina Dhopeshwarkar. 149

Patient-Sharing Relations in the Treatment of Diabetes and Their Implications for Health Information Exchange: Claims-Based Analysis (e12172)
Georg Duftschmid, Christoph Rinner, Simone Sauter, Gottfried Endel, Peter Klimek, Christoph Mitsch, Harald Heinzl. 161

Establishment of a New Platform for the Management of Patients After Cardiac Surgery: Descriptive Study (e13123)
Zhihui Zhu, Haibo Zhang, Yuehuan Li, Xu Meng, Yuqing Jiao, Ying Qin, Yan Li, Jie Han, Kun Liu, Chongyang Liu. 182

Identifying Clinical Terms in Medical Text Using Ontology-Guided Machine Learning (e12596)
Aryan Arbabi, David Adams, Sanja Fidler, Michael Brudno. 190

Fast Healthcare Interoperability Resources, Clinical Quality Language, and Systematized Nomenclature of Medicine—Clinical Terms in Representing Clinical Evidence Logic Statements for the Use of Imaging Procedures: Descriptive Study (e13590)
Eseosa Odigie, Ronilda Lacson, Ali Raja, David Osterbur, Ivan Ip, Louise Schneider, Ramin Khorasani. 205

Development of a Consumer Health Vocabulary by Mining Health Forum Texts Based on Word Embedding: Semiautomatic Approach (e12704)
Gen Gu, Xingting Zhang, Xingeng Zhu, Zhe Jian, Ken Chen, Dong Wen, Li Gao, Shaodian Zhang, Fei Wang, Handong Ma, Jianbo Lei. 215

Use of Telemedicine to Screen Patients in the Emergency Department: Matched Cohort Study Evaluating Efficiency and Patient Safety of Telemedicine (e11233)
Nicholas Rademacher, Gai Cole, Kevin Psoter, Gabor Kelen, Jamie Fan, Dennis Gordon, Junaid Razzak. 224

Use of Electronic Health Records to Develop and Implement a Silent Best Practice Alert Notification System for Patient Recruitment in Clinical Research: Quality Improvement Initiative (e10020)
Connor Devoe, Harriett Gabbidon, Nina Schussler, Lauren Cortese, Emily Caplan, Colin Gorman, Kamal Jethwani, Joseph Kvedar, Stephen Agboola. 235

Privacy-Preserving Analysis of Distributed Biomedical Data: Designing Efficient and Secure Multiparty Computations Using Distributed Statistical Learning Theory (e12702)
Fida Dankar, Nisha Madathil, Samar Dankar, Sabri Boughorbel. 243

Medication Adherence Prediction Through Online Social Forums: A Case Study of Fibromyalgia (e12561)
Kyle Haas, Zina Ben Miled, Malika Mahoui. 274

A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study (e13167)
Hervé Spechbach, Johanna Gerlach, Sanae Mazouri Karker, Nikos Tsourakis, Christophe Combescure, Pierrette Bouillon. 284

Viewpoint

Consumer-Mediated Data Exchange for Research: Current State of US Law, Technology, and Trust (e12348)
Yiscah Bracha, Jacqueline Bagwell, Robert Furberg, Jonathan Wald. 256

Review

The Use of Artificially Intelligent Self-Diagnosing Digital Platforms by the General Public: Scoping Review

Stephanie Aboueid[1], MSc, RD; Rebecca H Liu[2], MSc, PhD; Binyam Negussie Desta[1], MPH; Ashok Chaurasia[1], PhD; Shanil Ebrahim[3,4], PhD

[1]Applied Health Sciences, University of Waterloo, Waterloo, ON, Canada

[2]Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada

[3]Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[4]Meta-Research Innovation Center at Stanford, Stanford University, Santa Clara, CA, United States

Corresponding Author:
Stephanie Aboueid, MSc, RD
Applied Health Sciences
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G5
Canada
Phone: 1 6134061899
Email: seaboueid@uwaterloo.ca

Abstract

Background: Self-diagnosis is the process of diagnosing or identifying a medical condition in oneself. Artificially intelligent digital platforms for self-diagnosis are becoming widely available and are used by the general public; however, little is known about the body of knowledge surrounding this technology.

Objective: The objectives of this scoping review were to (1) systematically map the extent and nature of the literature and topic areas pertaining to digital platforms that use computerized algorithms to provide users with a list of potential diagnoses and (2) identify key knowledge gaps.

Methods: The following databases were searched: PubMed (Medline), Scopus, Association for Computing Machinery Digital Library, Institute of Electrical and Electronics Engineers, Google Scholar, Open Grey, and ProQuest Dissertations and Theses. The search strategy was developed and refined with the assistance of a librarian and consisted of 3 main concepts: (1) self-diagnosis; (2) digital platforms; and (3) public or patients. The search generated 2536 articles from which 217 were duplicates. Following the Tricco et al 2018 checklist, 2 researchers screened the titles and abstracts (n=2316) and full texts (n=104), independently. A total of 19 articles were included for review, and data were retrieved following a data-charting form that was pretested by the research team.

Results: The included articles were mainly conducted in the United States (n=10) or the United Kingdom (n=4). Among the articles, topic areas included accuracy or correspondence with a doctor's diagnosis (n=6), commentaries (n=2), regulation (n=3), sociological (n=2), user experience (n=2), theoretical (n=1), privacy and security (n=1), ethical (n=1), and design (n=1). Individuals who do not have access to health care and perceive to have a stigmatizing condition are more likely to use this technology. The accuracy of this technology varied substantially based on the disease examined and platform used. Women and those with higher education were more likely to choose the right diagnosis out of the potential list of diagnoses. Regulation of this technology is lacking in most parts of the world; however, they are currently under development.

Conclusions: There are prominent research gaps in the literature surrounding the use of artificially intelligent self-diagnosing digital platforms. Given the variety of digital platforms and the wide array of diseases they cover, measuring accuracy is cumbersome. More research is needed to understand the user experience and inform regulations.

(JMIR Med Inform 2019;7(2):e13445) doi:10.2196/13445

KEYWORDS

diagnosis; artificial intelligence; symptom checkers; diagnostic self evaluation; self-care

Introduction

Background

Researching health information on the internet has become common practice by the general public [1-3]. Those who do not have access to health care services are more likely to use the internet for health information [4]. In some cases, browsing the internet for health information can have certain benefits such as improving health outcomes by increasing the availability of information, providing social support, and improving self-efficacy [5,6]. However, potential negative consequences still exist; the information may not be reliable, and the individual seeking information may have low health literacy [6]. For example, an individual may not be able to critically analyze the health information and assess the applicability of the information to their case, which could result in detrimental effects on their health [6]. Therefore, health information widely circulated on the internet should be interpreted with caution [7].

Significant technological advances have resulted in the rise of more sophisticated digital health platforms, which could potentially mitigate this issue, especially those involving artificial intelligence (AI). Interest in AI appears to be relatively recent; however, the term dates back to the 1950s and is described as the theory and development of computer systems that can perform tasks that would normally require human intelligence [8,9]. Notably, AI has become incorporated in computerized diagnostic decision support systems, which were initially developed for health professionals. These platforms have now become readily available to the general public and are known as *self-diagnosing apps* or *symptom checkers*, which include the Mayo Clinic symptom checker, Babylon Health, the Ada health app, and the K Health app. On the basis of the medical information and symptoms provided by an individual, these digital platforms perform 2 main functions: (1) provide individuals with a list of potential diagnoses and (2) assist with triage [10]. While the accuracy of symptom checkers is still under question [11,12], this technology has been gaining traction globally [13,14] owing to its potential in addressing the lack of access to primary care providers (PCPs) and unnecessary medical visits—prominent issues in Canada and most parts of the world [15-18].

Objectives

Although accuracy is important to consider, it is of equal importance to understand the overall body of knowledge that surrounds this technology, including legal and ethical implications and user experiences. In light of this, it is imperative to systematically map the literature available on artificially intelligent self-diagnosing digital platforms to identify the areas of research pertaining to this topic and to outline the key gaps in knowledge. This information can support the growing interest in leveraging AI technology in health care systems. As such, this scoping review aimed to answer the following question: What is known about the use of artificially intelligent self-diagnosing digital platforms by the general public and what are the main knowledge gaps in the literature?

Methods

Eligibility Criteria

In this review, self-diagnosing digital platforms were defined as platforms that utilize algorithms to provide a list of potential diagnoses to the user based on the medical information and symptoms provided. Although this scoping review does not entail quality assessment, it follows a sound methodological approach to map out the results in a concise manner for knowledge users. This scoping review follows the 2018 checklist developed by Tricco et al [19] for reporting scoping reviews. Ethics approval was not required.

The 3 main overarching concepts that guided this search were (1) self-diagnosis; (2) digital platforms; and (3) public or patients. Given the relatively new emergence of this technology and its use by the general public, the search was not limited by a publication date. Articles that were included in the review were those that (1) pertained to the use of self-diagnosing digital platforms by the lay public or patients and (2) were written in English or French. Exclusion criteria were articles that (1) focused on the use of self-diagnosing AI technology by health professionals; (2) described the back-end development of a self-diagnosing platform (eg, neural networks and architecture); (3) focused on digital health platforms that provide general health information, advice for disease management or triage; (4) focused on a tool that entails a validated questionnaire rather than an algorithm; and/or (5) examined test kits or digital platforms requiring an image upload. To allow for a wide array of results to be included, quantitative, qualitative, and mixed-methods studies or reports were eligible for inclusion.

Information Sources and Search

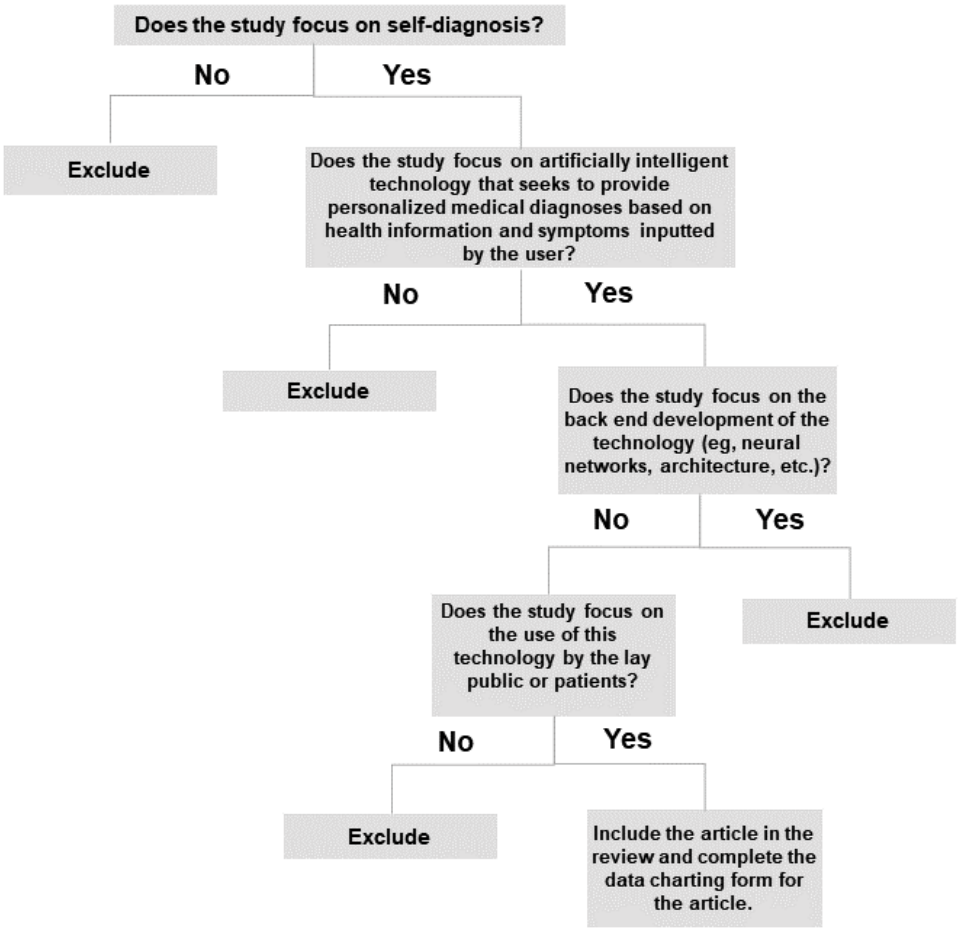
This scoping review systematically searched citation databases and the gray literature for relevant published and unpublished articles. The citation databases included PubMed (Medline), Scopus, Association for Computing Machinery Digital Library, Institute of Electrical and Electronics Engineers, and Google Scholar. To supplement the gray literature retrieved through Google Scholar [20], OpenGrey and ProQuest Dissertations and Theses were also searched. The final search strategy for each data source was defined and refined with the assistance of a librarian (Rebecca Hutchinson, University of Waterloo) and was finalized on November 19, 2018. The final search strategy for PubMed (Medline) can be found in Multimedia Appendix 1. The final search results were exported into RefWorks for screening.

Selection of Sources of Evidence

Once duplicates were removed in RefWorks, the screening process was conducted independently by 2 researchers (SA and RHL). The decision tree in Figure 1 was used as a guide to screen titles and abstracts (or executive summaries for reports and commentaries). Articles that were extracted from the title and abstract screening stage were read in their entirety (full-text review). For the full-text screening step, 2 researchers (SA and RHL) screened the same 30 articles to assess inter-rater reliability. Any uncertainty and disagreements were discussed and resolved through consensus. Following full-text review,

the reference lists of eligible articles were systematically screened. Similarly, for any review paper screened at the full-text review stage, references were screened for potentially relevant articles meeting the inclusion criteria.

Figure 1. Decision tree for assessing article eligibility.



Data Charting Process

Once the final number of articles was determined, a scan through these articles allowed the research team to gain a high-level understanding of the topics of interest in which self-diagnosing digital platforms were being examined (eg, accuracy and regulatory concerns). This allowed for the development of a data-charting form that captured all the relevant information, irrespective of the article type (eg, clinical trial or a qualitative study on user experience). The data-charting form was pretested with the same 5 articles to assess consistency. No changes were made to the form following this exercise.

Data Extraction

The variables collected through the data-charting form included the following: country, year of publication, main objective, the main area of study (eg, clinical, legal, and ethical), study design, data sources used (if any), target population (if any), sample size and sample characteristics (if any), methods/statistical analyses (if applicable), main findings, and study limitations (if applicable).

Synthesis of Results

Scoping reviews provide knowledge users with a concise overview on the literature available on a given topic of interest [21]. Given the heterogeneity of the studies included in this review, studies were grouped based on a specific area of study. A concept map was used to illustrate the breadth of studies surrounding self-diagnosing AI technology. Tables were used to provide an overview on the types of articles found in the literature and the data extracted from each article. A thematic synthesis was used to outline the knowledge gaps in the literature and other key considerations.

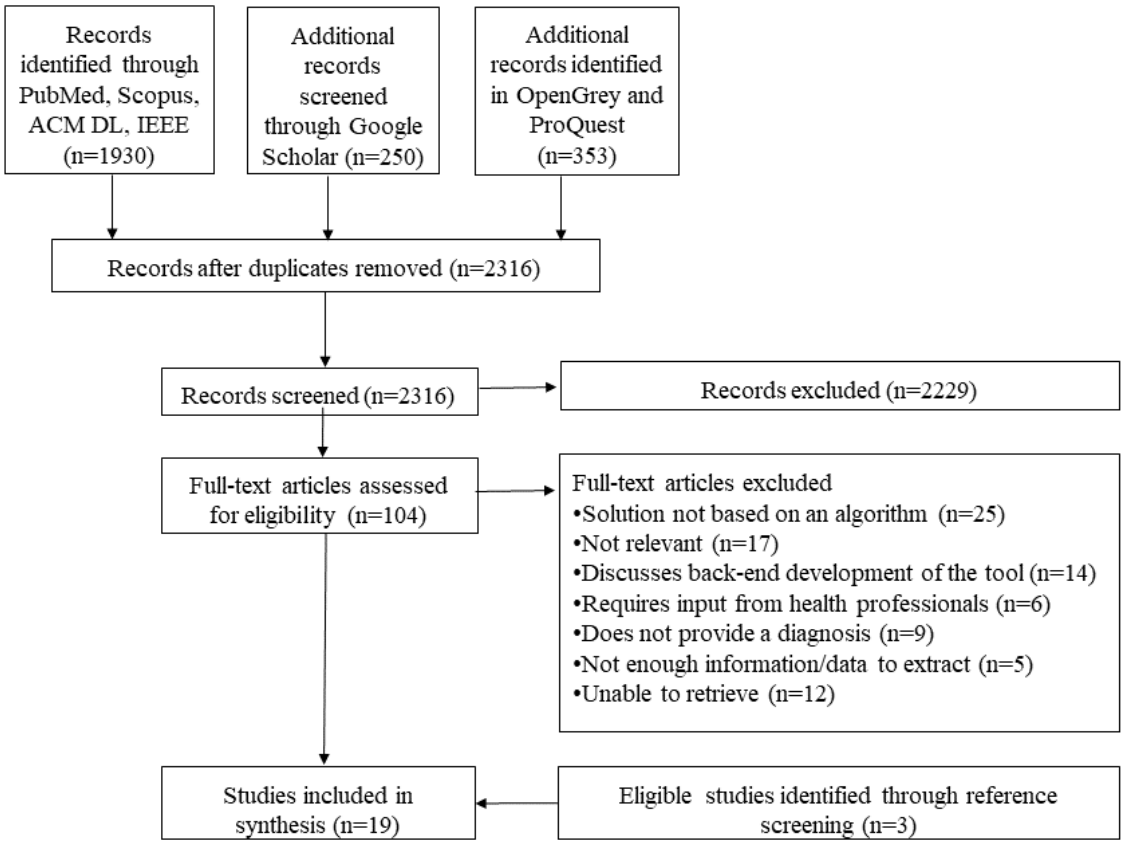
Results

Selection of Sources of Evidence

Figure 2 depicts the flow chart, which illustrates the selection process at each screening step. Our search identified a total of 2536 from which 217 were duplicates. In addition, 2 researchers independently screened the titles and abstracts of 2316 articles from which 2229 were excluded based on relevance and eligibility criteria. A total of 104 full-text articles were retrieved and assessed for eligibility. Of these, 76 articles were excluded for the following reasons: described the back-end development of the digital platform or the algorithm, examined the use of digitized questionnaires rather than algorithm-based digital platforms, the digital platform required the input of health professionals, provided the risk of disease, monitored symptoms,

technology designed for health professionals, not in scope, and did not provide enough data or information. We excluded 12 additional articles because we were unable to retrieve them. Through reference screening of the included articles, we identified 17 potentially relevant articles from which 3 articles were included in the review. A total of 19 articles were considered eligible for this review. Inter-rater reliability was assessed at the full-text stage which resulted in a score of 0.82, an almost perfect agreement score, between the 2 reviewers (SA and RHL) [22,23].

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of included articles. ACM DL: Association for Computing Machinery Digital Library; IEEE: Institute of Electrical and Electronics Engineers.



Characteristics of Sources of Evidence

The concept map in Figure 3 provides an illustrative overview of the main topic areas surrounding the use of artificially intelligent self-diagnosing digital platforms by the general public. The articles were mainly conducted in the United States (n=10) or the United Kingdom (n=4). In total, 2 of the articles were commentaries and the rest focused on the following areas: accuracy or correspondence with a doctor's diagnosis, regulation, sociological perspectives, experience, theory, privacy and security, ethics, and design. The concept map also outlines the main themes that emerged from the articles and the health conditions examined.

Figure 3. Concept map of the literature surrounding the use of artificially intelligent self-diagnosing digital platforms by the general public. DCM: degenerative cervical myelopathy; ENT: ear, nose, and throat; FDA: Food and Drug Administration; HIPAA: Health Insurance Portability and Accountability Act; NHS: National Health Service.

Results of Individual Sources of Evidence

Multimedia Appendix 2 provides an overview of all included articles and outlines the following variables: the article type, topic area examined, main objective, and main findings [24-42].

Synthesis of Results

Table 1 provides additional information on studies that entailed participant recruitment to answer their research question. These articles tended to focus on accuracy of the digital platform or user experience.

Table 1. Synthesis of results of studies with participants.

First author, year, reference, country	Sample size (n)	Target population	Data collection	Digital platforms used	Methods
Bisson, 2014 [26], United States	572	Individuals with knee pain	Primary data collection from patients and electronic medical records (EMRs)	A Web-based program developed by the research team	Sensitivity and specificity of the program's ability to provide a correct diagnosis for knee pain was tested, out of a possible 21 conditions in which the algorithm was trained to diagnose
Bisson, 2016 [27], United States	328	Individuals with knee pain	Primary data collection from patients and EMRs	A Web-based program developed by the research team	Sensitivity and specificity were calculated
Copeland, 2018 [29], United States	13	Users who tested the protocol (specifics not provided)	Primary data collection using the System Usability Scale and the Usability Metric for User Experience	Prototype developed by the research team	Descriptive statistics
Farmer, 2011 [32], United Kingdom	61	Patients coming in to the Ear, Nose, Throat surgeon's office	Primary data collected from patients over 1 month	Boots WebMD Symptom	Not provided
Hageman, 2014 [33], United States	86	Patients coming in to an outpatient hand and upper extremity surgeon's office	Primary data collection from patients and physicians	WebMD Symptom Checker	The Pearson chi-square test was used to determine the level of correspondence of the provided diagnosis by the diagnostic application and the final diagnosis of the physician
Lanseng, 2007 [36], Norway	160	Individuals between the ages of 18 and 65 years	Primary data collection using the Technology Readiness Survey (TRI)	N/A[a]	A survey with an internet - based medical self - diagnosis application as the focal technology was conducted; The research hypotheses were tested by completing a scenario and then following-up with a questionnaire
Luger, 2014 [37], United States	79	Older adults (aged 50 years or older)	Primary data collection of think-aloud protocols	WebMD Symptom Checker	Participants received one of 2 vignettes that depicted symptoms of illness. Participants talked out loud about their thoughts and actions while attempting to diagnose the symptoms with and without the help of common internet tools (Google and WebMD's Symptom Checker); Think-aloud content of participants was then compared with those who were accurate in their diagnosis versus those who were not.
Powley, 2016 [40], United Kingdom	34	Consecutive patients with newly presenting clinically apparent synovitis or a new onset of symptoms consistent with inflammatory arthritis	Primary data collection from patients	National Health Service (NHS) and WebMD Symptom Checkers	Patients were asked questions about their internet use in relation to their presenting symptoms. Subsequently, they completed the NHS and the WebMD symptom checkers and their answers as well as outcomes were recorded.

[a]Not applicable.

Discussion

Summary of Evidence and Knowledge Gaps

In this scoping review, 19 articles were included that examined artificially intelligent self-diagnosing digital platforms from various perspectives. Despite the popularity and accessibility of self-diagnosing AI technology by the public, it is noteworthy that research examining the accuracy of these platforms is limited. As such, it is unclear whether these platforms hinder or improve the health of users. Although some argue that the use of this technology may cause an individual to delay seeking care, it is important to recognize that delayed diagnoses are prevalent even without the use of this technology [40,42,43]. Many factors contribute to a delayed diagnosis with the top-ranked issues being poor communication between secondary and primary care, a mismatch between patients' medical needs and health care supply, and a lack of access or use of health services [42,44]. For example, Behrbalk et al found that the average time delay from initiation of symptoms to the diagnosis of cervical spondylotic myelopathy (CSM) was 2.2 (SD 2.3) years [43]. Although symptom checkers can potentially address delayed diagnoses, a review showed that this technology was suboptimal in diagnosing CSM [30].

Moreover, these platforms generally provide a list of potential diagnoses rather than a single diagnosis. In this case, the user must decide which condition describes their current state best. The likelihood of a user to accurately choose the right diagnosis is associated with the sociodemographic profile/variables of a user, such as education and gender [33]. For example, women and those with higher education were more likely to choose the correct diagnosis [33]. Therefore, although having a timely diagnosis is important, it may be counterproductive if the user considers the wrong treatment options owing to a misdiagnosis. Moreover, the patient may still require a visit to a PCP to receive treatment or a prescription. Issues may arise if patients already have a diagnosis in mind when visiting their PCP as it could translate into disagreements regarding their condition.

This scoping review suggests that there are prominent knowledge gaps in the literature; as such, a systematic review may not be worthwhile on this topic. Rather, concerted efforts are needed in producing research in this area related to accuracy, user experience, regulation, doctor-patient relationship, PCP perspectives, and ethics. Specifically, extensive research is needed in evaluating the accuracy of this technology while accounting for the fact that some platforms are designed for a wide area of conditions and others are specialized—as such, these platforms need to be evaluated accordingly. It is also important to distinguish the difference between accuracy and correspondence with a PCP's diagnosis as PCPs may misdiagnose or miss a diagnosis [45-47]. Importantly, when developing self-diagnosing AI digital platforms, it is important to test them on users with a wide range of backgrounds and level of experience with technology. This will ensure that a high proportion of users will end up choosing the right diagnosis.

Along with the importance of accuracy in self-diagnosing applications, there also needs to be guidance on how these platforms should be regulated. Although regulations related to self-diagnosing AI technologies should focus on patient safety as well as privacy and security, they should not hinder innovation in this area; rather, they should allow innovative advancements that are safe and improve access to timely diagnosis. Overall, more knowledge is needed on how different types of users interact with this technology and how its use can impact the PCP-patient relationship. There is also a need for clarity on data management shared by users. Ethical concerns surrounding the digital economy is a main area of concern, and there is currently a debate surrounding the trade-offs pertaining to the use of these platforms.

Limitations

Some limitations of this scoping review warrant mention. Artificially intelligent self-diagnosing platforms that require individuals to upload an image or a scan were excluded from the review. Test kits or platforms that would require the user to perform medical tests were also excluded. Our scoping review's focus was on platforms that required the least amount of effort from the user (ie, simply entering their symptoms into the platform to obtain potential diagnoses). It is also possible that some potentially relevant articles were missed because they could not be retrieved. To counteract this limitation, the authors systematically reviewed the references of relevant articles and held multiple meetings to assess consistency and to discuss any discrepancies in the screening process.

Conclusions

Given self-diagnosing AI technology's potential, it is worth understanding how it can be leveraged by health care systems to reduce costs and unnecessary medical visits. This scoping review aimed to map the literature surrounding the use of artificially intelligent self-diagnosing platforms. Given the direct-to-consumer approach of these platforms, it is worrisome that only a few studies have focused on the use of this technology. It is important that future research and resources are directed to understanding the accuracy and regulation of self-diagnosing AI digital platforms. These regulations may take different forms such as creating an application library which includes a list of platforms that have been deemed safe and provide highly accurate diagnoses from a credible health agency or organization. It should be noted that patient engagement is necessary in the development of these platforms to ensure that they allow a high proportion of individuals—irrespective of gender and education—to choose the right diagnosis. Importantly, user experience is crucial to consider as the public may be skeptical of this technology.

Acknowledgments

The authors thank Rebecca Hutchinson, librarian at the University of Waterloo, for assisting in creating the search strategy. This study was funded by the Canadian Institutes of Health Research. The funder had no role in study design, study selection, data synthesis, decision to publish, or preparation of the manuscript.

Authors' Contributions

SA conceived and designed the review. RHL, BND, AC, and SE participated in the conception of the study. SA, RHL, and BND carried out the review. All authors read and improved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy for PubMed (Medline).

[PDF File (Adobe PDF File), 57KB - medinform_v7i2e13445_app1.pdf]

Multimedia Appendix 2

Overview of included studies related to self-diagnosing artificial intelligence digital platforms.

[PDF File (Adobe PDF File), 103KB - medinform_v7i2e13445_app2.pdf]

References

1. Statistics Canada. 2009 Canadian Internet use survey URL: https://www150.statcan.gc.ca/n1/daily-quotidien/100510/dq100510a-eng.htm [accessed 2019-04-06] [WebCite Cache ID 77RGfEkhd]
2. Beck F, Richard JB, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
3. Pew Research Centre. Health Online 2013 URL: https://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-04-06] [WebCite Cache ID 77RGnsxqp]
4. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]
5. 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]
6. Tonsaker T, Bartlett G, Trpkov C. Health information on the internet: gold mine or minefield? Can Fam Physician 2014 May;60(5):407-408 [FREE Full text] [Medline: 24828994]
7. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
8. Turing A. Computing machinery and intelligence. Mind 1950;49:433-460 [FREE Full text]
9. Senate Canada. Standing Senate Committee on Social Affairs, Science and Technology. 2017. Challenge ahead: Integrating Robotics, Artificial Intelligence and 3D Printing Technologies into Canada?s Healthcare Systems URL: https://sencanada.ca/content/sen/committee/421/SOCI/reports/RoboticsAI3DFinal_Web_e.pdf [accessed 2019-04-06] [WebCite Cache ID 77RH6QKQ0]
10. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. Br Med J 2015;351:h3480 [FREE Full text] [Medline: 26157077]
11. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018 Dec 24;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]
12. Millenson ML, Baldwin JL, Zipperer L, Singh H. Beyond Dr Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis (Berl) 2018 Dec 25;5(3):95-105. [doi: 10.1515/dx-2018-0009] [Medline: 30032130]
13. Business Insider. 2018. An AI startup that claimed it can beat doctors in an exam is putting $100 million into creating 500 new jobs URL: https://www.businessinsider.com/babylon-health-invests-100-million-500-new-jobs-ai-research-2018-9 [accessed 2019-04-06] [WebCite Cache ID 77RH8Q52O]
14. Digital Health. 2018. Babylon goes to Canada with Telus Health partnership URL: https://www.digitalhealth.net/2018/09/babylon-canadian-partnership/ [accessed 2019-04-06] [WebCite Cache ID 77RHHIlTb]

15. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. Lancet 2017 Dec 08;390(10090):156-168 [FREE Full text] [doi: 10.1016/S0140-6736(16)32585-5] [Medline: 28077234]
16. Morgan DJ, Dhruva SS, Wright SM, Korenstein D. 2016 update on medical overuse: a systematic review. JAMA Intern Med 2016 Dec 01;176(11):1687-1692 [FREE Full text] [doi: 10.1001/jamainternmed.2016.5381] [Medline: 27654002]
17. Canadian Institute of Health Information. 2017. Unnecessary Care in Canada URL: https://www.cihi.ca/sites/default/files/document/choosing-wisely-baseline-report-en-web.pdf [accessed 2019-04-06] [WebCite Cache ID 77RHLS5ur]
18. Organisation for Economic Co-operation and Development. 2019. Health Care Quality Indicators URL: http://www.oecd.org/health/health-systems/oecdhealthcarequalityindicatorsproject.htm [accessed 2019-04-06] [WebCite Cache ID 77RHoSJY9]
19. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
20. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. PLoS One 2015;10(9):e0138237 [FREE Full text] [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]
21. Canadian Institutes of Health Research. 2016. A Guide to Knowledge Synthesis URL: http://www.cihr-irsc.gc.ca/e/29418.html [accessed 2019-04-06] [WebCite Cache ID 77RHreFWN]
22. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
23. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013;13:61 [FREE Full text] [doi: 10.1186/1471-2288-13-61] [Medline: 23627889]
24. Bauer M, Glenn T, Monteith S, Bauer R, Whybrow PC, Geddes J. Ethical perspectives on recommending digital technology for patients with mental illness. Int J Bipolar Disord 2017 Dec;5(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0073-9] [Medline: 28155206]
25. Weldegebrial T. Marshall University, Department of Health Informatics. 2016. Regulations for eHealth & mHealth Apps before Deployment URL: https://www.researchgate.net/publication/318887422_Regulations_for_eHealth_mHealth_Apps_before_Deployment [accessed 2019-04-06] [WebCite Cache ID 77RI024W0]
26. Bisson LJ, Komm JT, Bernas GA, Fineberg MS, Marzo JM, Rauh MA, et al. Accuracy of a computer-based diagnostic program for ambulatory patients with knee pain. Am J Sports Med 2014 Oct;42(10):2371-2376. [doi: 10.1177/0363546514541654] [Medline: 25073597]
27. Bisson LJ, Komm JT, Bernas GA, Fineberg MS, Marzo JM, Rauh MA, et al. How accurate are patients at diagnosing the cause of their knee pain with the help of a web-based symptom checker? Orthop J Sports Med 2016 Feb;4(2):2325967116630286 [FREE Full text] [doi: 10.1177/2325967116630286] [Medline: 26962542]
28. Boulos MNK, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
29. Copeland C, Morreale P, Li J. m-Health application Interface design for symptom checking. 2018 Presented at: 10th International Conference on e-Health; July 17-19, 2018; Madrid, Spain URL: http://tinyurl.com/y5jotoby
30. Davies BM, Munro CF, Kotter MR. A novel insight into the challenges of diagnosing degenerative cervical myelopathy using web-based symptom checkers. J Med Internet Res 2019 Jan 11;21(1):e10868 [FREE Full text] [doi: 10.2196/10868] [Medline: 30300137]
31. Flaherty JL. Digital diagnosis: privacy and the regulation of mobile phone health applications. Am J Law Med 2014;40(4):416-441. [Medline: 27530051]
32. Farmer SE, Bernardotto M, Singh V. How good is internet self-diagnosis of ENT symptoms using Boots WebMD symptom checker? Clin Otolaryngol 2011 Oct;36(5):517-518. [doi: 10.1111/j.1749-4486.2011.02375.x] [Medline: 22032458]
33. Hageman MG, Anderson J, Blok R, Bossen JK, Ring D. Internet self-diagnosis in hand surgery. Hand (N Y) 2015 Sep;10(3):565-569 [FREE Full text] [doi: 10.1007/s11552-014-9707-x] [Medline: 26330798]
34. Jutel A, Lupton D. Digitizing diagnosis: a review of mobile applications in the diagnostic process. Diagnosis (Berl) 2015 Jun 01;2(2):89-96. [doi: 10.1515/dx-2014-0068] [Medline: 29540025]
35. Kao C, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
36. Lanseng E, Andreassen T. Electronic healthcare: a study of people's readiness and attitude toward performing self‐diagnosis. Int J of Service Industry Mgmt 2007 Aug 14;18(4):394-417. [doi: 10.1108/09564230710778155]
37. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]
38. Lupton D, Jutel A. 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. Soc Sci Med 2015 May;133:128-135. [doi: 10.1016/j.socscimed.2015.04.004] [Medline: 25864149]

39. Morita T, Rahman A, Hasegawa T, Ozaki A, Tanimoto T. The potential possibility of symptom checker. Int J Health Policy Manag 2017 Dec 05;6(10):615-616 [FREE Full text] [doi: 10.15171/ijhpm.2017.41] [Medline: 28949479]
40. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Dec 24;17(1):362 [FREE Full text] [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
41. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf 2008 May;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]
42. Tudor Car L, Papachristou N, Bull A, Majeed A, Gallagher J, El-Khatib M, et al. Clinician-identified problems and solutions for delayed diagnosis in primary care: a PRIORITIZE study. BMC Fam Pract 2016 Dec 09;17:131 [FREE Full text] [doi: 10.1186/s12875-016-0530-z] [Medline: 27613564]
43. Behrbalk E, Salame K, Regev GJ, Keynan O, Boszczyk B, Lidar Z. Delayed diagnosis of cervical spondylotic myelopathy by primary care physicians. Neurosurg Focus 2013 Jul;35(1):E1. [doi: 10.3171/2013.3.FOCUS1374] [Medline: 23815245]
44. Aboueid S, Meyer SB. Factors affecting access and use of preventive and weight management care: A public health lens. Healthc Manage Forum 2019 Apr 03:840470418824345. [doi: 10.1177/0840470418824345] [Medline: 30943766]
45. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. BMJ Qual Saf 2017 Jun;26(6):484-494 [FREE Full text] [doi: 10.1136/bmjqs-2016-005401] [Medline: 27530239]
46. Panesar SS, deSilva D, Carson-Stevens A, Cresswell KM, Salvilla SA, Slight SP, et al. How safe is primary care? A systematic review. BMJ Qual Saf 2016 Dec;25(7):544-553. [doi: 10.1136/bmjqs-2015-004178] [Medline: 26715764]
47. Nurek M, Kostopoulou O, Delaney BC, Esmail A. Reducing diagnostic errors in primary care. A systematic meta-review of computerized diagnostic decision support systems by the LINNEAUS collaboration on patient safety in primary care. Eur J Gen Pract 2015 Sep;21 Suppl:8-13 [FREE Full text] [doi: 10.3109/13814788.2015.1043123] [Medline: 26339829]

Abbreviations

AI: artificial intelligence
CSM: cervical spondylotic myelopathy
PCP: primary care provider

Edited by C Lovis; submitted 20.01.19; peer-reviewed by D Mendes, MS Aslam; comments to author 03.03.19; revised version received 04.03.19; accepted 27.03.19; published 01.05.19.

Please cite as:
Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S
The Use of Artificially Intelligent Self-Diagnosing Digital Platforms by the General Public: Scoping Review
JMIR Med Inform 2019;7(2):e13445
URL: http://medinform.jmir.org/2019/2/e13445/
doi:10.2196/13445
PMID:31042151

Review

Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review

Seyedmostafa Sheikhalishahi[1,2], MSc; Riccardo Miotto[3], PhD; Joel T Dudley[3], PhD; Alberto Lavelli[4], MSc; Fabio Rinaldi[5], PhD; Venet Osmani[1], PhD

[1]eHealth Research Group, Fondazione Bruno Kessler Research Institute, Trento, Italy

[2]Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

[3]Institute for Next Generation Healthcare, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]NLP Research Group, Fondazione Bruno Kessler Research Institute, Trento, Italy

[5]Institute of Computational Linguistics, University of Zurich, Zurich, Switzerland

Corresponding Author:
Venet Osmani, PhD
eHealth Research Group
Fondazione Bruno Kessler Research Institute
Via Sommarive 18, Povo
Trento, 38123
Italy
Phone: 39 0461 31 2479
Email: vosmani@fbk.eu

Abstract

Background: Novel approaches that complement and go beyond evidence-based medicine are required in the domain of chronic diseases, given the growing incidence of such conditions on the worldwide population. A promising avenue is the secondary use of electronic health records (EHRs), where patient data are analyzed to conduct clinical and translational research. Methods based on machine learning to process EHRs are resulting in improved understanding of patient clinical trajectories and chronic disease risk prediction, creating a unique opportunity to derive previously unknown clinical insights. However, a wealth of clinical histories remains locked behind clinical narratives in free-form text. Consequently, unlocking the full potential of EHR data is contingent on the development of natural language processing (NLP) methods to automatically transform clinical text into structured clinical data that can guide clinical decisions and potentially delay or prevent disease onset.

Objective: The goal of the research was to provide a comprehensive overview of the development and uptake of NLP methods applied to free-text clinical notes related to chronic diseases, including the investigation of challenges faced by NLP methodologies in understanding clinical narratives.

Methods: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed and searches were conducted in 5 databases using “clinical notes,” “natural language processing,” and “chronic disease” and their variations as keywords to maximize coverage of the articles.

Results: Of the 2652 articles considered, 106 met the inclusion criteria. Review of the included papers resulted in identification of 43 chronic diseases, which were then further classified into 10 disease categories using the *International Classification of Diseases, 10th Revision*. The majority of studies focused on diseases of the circulatory system (n=38) while endocrine and metabolic diseases were fewest (n=14). This was due to the structure of clinical records related to metabolic diseases, which typically contain much more structured data, compared with medical records for diseases of the circulatory system, which focus more on unstructured data and consequently have seen a stronger focus of NLP. The review has shown that there is a significant increase in the use of machine learning methods compared to rule-based approaches; however, deep learning methods remain emergent (n=3). Consequently, the majority of works focus on classification of disease phenotype with only a handful of papers addressing extraction of comorbidities from the free text or integration of clinical notes with structured data. There is a notable use of relatively simple methods, such as shallow classifiers (or combination with rule-based methods), due to the interpretability of predictions, which still represents a significant issue for more complex methods. Finally, scarcity of publicly available data may also have contributed to insufficient development of more advanced methods, such as extraction of word embeddings from clinical notes.

Conclusions: Efforts are still required to improve (1) progression of clinical NLP methods from extraction toward understanding; (2) recognition of relations among entities rather than entities in isolation; (3) temporal extraction to understand past, current, and future clinical events; (4) exploitation of alternative sources of clinical knowledge; and (5) availability of large-scale, de-identified clinical corpora.

(JMIR Med Inform 2019;7(2):e12239) doi:10.2196/12239

KEYWORDS

electronic health records; clinical notes; chronic diseases; natural language processing; machine learning; deep learning; heart disease; stroke; cancer; diabetes; lung disease

Introduction

Overview

The burden of chronic diseases, such as cancers, diabetes, and hypertension, is widely accepted as one of the principal challenges of health care. While immense progress has been made in the discovery of new treatments and prevention strategies, this challenge not only persists, but its incidence is exhibiting an upward trend [1], with significant impact on patient quality of life and care costs. Consequently, there is a need for novel approaches to complement and go beyond current evidence-based medicine that can reduce the impact of chronic conditions on modern society.

A promising direction is the secondary use of electronic health records (EHRs) to analyze patient data, advance medical research, and better inform clinical decision making. Methods based in analysis of EHRs [2] are resulting in improved understanding of patient clinical trajectories [3] while enabling better patient stratification and risk prediction [4-6]. In particular, use of machine learning and especially deep learning to process EHRs is creating a unique opportunity to derive previously unknown clinical insights [7]. This is especially relevant for chronic diseases as their longitudinal nature provides a very large and continuous stream of data, where clinically meaningful patterns can be extracted and used to guide clinical decisions, including delaying or preventing disease onset.

However, EHRs are challenging to represent and model due to their high dimensionality, noise, heterogeneity, sparseness, incompleteness, random errors, and systematic biases. Moreover, a wealth of information about patient clinical history is generally locked behind free-text clinical narratives [8] since writing text remains the most natural and expressive method to document clinical events. Development of natural language processing (NLP) methods is essential to automatically transform clinical text into structured clinical data that can be directly processed using machine learning algorithms. Use of NLP in the clinical domain is seeing an increasing uptake with diverse applications, including identification of biomedical concepts from radiology reports [9], nursing documentation [10], and discharge summaries [11]. Frameworks based on NLP applied to clinical narratives, however, have not been widely used in clinical settings to help decision support systems or workflows.

Motivation

Clinically relevant information from clinical notes has been historically extracted via manual review by clinical experts, leading to scalability and cost issues. This is of particular relevance for chronic diseases since clinical notes dominate over structured data (for example, Wei et al [12] graphically quantify the amount of clinical notes over structured data for chronic diseases such as rheumatoid arthritis, Parkinson disease, and Alzheimer disease). Availability of these data creates an immense opportunity for NLP to automatically extract clinically meaningful information that may delay or prevent disease onset, giving rise, however, to several challenges. In this paper we aimed to identify directions that could speed up the adoption of NLP of clinical notes for chronic diseases and provide an understanding of the current challenges and state of the art.

Systematic reviews related to processing of clinical notes have been published in the past [13-18]; however, none have focused specifically on chronic diseases, making it difficult to derive conclusions and recommendations on this specific and very diverse domain. In particular, this paper investigates NLP challenges related to 43 unique chronic diseases identified by our systematic review and discusses the trends of applying various NLP methods for clinical translational research. Based on the outcomes of this review, we also devised a number of recommendations on future research directions, including (1) evolution of clinical NLP methods from extraction toward understanding; (2) recognition of relations among entities, rather than entities in isolation; (3) temporal extraction in order to understand past, current, and future clinical events; (4) exploitation of alternative sources of clinical knowledge; and (5) availability of large-scale deidentified and annotated clinical corpora.

Methods

Search Strategy and Information Sources

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [19]. We carried out a search of several databases to identify all potentially relevant articles published from January 1, 2007, to February 6, 2018, including Scopus, Web of Science (including MEDLINE) and PubMed, and the Association for Computing Machinery (ACM) Digital Library. We have limited the search to journal articles written in English. In all the searches we used the combination of the following groups of keywords: (1) “clinical notes,” “medical notes,” or “clinical narratives”; (2) “natural language processing,” “medical language processing,” “text mining,” or “information extraction”; and (3) “chronic disease,” “heart disease,” “stroke,” “cancer,” “diabetes,” or “lung disease” (where the last set of keywords reflects the top five chronic diseases). The search keywords were selected to

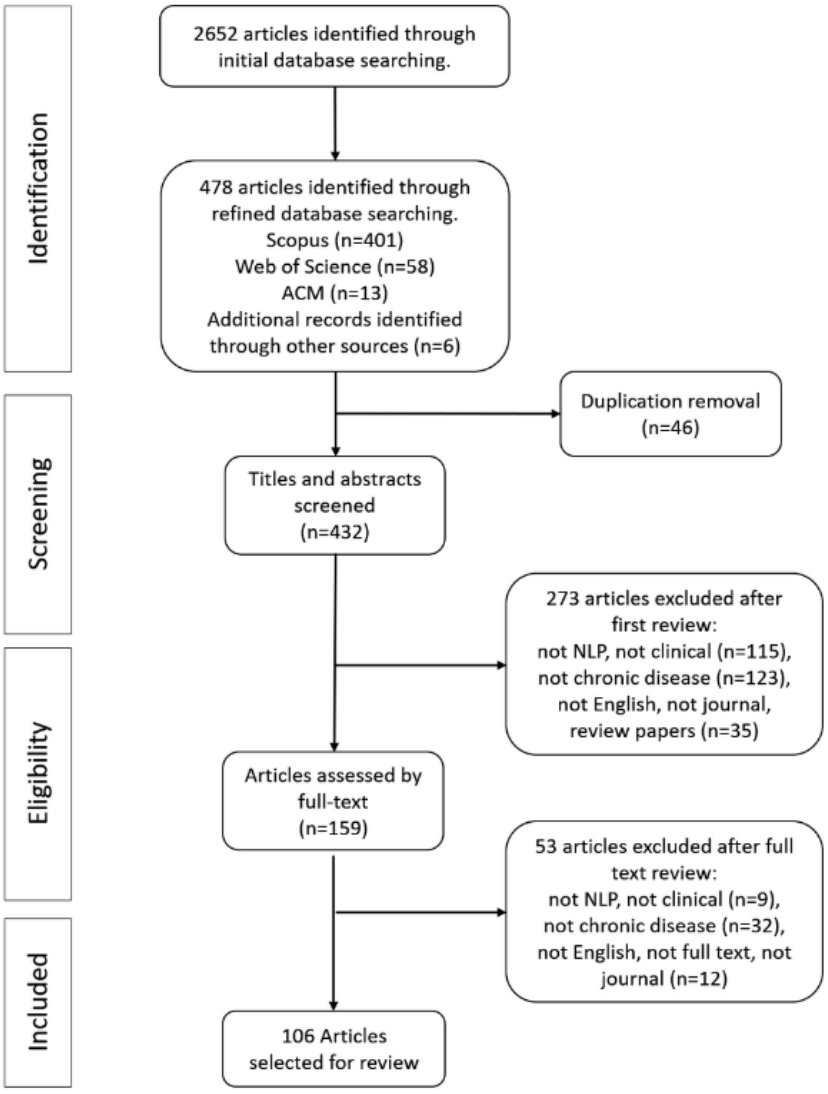
be exhaustive to maximize coverage of the articles. The exact queries are provided in Multimedia Appendix 1.

Article Selection

In the initial queries we also included the following terms: "electronic health records," "EHR," "electronic medical records," and "EMR." This led to a total of 2652 retrieved articles. However, upon reviewing these articles, we noticed that the scope was too broad, providing results outside of focus of this review. Consequently, we narrowed the search strategy to the keywords specified in the previous section, obtaining a total of 478 articles, with 401 articles from Scopus, 58 from Web of Science (including PubMed), 13 from ACM Digital Library, and 6 added manually, including 4 conference papers. After removing 46 duplicates, 432 articles were retained, and two authors (MS and VO) reviewed their titles and abstracts (216 articles each). After this screening phase, 159 articles were retained for further analysis.

In the second screening stage, five authors independently reviewed the 159 full-text articles, resulting in 106 articles fulfilling our criteria that are discussed in this review. The most common reason for exclusion was that the work was not directly related to chronic diseases (n=32); another reason was the work was not topical (eg, the article was not a journal paper or we could not retrieve the text). A flowchart and description of the selection process are provided in Figure 1 and Multimedia Appendix 2, respectively.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses article selection flowchart. ACM: Association for Computing Machinery; NLP: natural language processing.



Results

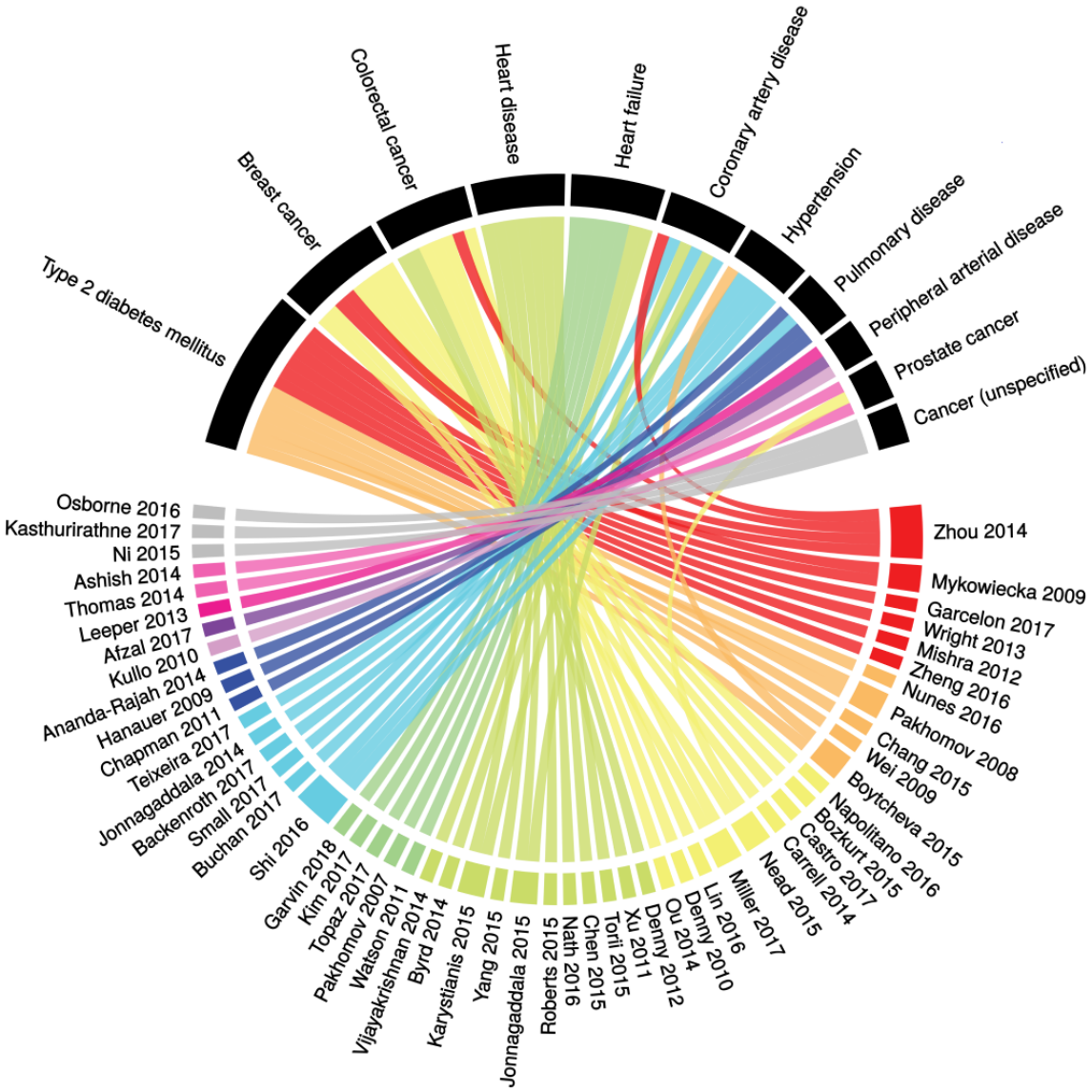
Categorization of Diseases

The 106 articles reviewed were largely related to 43 unique chronic diseases (as shown in Multimedia Appendix 2). One of our aims was to understand the extent of NLP for specific disease categories and their associated clinical notes. Therefore, we grouped the 43 unique chronic diseases into 10 disease categories using the *International Classification of Diseases, 10th Revision* (ICD-10) as shown in Table 1.

Table 1. Classifications of chronic conditions studied (n=102) and the corresponding number of papers found.

Classification of chronic condition	Studies, n (%)	Conditions included
Diseases of the circulatory system	38 (35.8)	Congestive heart disease (2), coronary artery disease (6), heart disease (6), heart failure (7), hypertension (5), peripheral arterial disease (3), pulmonary disease (4)
Neoplasms	34 (32.1)	Breast cancer (8), colorectal cancer (7), prostate cancer (4), lymphoma (2)
Endocrine, nutritional, and metabolic diseases	14 (13.2)	Type 2 diabetes mellitus (12), obesity (2)
Other diseases	16 (15.1)	Diseases of the digestive system (3), diseases of the genitourinary system (3), diseases of the musculoskeletal system and connective tissue (3), diseases of the respiratory system (2), mental and behavioral disorders (2), multidisease (3)

Figure 2. Relationship between chronic diseases (black sectors) and articles included in the review (for clarity we have included only diseases that are addressed by three or more articles).



The top three disease groups were (1) diseases of the circulatory system (n=38) (such as coronary artery disease [20] and hypertension [21]); (2) neoplasms (n=34) (such as breast cancer [22] and prostate cancer [23]); and (3) endocrine, nutritional, and metabolic diseases (n=14) (such as type 2 diabetes [24] and obesity [25]). An overview of the diseases studied and the corresponding articles is shown in Figure 2.

An unexpected finding is that despite the higher incidence of metabolic diseases in the general population [26] compared with diseases of circulatory system [27], the use of NLP in clinical narratives of these diseases exhibits an opposite trend. Diseases of the circulatory system are represented in much greater numbers with respect to metabolic diseases (n=38 vs n=14, respectively). We hypothesize that the structure of data

contained in EHRs may explain this finding. Medical records related to metabolic diseases typically contain much more structured data (for example, numerical values for various physiological and physical parameters) than medical records for diseases of the circulatory system, which focus more on unstructured data [28]. This creates a more pressing need to use NLP to extract information from notes related to diseases of the circulatory system, whereas EHRs of patients with metabolic diseases in large part may already contain data that can be used by algorithms with minimal preprocessing. In the sections that follow we summarize the most representative papers (the complete list is provided in Multimedia Appendix 2).

Disease Groups

Diseases of the Circulatory System

Cardiovascular Diseases

Most of the work in this area focused on using NLP to estimate the risk of heart disease. As an example, Chen et al [29] developed a hybrid pipeline based on both machine learning and rules to identify medically relevant information related to heart disease risk and track the disease progression over sets of longitudinal patient records, including clinical notes (similarly to Torri et al [30]). Karystianis et al [31] and Yang et al [32] evaluated the identification of heart disease risk factors from the clinical notes of diabetic patients. In a slightly different approach, Roberts et al [33] focused on estimating heart disease risk based on classification of 8 risk triggers (for example, aspirin). Other studies in this area have focused on evaluating the use of aspirin as a risk factor [34,35], extracting heart function measurements from echocardiograms [36], deep vein thrombosis and pulmonary embolism [37], and low-density lipoprotein level and statins use [38].

Risk of stroke and major bleeding in patients with atrial fibrillation has been predicted using structured data and clinical notes [39], while patients with heart failure have been identified using clinical notes only [40]. Moreover, medical reports written in the Italian language have been used to identify arrhythmia events [41].

Peripheral and Coronary Arterial Disease

Several studies used NLP to extract cases of peripheral arterial disease (PAD) and critical limb ischemia from clinical notes [42,43], including a genome-wide associated study, focusing on PAD to identify drugs, diseases, signs/symptoms, anatomical sites, and procedures [44]. Leeper et al [45] used NLP to identify PAD patients to conduct a safety surveillance study on exposure to Cilostazol, finding complications of malignant arrhythmia and sudden death not observed in association with the drug. Furthermore, Clinical Text Analysis Knowledge Extraction System (cTAKES) has been used to process clinical history of diabetic patients to predict development of PAD [46].

Hypertension

Work on hypertension has been principally focused on NLP to extract relevant indicators, comorbidities, and drug therapies [21]. Analysis of clinical narratives in the Bulgarian language of 100 million outpatient notes was used to extract numerical blood pressure values with a high sensitivity and recall [47], while term hypertension was extracted from free-text notes, using a rule-based, open-source tool [48]. Clinical notes and several types of medical documents were also used to identify hypertensive individuals using open-source medication information extraction (IE) system MedEx [49].

Right-Sided, Left-Sided, and Congestive Heart Failure

Byrd et al [50] and Jonnagaddala et al [20] proposed a hybrid NLP model to identify Framingham heart failure signs and symptoms from clinical notes and EHRs (ie, classifying whether Framingham criteria are asserted). Left ventricular ejection fraction was extracted from free-text echocardiogram reports [51], while unstructured, longitudinal EHRs of diabetic patients were used to extract relevant information of heart disease, using naïve Bayes and conditional random field (CRF) classifiers [52].

Wang et al [53] proposed a system for the identification of congestive heart failure (CHF) from EHRs, which they prospectively validated. Furthermore, left ventricular ejection fraction plus the associated qualitative and quantitative values were used to identify patients at risk of CHF [54], while free-text notes were used to distinguish left and right heart failure [55].

Heart Failure Identification

Topaz et al [56] developed an algorithm to identify heart failure (HF) patients with ineffective self-management of diet, physical activity, adherence to medication, and clinical appointments using discharge summary notes, while Garvin et al [57] focused on the quality of care for HF patients. Vijayakrishnan et al [58] explored the application of a previously validated text and data-mining tool to identify the presence of HF signs and symptoms criteria in the EHRs of a large primary care population. They found that HF signs and symptoms were documented much more frequently among the eventual HF cases, years before the first diagnosis as well, thus suggesting a potential future role for early detection of HF. Last, regular expressions were used to identify predefined psychosocial factors that served as predictors of the likelihood to be readmitted to the hospital after a case of HF [59].

Neoplasms

Overview

This section reviews a number of cancer-related studies, including detection of multiple types of cancer [60,61], extracting tumor characteristics and tumor-related information [62-64], disease trajectories of patients with cancer [65], cancer recurrence [23,66], and detection of stage of cancer [67,68].

Kasthurirathne et al [60] evaluated the performance of common classification algorithms to detect cancer cases from free-text pathology reports using nondictionary approaches. Yim et al [62] explored a machine learning algorithm to extract tumor characteristics by applying reference resolution on radiology reports. Jensen et al [65] developed a methodology that allows disease trajectories of cancer patients to be estimated from the clinical text. Napolitano et al [67] facilitated the extraction of information relevant to cancer staging, proposing a model for semistructured reports that outperformed the model for unstructured reports alone.

A number of studies have focused on different applications of NLP in pathology, histopathology, and radiology reports [69], including extracting relevant domain entities from narrative cancer pathology reports [70], negation detection of medical entities in pathology reports [71], sentence translation from pathology reports into graph representations [72], extracting information from pathology reports and pathology classifications [73,74], and named entity recognition from histopathology notes [75].

The three most common types of cancers found are breast cancer (n=8), colorectal cancer (n=7), and prostate cancer (n=4).

Breast Cancer

Carrell et al [66] proposed an NLP system to process clinical text to identify breast cancer recurrences, while Castro et al [22] addressed the automated Breast Imaging-Reporting and Data System (BI-RADS) categories extraction from breast radiology reports. Miller et al [76] proposed a tool for coreference resolution in clinical texts evaluated within the domain (colon cancer) and between domains (breast cancer). Mykowiecka et al [77] propose a rule-based IE system evaluated on mammography reports. Bozkurt et al [78] developed NLP methods to recognize lesions in free-text mammography reports and extract their corresponding relationships, producing a complete information frame for each lesion.

Colorectal and Prostate Cancer

EHRs and NLP were used to identify patients in need of colorectal cancer screening [79] and detect colonoscopy-related concepts as well as temporal-related information [80]. Additionally, EHRs and NLP were used to also identify patients with prostate biopsies positive for prostatic adenocarcinoma [81].

Liver and Pancreatic Cancer

Ping et al [82] extracted textual information concerning a set of predefined clinical concepts from a variety of clinical reports for patients with liver cancer, while Al-Haddad et al [83] identified patients with confirmed surgical pathology diagnoses of intraductal papillary mucinous neoplasms.

Endocrine, Nutritional, and Metabolic Diseases

Applications of NLP in the domain of endocrine, nutritional, and metabolic diseases include negation detection and mention of family history in free-text notes [84] and assigning temporal tags to medical concepts [85]; obesity [25,86] and diabetes identification [77,87-89]; and diabetes complications such as foot examination findings [90], vision loss [91], and quantifying the occurrence of hypoglycemia [24].

Two support vector machines (SVMs) were combined to automatically identify obesity types by extracting obesity and diabetes-related concepts from clinical text [86] in addition to patient identification [92]. An SVM-based system was developed and validated to identify EHR progress notes pertaining to diabetes [87], while foot examination findings from clinical reports [90] were used to predict quality of life [93]. Additionally, an analysis of a large EHR database was used to quantify occurrence of hypoglycemia [24].

Other Disease Categories

The remaining 16 papers focused on processing clinical notes of different types of chronic diseases. Three studies concern diseases of the musculoskeletal system and connective tissue, in particular classification of snippets of text related to axial spondyloarthritis in the EMRs of US military veterans using NLP and SVM [94], phenotyping systemic lupus erythematosus [95], and identification of rheumatoid arthritis patients via ontology-based NLP and logistic regression [96]. In the domain of diseases of the digestive system, Chen et al [97] used natural language features from pathology reports to identify celiac disease patients, Soguero-Ruiz et al [98] used feature selection and SVMs to detect early complications after colorectal cancer, and Chang et al [99] integrated rule-based NLP on notes with ICD-9s and lab values in an algorithm to better define and risk-stratify patients with cirrhosis.

Two papers evaluated deep learning in a multidisease domain. In particular, Miotto et al [3] derived a general purpose patient representation from aggregated EHRs (structured clinical data and clinical notes) based on neural networks that facilitates clinical predictive modeling given the patient status. Clinical notes were parsed using the National Center for Biomedical Ontology's Open Biomedical Annotator to extract medical terms and further processed using topic modeling (latent Dirichlet allocation). Shi et al [100] proposed assessing disease risk from patient clinical notes using word embeddings and convolutional neural networks with full connection layer.

Neural networks were also used to process clinical notes for phenotyping psychiatric diagnosis [101]. In particular, this model included two neural networks, one highly accurate at rejecting patients but poor at identifying suitable ones and the other one with the opposite capabilities. In the same domain of mental and behavioral disorders, comorbidity networks were derived from the patient notes at the largest Danish psychiatric hospital in order to extract disease correlations [102].

IE from clinical notes based on NLP was also used to (1) screen computed tomography reports for invasive pulmonary mold [103], (2) discover the co-occurrences of chronic obstructive pulmonary disease with other medical terms [104], (3) quantify the relationship between aggregated preoperative risk factors and cataract surgery complications [105], (4) detect patients with multiple sclerosis from the clinical notes prior to the initial recognition by their health care providers [106], and (5) identify patients on dialysis in the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) publicly available dataset [107].

Last, Pivovarov and Elhadad [108] used clinical notes of patients with chronic kidney disease to validate a novel model to compute the similarity of two medical concepts by combining complementary information derived from usage patterns of clinical documentation, accepted definitions, and position of the concepts in an ontology.

Information Extraction Methods

In order to understand trends in NLP methods for chronic diseases, in this review we have analyzed papers with respect to the methods employed (machine vs rule-based learning). While there is an increasing use of machine learning methods

in comparison to rule-based (as shown in Figure 3), it is not as pronounced as we had expected considering the superior performance of machine learning algorithms shown in the NLP literature [109]. This result may reflect the fact that we are still currently witnessing a transition from rule-based methods to machine learning algorithms, with rule-based methods used as a baseline to compare the performance of machine learning approaches.

Our review identified 16 papers that employed hybrid approaches combining rule-based and machine learning methods. Out of these, 2 papers describe work to identify diseases, risk factors, medications, and time attributes. In particular, a hybrid pipeline based on CRFs, SVMs, and rule-based approaches was used to identify negation information and normalize temporal expressions [29], while a series of SVM models in conjunction with manually built lexicons were used to classify triggers specific to each risk factor [33].

We identified 24 papers that focused on comparison between performance of rule-based and machine learning methods. Typically, the rule-based methods were used as a baseline to test the performance against machine learning algorithms.

As for rule-based approaches, the methods in this review include dictionary lookup [110-112], terminology identification based on domain ontologies [3,42,45,58], various types of manually defined rules [37,113], and regular expressions patterns [114,115].

The most widely used machine learning approach is SVMs, having been used for predicting heart disease in medical records [32,46], identifying EHR progress notes pertaining to diabetes [94], and categorizing breast radiology reports according to BI-RADS [22].

Naïve Bayes was the second most frequent approach, being used to predict heart disease in medical records [30,80], classify smoking status [52], search EMR records to identify multiple sclerosis [106], and classify EMR records for obesity [86] and cancer [60,65,67]. CRFs are the third most frequent approach, have been used to predict heart disease in medical records [29,32], identify EHR progress notes pertaining to diabetes [85], categorize breast radiology reports [22], and identify tumor attributes in radiology reports [63]. Lastly, random forests were used for predicting heart disease [53], classifying cancer types [60], and identifying hypertension [49].

Figure 3. Natural language processing rule-based methods versus machine learning for chronic diseases.

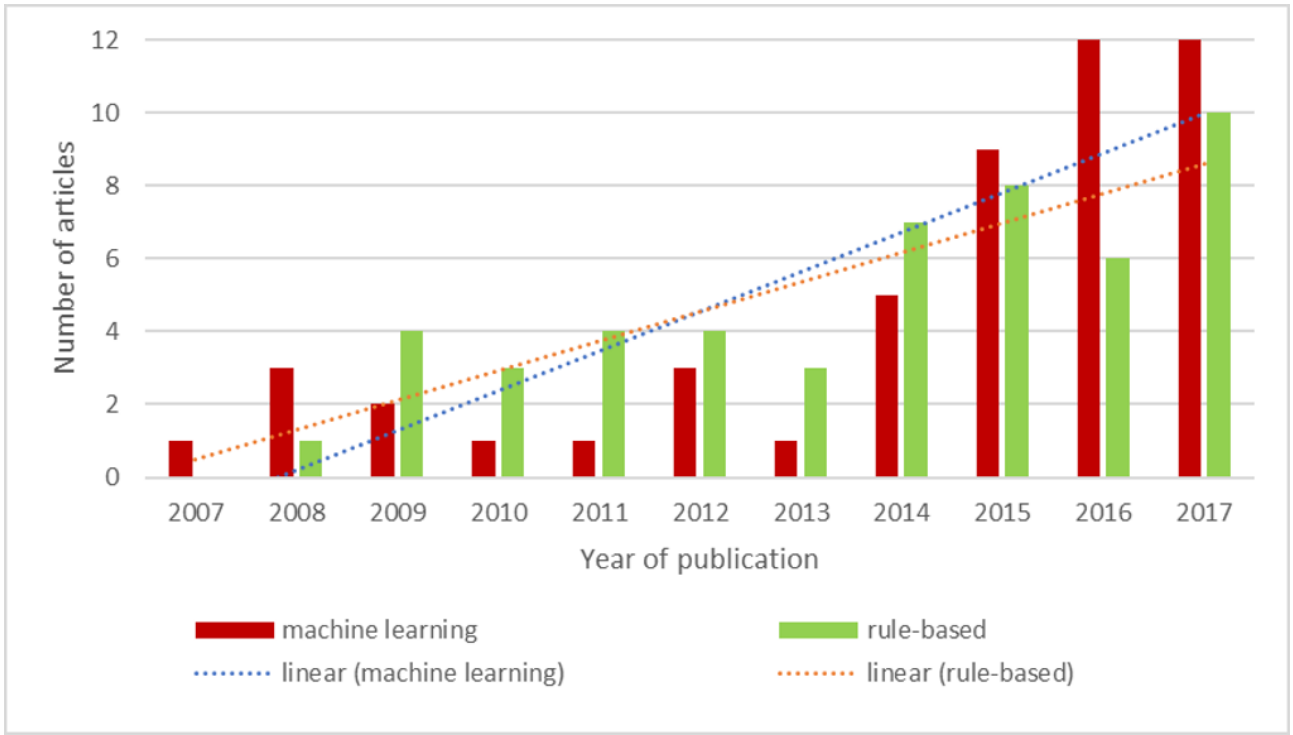


Table 2. Most frequently used natural language processing methods and the corresponding number of papers.

Method	Papers (n)
Support vector machine	18
Naïve Bayes	11
Conditional random fields	7
Random forest	4
Maximum entropy	3
Decision tree	3
Deep neural networks	3
Logistic regression	3
Rule-based methods	74

It is interesting to note that there are only 3 papers using approaches based on deep learning [3,100,101], as shown in Table 2. In particular, Geraci et al [101] apply deep neural networks to EMRs to identify suitable candidates for a study on youth depression. Miotto et al [3] present a method to derive a patient representation that facilitates clinical predictive modeling from aggregate EHRs, including clinical narratives. They represented free-text notes using topic modeling. This method significantly outperformed those achieved by standard feature learning strategies. Finally, Shi et al [100] propose a disease assessment model based on clinical notes, using convolutional neural network for disease risk assessment. The experiment involved patients with cerebral infarction, pulmonary infection, and coronary atherosclerotic heart disease.

Natural Language Processing Tasks, Methods, and Datasets

The NLP works described in the reviewed papers and associated approaches reveal that the most frequently described tasks are text classification and entity recognition. The majority of the papers describe text classification tasks using standard approaches in NLP such as SVM (n=12) and naïve Bayes (n=4). Entity recognition approaches are based on manually developed resources (dictionary, regular expressions, handwritten rules) as well as methods based on machine learning. As for the former, there are dictionary-based approaches (n=5) and those relying on regular expressions (n=12). As for the latter, the approaches are mainly based on standard machine language techniques such as CRF and deep learning. A few papers describe approaches to coreference resolution (n=2) and negation detection (n=3). Coreference resolution is addressed using SVM, while negation detection is based on SVM (n=2) or manual rules (n=1).

Regarding datasets, the majority of the papers describe experiments run on datasets that are not publicly available (typically clinical data collected at research-based health care institutions and exploited by in-house NLP teams). On the other hand, out of 16 papers involving publicly available corpora, 12 exploit the Informatics for Integrating Biology and the Bedside (i2b2) datasets. The other 4 public datasets used are MIMIC-II [107], PhenoCHF [116], Temporal Histories of Your Medical Event (THYME), and Cancer Deep Phenotype Extraction (DeepPhe) [76].

Comparisons to Other Systematic Reviews

Interest in using NLP for the automated processing of medical records, and in particular of free-text clinical notes, is increasing, exemplified by a number of recent reviews of the field. Yet none of these works focuses solely on chronic diseases, where the amount of patient clinical notes tends to be larger than other domains or provides specific recommendations on how to advance the field toward a clinical adoption that helps in treating people with chronic conditions. Here we briefly provide a summary of previous works partially related to the work presented in this paper.

Ford et al [13] present a systematic review of 67 papers using IE techniques applied to medical records for the purpose of case detection (ie, finding occurrences of specific medical conditions). Similarly, Kreimeyer et al [117] review 86 papers focusing on clinical NLP systems and a set of 71 associated NLP tasks.

The work by Shivade et al [14] reviews 97 papers aiming at identifying patient cohorts for further medical studies. Different from our work, theirs is not limited to investigation of studies using NLP and text mining but includes rule-based approaches, which do not make use of the textual part of the medical records. They observe, however, that the use of machine learning and statistical and NLP methods is on the rise compared to rule-based systems.

Abbe et al [118] consider applications of text mining in psychiatry through a PRISMA-based review. The study evaluates the application of specific NLP techniques in relation to the goal of the studies, first qualitatively, and then with a cluster analysis of the topics of selected abstracts. It identifies four main themes in the publications taken into consideration: (1) psychopathology (2), patient perspective, (3) medical records, and (4) medical literature. The scope of this review only partially overlaps with our own, given the narrow thematic analysis and inclusion of studies that deal with IE from other textual resources, such as patient perspectives.

The review by Spasic et al [119] focuses on cancer research. The authors classify the studies by cancer type and type of processed document. They do not focus solely on studies based on medical records or other types of clinical documents but also include meta-studies that apply text mining techniques to

PubMed publications. They classify NLP applications in four categories: named entity recognition, IE, text classification, and information retrieval. Their investigation reveals a predominance of symbolic approaches (dictionary and rule-based).

The work by Pons et al [120] is a systematic review of NLP applications in the area of radiology. After initial preselection based on abstracts, a detailed review of the full text of the selected papers ultimately yields 67 publications, all deemed to consider practical applications of NLP in radiology. The selected publications are then grouped into five broad categories depending on the specific application: diagnostic surveillance, cohort building, query-case retrieval, quality assessment of radiological practice, and clinical support services. The authors provide a detailed comparative analysis of the performance reported in each publication, grouping them by application category.

Closest to our work is a systematic review by Wang et al [18] that has focused on IE applications; however, our review additionally includes methodologies used in analysis of clinical notes, providing a wider set of articles. We believe that our review has a broader and more recent coverage of chronic diseases, followed by detailed analysis for each disease, compared with previous reviews, which have focused on specific conditions such as cancer [119], psychiatry [118], radiology [120], or IE applications [18].

Publication Venues

The 106 articles considered in this review were published in 50 unique venues. Figure 4 illustrates how we manually sorted publication venues into three categories: (1) clinical medicine, (2) medical informatics, and (3) computer science. We observed that most of the studies were published in medical informatics journals. Figure 5 shows an increasing trend in number of publications over the years (except for the year 2018 due to partial-year retrieval) implying an increasing interest in the application of NLP in both clinical and informatics research for chronic diseases.

Figure 4. Categorization of the publication venues.

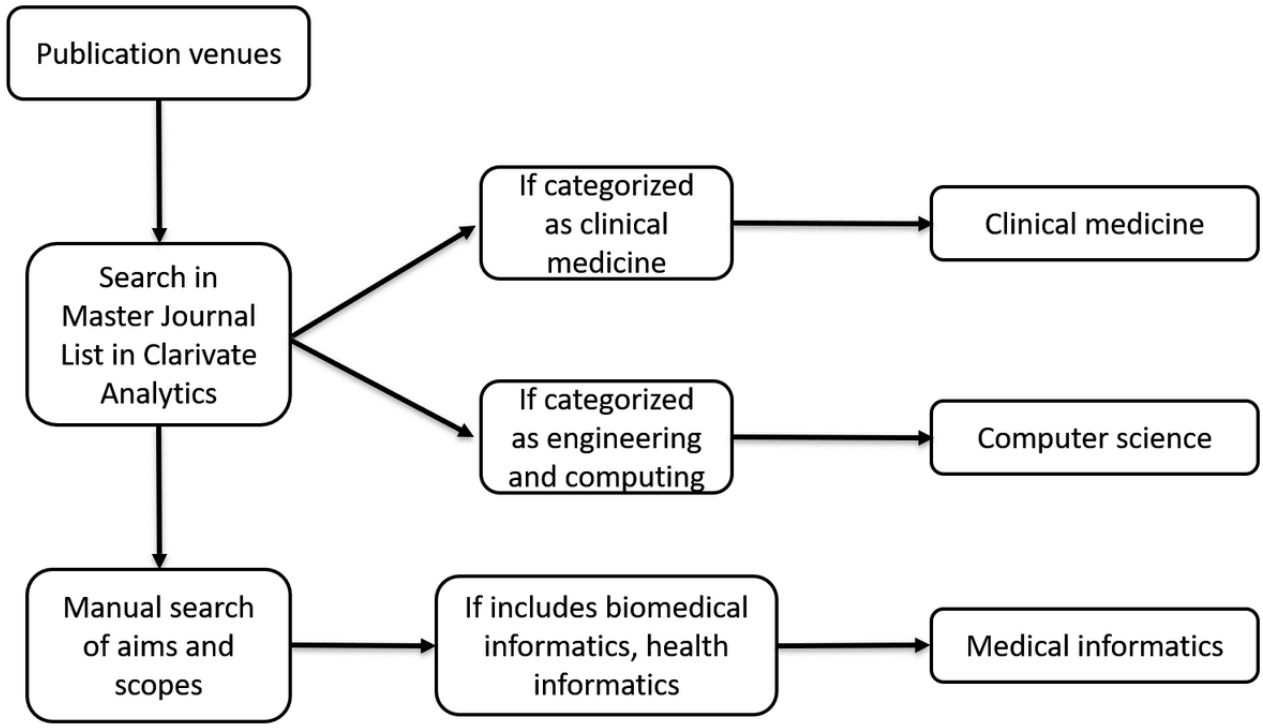
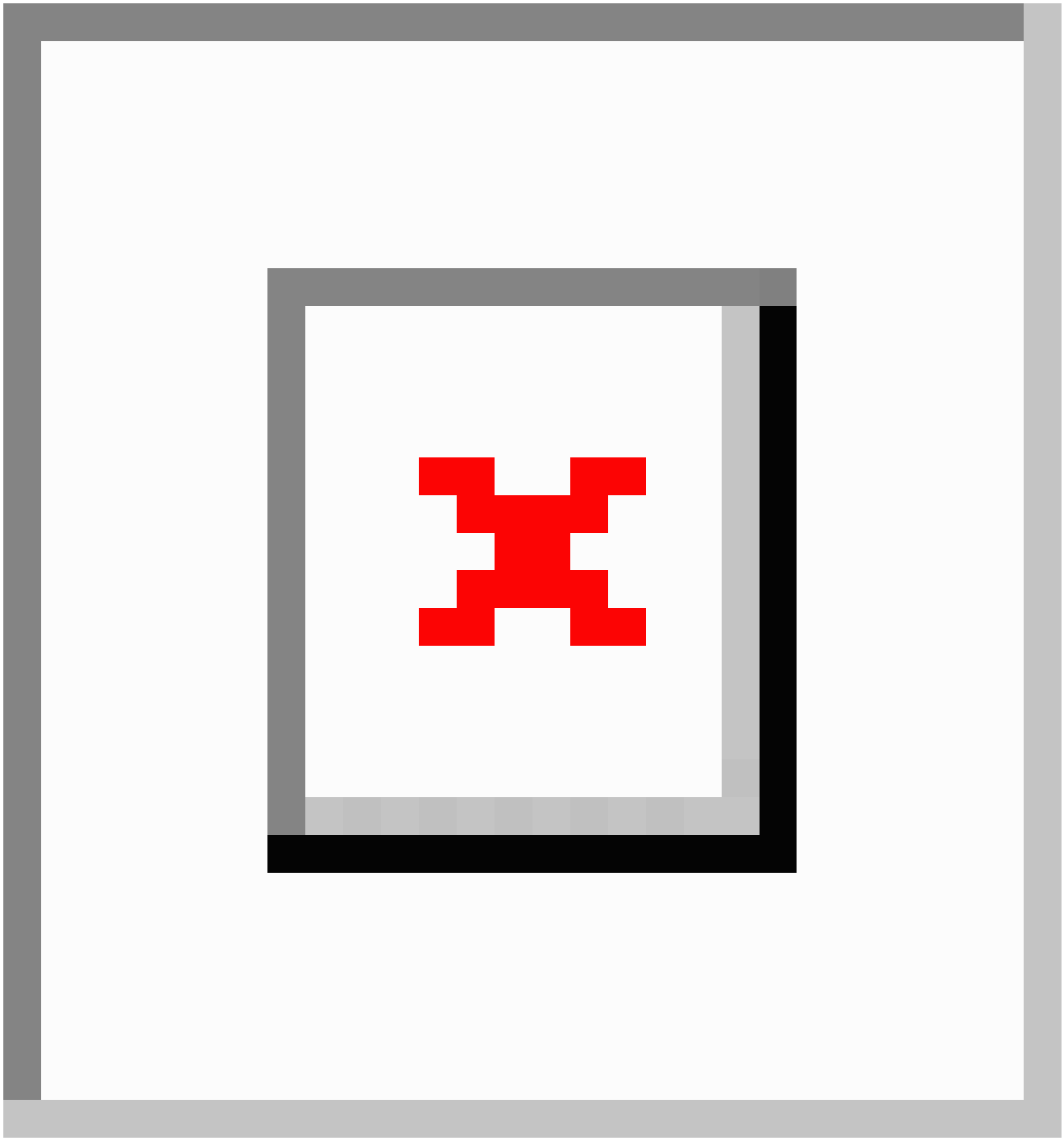


Figure 5. Distribution of included studies according to publication venues.



Discussion

Principal Findings

Our systematic review has shown that NLP has a wide range of applications for processing clinical notes of diverse chronic diseases (43 unique chronic diseases identified in the analysis). In this respect, there is a significant increase in the use of machine learning compared with rule-based methods. Despite the potential offered by deep learning, the majority of papers still rely on shallow classifiers. In fact, only a handful of studies (ie, 3 papers) made use of deep classifiers or general deep learning methods for NLP. This was unexpected, considering the potential of deep learning for text processing [121]. Our hypothesis is that since deep learning is still an emerging area, initial applications in the clinical domain may have been published in workshops, conference proceedings, and the e-print repository arXiv rather than journals, the focus of this review. In this respect, a keyword search in arXiv for “deep learning,” and “clinical notes,” “medical notes,” or “clinical narratives” for the previous five years (2013-2018) shows a significant growth of papers: 7 from 2013 to 2015, 13 in 2016, 19 in 2017, and 22 in 2018. In addition, the longer review time for journals has likely contributed to this outcome for the more recent papers. We expect this result to shift in the coming years as an increasing amount of work based on deep learning to process clinical notes is published in peer-reviewed journals.

Another finding from our review is that the majority of papers reviewed identify risk factors for a particular disease and classify

a clinical note by a certain disease phenotype. However, there are only a handful of papers that extract comorbidities from the free-text or integrate clinical notes with structured data for prediction and longitudinal modeling of trajectories of patients with chronic diseases. Such an outcome could be related to the use of data analysis methods and algorithms (such as shallow classifiers and rule-based approaches highlighted earlier) that do not have the capability to capture temporal and longitudinal relationships between clinical variables and in turn capture disease evolution. Tools (such as MetaMap) and methods (such as mapping n-grams to ontologies) used may have been other influencing factors. While these tools allow extracting meaningful medical information from the text, inherently they reduce the possibility to derive more complex relationships, principally due to phrase structure (for example “breast and lung cancer” may be identified only as “breast” and “lung cancer” rather than both “breast cancer” and “lung cancer”). However, the use of relatively simple methods is advantageous in terms of interpretability of predictions—a highly important aspect in clinical domain—whereas it still represents a significant issue for more complex methods.

Our review has retrieved only a few studies on the topic of extracting word embeddings from clinical notes. This may be due to insufficient available data to train the algorithms as well as the fact that embedding methods have been developed only recently. The issue of insufficient training data could be addressed using transfer learning methods, while using precomputed embeddings for specific diseases or categories of diseases could be useful to effectively capture longitudinal relationships.

Our review has shown that SVM and naïve Bayes algorithms were most often used for machine learning–based tasks or in combination with rule-based methods. This may be due to the popularity of these algorithms as well as because naïve Bayes, being a relatively simple algorithm, requires relatively small amount of training data (in comparison with deep classifiers, for example). Although it is not feasible to directly compare algorithmic performance of the studies that we considered (due to both diversity of data and challenges addressed), we have noted that the most commonly reported performance measures were sensitivity (recall), positive predictive value (precision), and *F* score.

Finally, our review has reinforced the fact that availability of public datasets remains scarce. This outcome was largely expected given the sensitivity of clinical data in addition to all the legal and regulatory issues, including the Health Insurance Portability and Accountability Act and the Data Protection Directive (Directive 95/46/EC) of the European Law (superseded by the General Data Protection Regulation 2016/679). As a result, the studies reviewed in this paper typically came from research-based health care institutions with in-house NLP teams having access to clinical data. Therefore, the need remains for shared tasks such as i2b2 and access to data that would increase participation in clinical NLP and contribute to improvements of NLP methods and algorithms targeting clinical applications.

Limitations

This review has examined the last 11 years of clinical IE applications literature and may have the following limitations. The review is limited to journal articles written in the English language, and papers written in other languages, especially papers that consider clinical narratives, may provide additional results. In addition, papers using clinical articles from non-EHR systems have not been considered. Finally, focusing on the clinical domain may have introduced a bias with respect to the methods reviewed (rule-based vs machine learning), as rule-based methods are more prevalent in the clinical domain compared with other domains [122].

Recommendations

Our review has shown that there is a clear necessity for clinical NLP methods to evolve beyond extraction of clinical concepts and focus more on concept understanding (ie, not only understanding of relationships between concepts but incorporation of clinical facts, domain knowledge, and general knowledge in the reasoning process). In this review, we have not encountered work that attempts to bridge the gap between concept extraction and concept understanding.

We have devised the following specific recommendations:

1. Focus on recognition of relationships among clinical concepts and entities. While progress has been made in recognizing entities in textual narratives (such as diseases, drugs, procedures), further efforts must be focused on automatic inference of relationships between these entities (for example, drug A causes adverse event B for chronic disease C), which in turn would allow deeper understanding of clinical text.
2. Temporal extraction, automated mark-up and normalization of temporal information from natural language texts, is an important aspect. This is especially relevant for clinical text as disease progression and clinical events are typically recorded chronologically, with specific events being significant only in a particular temporal context. As such, significant attention should be given to temporal extraction considering its implication in clinical context, especially since none of the works in this review dealt with temporal extraction (or used crude methods such as timestamps of clinical notes).
3. Scarcity of annotated clinical corpora has raised the need to exploit alternative sources of domain knowledge. In addition to mainstream sources such as biomedical literature, encyclopedias, and textbooks, automatic diagnostic and decision support systems could be exploitable (such as DXplain [123]). Transfer learning, a method of transferring knowledge from existing corpora in other domains to the clinical domain, also holds great potential and should be investigated in more detail.
4. Significant advances in effective clinical NLP will depend on large-scale corpora becoming available to researchers. While shared tasks such as i2b2 and its successor n2c2 are steps in the right direction, further incentives will be required such as developing mechanisms that would empower patients to donate their anonymized data or even

providing algorithms that run on clinical text inside care institutions.

Acknowledgments

This work was partially supported by the European Union's Horizon 2020 research and innovation program under grant agreement #769765.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[PDF File (Adobe PDF File), 178KB - medinform_v7i2e12239_app1.pdf]

Multimedia Appendix 2

Complete list of reviewed papers, chronic diseases and their classifications, algorithms used, publication venues, and excluded papers.

[XLSX File (Microsoft Excel File), 107KB - medinform_v7i2e12239_app2.xlsx]

References

1. World Health Organization. WHO Global status report on noncommunicable diseases 2014 URL: https://www.who.int/nmh/publications/ncd-status-report-2014/en/ [accessed 2019-03-29] [WebCite Cache ID 77Fa8uXax]
2. Kruse CS, Kothman K, Anerobi K, Abanaka L. Adoption factors of the electronic health record: a systematic review. JMIR Med Inform 2016 Jun 01;4(2):e19. [doi: 10.2196/medinform.5525] [Medline: 27251559]
3. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 Dec 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
4. Jensen P, Jensen L, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 02;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
5. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208. [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
6. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. J Med Internet Res 2018 Jan 30;20(1):e22 [FREE Full text] [doi: 10.2196/jmir.9268] [Medline: 29382633]
7. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2017 May 06. [doi: 10.1093/bib/bbx044] [Medline: 28481991]
8. Jensen K, Soguero-Ruiz C, Oyvind MK, Lindsetmo R, Kouskoumvekaki I, Girolami M, et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. Sci Rep 2017 Dec 07;7:46226 [FREE Full text] [doi: 10.1038/srep46226] [Medline: 28387314]
9. Flynn R, Macdonald TM, Schembri N, Murray GD, Doney ASF. Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes. Pharmacoepidemiol Drug Saf 2010 Aug;19(8):843-847. [doi: 10.1002/pds.1981] [Medline: 20602346]
10. Popejoy LL, Khalilia MA, Popescu M, Galambos C, Lyons V, Rantz M, et al. Quantifying care coordination using natural language processing and domain-specific ontology. J Am Med Inform Assoc 2015 Apr;22(e1):e93-e103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002702] [Medline: 25324557]
11. Yang H, Spasic I, Keane JA, Nenadic G. A text mining approach to the prediction of disease status from clinical discharge summaries. J Am Med Inform Assoc 2009;16(4):596-600 [FREE Full text] [doi: 10.1197/jamia.M3096] [Medline: 19390098]
12. Wei W, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J Am Med Inform Assoc 2016 Apr;23(e1):e20-e27 [FREE Full text] [doi: 10.1093/jamia/ocv130] [Medline: 26338219]
13. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc 2016 Dec;23(5):1007-1015 [FREE Full text] [doi: 10.1093/jamia/ocv180] [Medline: 26911811]

14. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc 2014;21(2):221-230 [FREE Full text] [doi: 10.1136/amiajnl-2013-001935] [Medline: 24201027]
15. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. Int J Methods Psychiatr Res 2016 Dec;25(2):86-100. [doi: 10.1002/mpr.1481] [Medline: 26184780]
16. Spasić I, Livsey J, Keane JA, Nenadić G. Text mining of cancer-related information: review of current status and future directions. Int J Med Inform 2014 Sep;83(9):605-623 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.009] [Medline: 25008281]
17. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural language processing in radiology: a systematic review. Radiology 2016 May;279(2):329-343. [doi: 10.1148/radiol.16142770] [Medline: 27089187]
18. Wang Y, Wang ZF, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Dec;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
20. Jonnagaddala J, Liaw S, Ray P, Kumar M, Chang N, Dai H. Coronary artery disease risk assessment from unstructured electronic health records using text mining. J Biomed Inform 2015 Dec;58 Suppl:S203-S210 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.003] [Medline: 26319542]
21. Backenroth D, Chase HS, Wei Y, Friedman C. Monitoring prescribing patterns using regression and electronic health records. BMC Med Inform Decis Mak 2017 Dec 19;17(1):175 [FREE Full text] [doi: 10.1186/s12911-017-0575-5] [Medline: 29258594]
22. Castro SM, Tseytlin E, Medvedeva O, Mitchell K, Visweswaran S, Bekhuis T, et al. Automated annotation and classification of BI-RADS assessment from radiology reports. J Biomed Inform 2017 Dec;69:177-187 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.011] [Medline: 28428140]
23. Heintzelman NH, Taylor RJ, Simonsen L, Lustig R, Anderko D, Haythornthwaite JA, et al. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text. J Am Med Inform Assoc 2013 Sep;20(5):898-905 [FREE Full text] [doi: 10.1136/amiajnl-2012-001076] [Medline: 23144336]
24. Nunes AP, Yang J, Radican L, Engel SS, Kurtyka K, Tunceli K, et al. Assessing occurrence of hypoglycemia and its severity from electronic health records of patients with type 2 diabetes mellitus. Diabetes Res Clin Pract 2016 Nov;121:192-203. [doi: 10.1016/j.diabres.2016.09.012] [Medline: 27744128]
25. Ware H, Mullett CJ, Jagannathan V. Natural language processing framework to assess clinical conditions. J Am Med Inform Assoc 2009;16(4):585-589 [FREE Full text] [doi: 10.1197/jamia.M3091] [Medline: 19390100]
26. Golden SH, Robinson KA, Saldanha I, Anton B, Ladenson PW. Clinical review: prevalence and incidence of endocrine and metabolic disorders in the United States: a comprehensive review. J Clin Endocrinol Metab 2009 Jun;94(6):1853-1878 [FREE Full text] [doi: 10.1210/jc.2008-2291] [Medline: 19494161]
27. Sanchis-Gomar F, Perez-Quilis C, Leischik R, Lucia A. Epidemiology of coronary heart disease and acute coronary syndrome. Ann Transl Med 2016 Jul;4(13):256 [FREE Full text] [doi: 10.21037/atm.2016.06.33] [Medline: 27500157]
28. Liao KP, Ananthakrishnan AN, Kumar V, Xia Z, Cagan A, Gainer VS, et al. Methods to develop an electronic medical record phenotype algorithm to compare the risk of coronary artery disease across 3 chronic disease cohorts. PLoS One 2015;10(8):e0136651 [FREE Full text] [doi: 10.1371/journal.pone.0136651] [Medline: 26301417]
29. Chen Q, Li H, Tang B, Wang X, Liu X, Liu Z, et al. An automatic system to identify heart disease risk factors in clinical texts over time. J Biomed Inform 2015 Dec;58 Suppl:S158-S163 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.002] [Medline: 26362344]
30. Torii M, Fan J, Yang W, Lee T, Wiley MT, Zisook DS, et al. Risk factor detection for heart disease by applying text analytics in electronic medical records. J Biomed Inform 2015 Dec;58 Suppl:S164-S170 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.011] [Medline: 26279500]
31. Karystianis G, Dehghan A, Kovacevic A, Keane JA, Nenadic G. Using local lexicalized rules to identify heart disease risk factors in clinical notes. J Biomed Inform 2015 Dec;58 Suppl:S183-S188 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.013] [Medline: 26133479]
32. Yang H, Garibaldi JM. A hybrid model for automatic identification of risk factors for heart disease. J Biomed Inform 2015 Dec;58 Suppl:S171-S182 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.006] [Medline: 26375492]
33. Roberts K, Shooshan SE, Rodriguez L, Abhyankar S, Kilicoglu H, Demner-Fushman D. The role of fine-grained annotations in supervised recognition of risk factors for heart disease from EHRs. J Biomed Inform 2015 Dec;58 Suppl:S111-S119 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.010] [Medline: 26122527]
34. Pakhomov S, Shah N, Hanson P, Balasubramaniam S, Smith S. Automated processing of electronic medical records is a reliable method of determining aspirin use in populations at risk for cardiovascular events. Inform Prim Care 2010;18(2):125-133 [FREE Full text] [Medline: 21078235]

35. Zheng C, Rashid N, Koblick R, An J. Medication extraction from electronic clinical notes in an integrated health system: a study on aspirin use in patients with nonvalvular atrial fibrillation. Clin Ther 2015 Sep;37(9):2048-2052. [doi: 10.1016/j.clinthera.2015.07.002] [Medline: 26233471]
36. Patterson OV, Freiberg MS, Skanderson M, Brandt CA, DuVall SL. Unlocking echocardiogram measurements for heart disease research through natural language processing. BMC Cardiovasc Disord 2017 Dec 12;17(1):151 [FREE Full text] [doi: 10.1186/s12872-017-0580-8] [Medline: 28606104]
37. Tian Z, Sun S, Eguale T, Rochefort CM. Automated extraction of VTE events from narrative radiology reports in electronic health records: a validation study. Med Care 2017 Dec;55(10):e73-e80 [FREE Full text] [doi: 10.1097/MLR.0000000000000346] [Medline: 25924079]
38. Ross EG, Shah N, Leeper N. Statin intensity or achieved LDL? Practice-based evidence for the evaluation of new cholesterol treatment guidelines. PLoS One 2016;11(5):e0154952 [FREE Full text] [doi: 10.1371/journal.pone.0154952] [Medline: 27227451]
39. Wang SV, Rogers JR, Jin Y, Bates DW, Fischer MA. Use of electronic healthcare records to identify complex patients with atrial fibrillation for targeted intervention. J Am Med Inform Assoc 2017 Mar 01;24(2):339-344. [doi: 10.1093/jamia/ocw082] [Medline: 27375290]
40. Pakhomov S, Weston S, Jacobsen S, Chute C, Meverden R, Roger V. Electronic medical records for clinical research: application to the identification of heart failure. Am J Manag Care 2007 Jun;13(6 Part 1):281-288 [FREE Full text] [Medline: 17567225]
41. Viani N, Larizza C, Tibollo V, Napolitano C, Priori SG, Bellazzi R, et al. Information extraction from Italian medical reports: an ontology-driven approach. Int J Med Inform 2018 Mar;111:140-148. [doi: 10.1016/j.ijmedinf.2017.12.013] [Medline: 29425625]
42. Afzal N, Mallipeddi VP, Sohn S, Liu H, Chaudhry R, Scott CG, et al. Natural language processing of clinical notes for identification of critical limb ischemia. Int J Med Inform 2018 Mar;111:83-89 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.12.024] [Medline: 29425639]
43. Afzal N, Sohn S, Abram S, Scott CG, Chaudhry R, Liu H, et al. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. J Vasc Surg 2017 Dec;65(6):1753-1761 [FREE Full text] [doi: 10.1016/j.jvs.2016.11.031] [Medline: 28189359]
44. Kullo IJ, Fan J, Pathak J, Savova GK, Ali Z, Chute CG. Leveraging informatics for genetic studies: use of the electronic medical record to enable a genome-wide association study of peripheral arterial disease. J Am Med Inform Assoc 2010;17(5):568-574 [FREE Full text] [doi: 10.1136/jamia.2010.004366] [Medline: 20819866]
45. Leeper NJ, Bauer-Mehren A, Iyer SV, Lependu P, Olson C, Shah NH. Practice-based evidence: profiling the safety of cilostazol by text-mining of clinical notes. PLoS One 2013;8(5):e63499 [FREE Full text] [doi: 10.1371/journal.pone.0063499] [Medline: 23717437]
46. Buchan K, Filannino M, Uzuner O. Automatic prediction of coronary artery disease from clinical narratives. J Biomed Inform 2017 Dec;72:23-32 [FREE Full text] [doi: 10.1016/j.jbi.2017.06.019] [Medline: 28663072]
47. Boytcheva S, Angelova G, Angelov Z, Tcharaktchiev D. Text mining and big data analytics for retrospective analysis of clinical texts from outpatient care. Cybern Inf Technol 2015;1(4):55-77 [FREE Full text] [doi: 10.1515/cait-2015-0055]
48. Jonnagaddala J, Liaw S, Ray P, Kumar M. HTNSystem: hypertension information extraction system for unstructured clinical notes. Lect Notes Comput Sci 2014:219-227 [FREE Full text] [doi: 10.1007/978-3-319-13987-6_21]
49. Teixeira PL, Wei W, Cronin RM, Mo H, VanHouten JP, Carroll RJ, et al. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. J Am Med Inform Assoc 2017 Jan;24(1):162-171 [FREE Full text] [doi: 10.1093/jamia/ocw071] [Medline: 27497800]
50. Byrd RJ, Steinhubl SR, Sun J, Ebadollahi S, Stewart WF. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. Int J Med Inform 2014 Dec;83(12):983-992 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.12.005] [Medline: 23317809]
51. Garvin JH, DuVall SL, South BR, Bray BE, Bolton D, Heavirland J, et al. Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architecture (UIMA) for heart failure. J Am Med Inform Assoc 2012;19(5):859-866 [FREE Full text] [doi: 10.1136/amiajnl-2011-000535] [Medline: 22437073]
52. Jonnagaddala J, Liaw S, Ray P, Kumar M, Dai H, Hsu C. Identification and progression of heart disease risk factors in diabetic patients from longitudinal electronic health records. Biomed Res Int 2015;2015:636371 [FREE Full text] [doi: 10.1155/2015/636371] [Medline: 26380290]
53. Wang Y, Luo J, Hao S, Xu H, Shin AY, Jin B, et al. NLP based congestive heart failure case finding: a prospective analysis on statewide electronic medical records. Int J Med Inform 2015 Dec;84(12):1039-1047. [doi: 10.1016/j.ijmedinf.2015.06.007] [Medline: 26254876]
54. Kim Y, Garvin JH, Goldstein MK, Hwang TS, Redd A, Bolton D, et al. Extraction of left ventricular ejection fraction information from various types of clinical reports. J Biomed Inform 2017 Dec;67:42-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.01.017] [Medline: 28163196]

55. American Heart Association. Types of Heart Failure URL: https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/types-of-heart-failure [accessed 2019-03-30] [WebCite Cache ID 77FcIDcn5]
56. Topaz M, Radhakrishnan K, Blackley S, Lei V, Lai K, Zhou L. Studying associations between heart failure self-management and rehospitalizations using natural language processing. West J Nurs Res 2017 Jan;39(1):147-165. [doi: 10.1177/0193945916668493] [Medline: 27628125]
57. Garvin JH, Kim Y, Gobbel GT, Matheny ME, Redd A, Bray BE, et al. Automating quality measures for heart failure using natural language processing: a descriptive study in the Department of Veterans Affairs. JMIR Med Inform 2018 Jan 15;6(1):e5 [FREE Full text] [doi: 10.2196/medinform.9150] [Medline: 29335238]
58. Vijayakrishnan R, Steinhubl SR, Ng K, Sun J, Byrd RJ, Daar Z, et al. Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record. J Card Fail 2014 Jul;20(7):459-464 [FREE Full text] [doi: 10.1016/j.cardfail.2014.03.008] [Medline: 24709663]
59. Watson AJ, O'Rourke J, Jethwani K, Cami A, Stern TA, Kvedar JC, et al. Linking electronic health record-extracted psychosocial data in real-time to risk of readmission for heart failure. Psychosomatics 2011;52(4):319-327 [FREE Full text] [doi: 10.1016/j.psym.2011.02.007] [Medline: 21777714]
60. Kasthurirathne SN, Dixon BE, Gichoya J, Xu H, Xia Y, Mamlin B, et al. Toward better public health reporting using existing off the shelf approaches: the value of medical dictionaries in automated cancer detection using plaintext medical data. J Biomed Inform 2017 Dec;69:160-176 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.008] [Medline: 28410983]
61. Osborne JD, Wyatt M, Westfall AO, Willig J, Bethard S, Gordon G. Efficient identification of nationally mandated reportable cancer cases using natural language processing and machine learning. J Am Med Inform Assoc 2016 Nov;23(6):1077-1084. [doi: 10.1093/jamia/ocw006] [Medline: 27026618]
62. Yim W, Kwan SW, Yetisgen M. Tumor reference resolution and characteristic extraction in radiology reports for liver cancer stage prediction. J Biomed Inform 2016 Dec;64:179-191 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.005] [Medline: 27729234]
63. Yim W, Kwan SW, Yetisgen M. Classifying tumor event attributes in radiology reports. J Assoc Inform Sci Technol 2017 Sep 14;68(11):2662-2674. [doi: 10.1002/asi.23937]
64. Wang H, Zhang W, Zeng Q, Li Z, Feng K, Liu L. Extracting important information from Chinese Operation Notes with natural language processing methods. J Biomed Inform 2014 Apr;48:130-136 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.017] [Medline: 24486562]
65. Jensen K, Soguero-Ruiz C, Oyvind MK, Lindsetmo R, Kouskoumvekaki I, Girolami M, et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. Sci Rep 2017 Dec 07;7:46226 [FREE Full text] [doi: 10.1038/srep46226] [Medline: 28387314]
66. Carrell DS, Halgrim S, Tran D, Buist DSM, Chubak J, Chapman WW, et al. Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. Am J Epidemiol 2014 Mar 15;179(6):749-758 [FREE Full text] [doi: 10.1093/aje/kwt441] [Medline: 24488511]
67. Napolitano G, Marshall A, Hamilton P, Gavin AT. Machine learning classification of surgical pathology reports and chunk recognition for information extraction noise reduction. Artif Intell Med 2016 Dec;70:77-83. [doi: 10.1016/j.artmed.2016.06.001] [Medline: 27431038]
68. Warner JL, Levy MA, Neuss MN, Warner JL, Levy MA, Neuss MN. ReCAP: feasibility and accuracy of extracting cancer stage information from narrative electronic health record data. J Oncol Pract 2016 Feb;12(2):157-169. [doi: 10.1200/JOP.2015.004622] [Medline: 26306621]
69. Zopf JJ, Langer JM, Boonn WW, Kim W, Zafar HM. Development of automated detection of radiology reports citing adrenal findings. J Digit Imaging 2012 Feb;25(1):43-49 [FREE Full text] [doi: 10.1007/s10278-011-9425-7] [Medline: 22042494]
70. Ou Y, Patrick J. Automatic structured reporting from narrative cancer pathology reports. 2014 Presented at: HIKM '14 Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management; 2014; Auckland.
71. Ou Y, Patrick J. Automatic negation detection in narrative pathology reports. Artif Intell Med 2015 May;64(1):41-50 [FREE Full text] [doi: 10.1016/j.artmed.2015.03.001] [Medline: 25990897]
72. Luo Y, Sohani AR, Hochberg EP, Szolovits P. Automatic lymphoma classification with sentence subgraph mining from pathology reports. J Am Med Inform Assoc 2014;21(5):824-832 [FREE Full text] [doi: 10.1136/amiajnl-2013-002443] [Medline: 24431333]
73. Ashish N, Dahm L, Boicey C. University of California, Irvine-Pathology Extraction Pipeline: the pathology extraction pipeline for information extraction from pathology reports. Health Informatics J 2014 Dec;20(4):288-305. [doi: 10.1177/1460458213494032] [Medline: 25155030]
74. Jouhet V, Defossez G, Burgun A, Le Beux P, Levillain P, Ingrand P, et al. Automated classification of free-text pathology reports for registration of incident cases of cancer. Methods Inf Med 2012;51(3):242-251. [doi: 10.3414/ME11-01-0005] [Medline: 21792466]
75. Afzal M, Hussain M, Khan WA, Ali T, Jamshed A, Lee S. Smart extraction and analysis system for clinical research. Telemed J E Health 2017 Dec;23(5):404-420. [doi: 10.1089/tmj.2016.0157] [Medline: 27782787]

76. Miller T, Dligach D, Bethard S, Lin C, Savova G. Towards generalizable entity-centric clinical coreference resolution. J Biomed Inform 2017 Dec;69:251-258 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.015] [Medline: 28438706]
77. Mykowiecka A, Marciniak M, Kupść A. Rule-based information extraction from patients' clinical data. J Biomed Inform 2009 Oct;42(5):923-936 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.007] [Medline: 19646551]
78. Bozkurt S, Lipson JA, Senol U, Rubin DL. Automatic abstraction of imaging observations with their characteristics from mammography reports. J Am Med Inform Assoc 2015 Apr;22(e1):e81-e92. [doi: 10.1136/amiajnl-2014-003009] [Medline: 25352567]
79. Denny JC, Choma NN, Peterson JF, Miller RA, Bastarache L, Li M, et al. Natural language processing improves identification of colorectal cancer testing in the electronic medical record. Med Decis Making 2012;32(1):188-197. [doi: 10.1177/0272989X11400418] [Medline: 21393557]
80. Denny JC, Peterson JF, Choma NN, Xu H, Miller RA, Bastarache L, et al. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. J Am Med Inform Assoc 2010;17(4):383-388 [FREE Full text] [doi: 10.1136/jamia.2010.004804] [Medline: 20595304]
81. Thomas AA, Zheng C, Jung H, Chang A, Kim B, Gelfond J, et al. Extracting data from electronic medical records: validation of a natural language processing program to assess prostate biopsy results. World J Urol 2014 Feb;32(1):99-103. [doi: 10.1007/s00345-013-1040-4] [Medline: 23417341]
82. Ping X, Tseng Y, Chung Y, Wu Y, Hsu C, Yang P, et al. Information extraction for tracking liver cancer patients' statuses: from mixture of clinical narrative report types. Telemed J E Health 2013 Sep;19(9):704-710. [doi: 10.1089/tmj.2012.0241] [Medline: 23869395]
83. Al-Haddad MA, Friedlin J, Kesterson J, Waters JA, Aguilar-Saavedra JR, Schmidt CM. Natural language processing for the development of a clinical registry: a validation study in intraductal papillary mucinous neoplasms. HPB (Oxford) 2010 Dec;12(10):688-695 [FREE Full text] [doi: 10.1111/j.1477-2574.2010.00235.x] [Medline: 21083794]
84. Garcelon N, Neuraz A, Benoit V, Salomon R, Burgun A. Improving a full-text search engine: the importance of negation detection and family history context to identify cases in a biomedical data warehouse. J Am Med Inform Assoc 2017 May 01;24(3):607-613. [doi: 10.1093/jamia/ocw144] [Medline: 28339516]
85. Chang N, Dai H, Jonnagaddala J, Chen C, Tsai RT, Hsu W. A context-aware approach for progression tracking of medical concepts in electronic medical records. J Biomed Inform 2015 Dec;58 Suppl:S150-S157 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.013] [Medline: 26432355]
86. Figueroa RL, Flores CA. Extracting information from electronic medical records to identify the obesity status of a patient based on comorbidities and bodyweight measures. J Med Syst 2016 Aug;40(8):191. [doi: 10.1007/s10916-016-0548-8] [Medline: 27402260]
87. Wright A, McCoy AB, Henkin S, Kale A, Sittig DF. Use of a support vector machine for categorizing free-text notes: assessment of accuracy across two institutions. J Am Med Inform Assoc 2013;20(5):887-890 [FREE Full text] [doi: 10.1136/amiajnl-2012-001576] [Medline: 23543111]
88. Mishra NK, Son RY, Arnzen JJ. Towards automatic diabetes case detection and ABCS protocol compliance assessment. Clin Med Res 2012 Aug;10(3):106-121 [FREE Full text] [doi: 10.3121/cmr.2012.1047] [Medline: 22634542]
89. 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]
90. Pakhomov SVS, Hanson PL, Bjornsen SS, Smith SA. Automatic classification of foot examination findings using clinical notes and machine learning. J Am Med Inform Assoc 2008 Mar;15(2):198-202 [FREE Full text] [doi: 10.1197/jamia.M2585] [Medline: 18096902]
91. Smith DH, Johnson ES, Russell A, Hazlehurst B, Muraki C, Nichols GA, et al. Lower visual acuity predicts worse utility values among patients with type 2 diabetes. Qual Life Res 2008 Dec;17(10):1277-1284. [doi: 10.1007/s11136-008-9399-1] [Medline: 18972222]
92. Wei W, Tao C, Jiang G, Chute CG. A high throughput semantic concept frequency based approach for patient identification: a case study using type 2 diabetes mellitus clinical notes. AMIA Annu Symp Proc 2010 Nov 13;2010:857-861 [FREE Full text] [Medline: 21347100]
93. Pakhomov S, Shah N, Hanson P, Balasubramaniam S, Smith SA, Smith SA. Automatic quality of life prediction using electronic medical records. AMIA Annu Symp Proc 2008 Nov 06:545-549 [FREE Full text] [Medline: 18998862]
94. Walsh JA, Shao Y, Leng J, He T, Teng C, Redd D, et al. Identifying axial spondyloarthritis in electronic medical records of US veterans. Arthritis Care Res (Hoboken) 2017 Dec;69(9):1414-1420 [FREE Full text] [doi: 10.1002/acr.23140] [Medline: 27813310]
95. Turner CA, Jacobs AD, Marques CK, Oates JC, Kamen DL, Anderson PE, et al. Word2Vec inversion and traditional text classifiers for phenotyping lupus. BMC Med Inform Decis Mak 2017 Aug 22;17(1):126 [FREE Full text] [doi: 10.1186/s12911-017-0518-1] [Medline: 28830409]

96. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. J Am Med Inform Assoc 2012 Jun;19(e1):e162-e169 [FREE Full text] [doi: 10.1136/amiajnl-2011-000583] [Medline: 22374935]
97. Chen W, Huang Y, Boyle B, Lin S. The utility of including pathology reports in improving the computational identification of patients. J Pathol Inform 2016;7:46 [FREE Full text] [doi: 10.4103/2153-3539.194838] [Medline: 27994938]
98. Soguero-Ruiz C, Hindberg K, Rojo-Alvarez JL, Skrovseth SO, Godtliebsen F, Mortensen K, et al. Support vector feature selection for early detection of anastomosis leakage from bag-of-words in electronic health records. IEEE J Biomed Health Inform 2016 Dec;20(5):1404-1415. [doi: 10.1109/JBHI.2014.2361688] [Medline: 25312965]
99. Chang EK, Yu CY, Clarke R, Hackbarth A, Sanders T, Esrailian E, et al. Defining a patient population with cirrhosis: an automated algorithm with natural language processing. J Clin Gastroenterol 2016;50(10):889-894. [doi: 10.1097/MCG.0000000000000583] [Medline: 27348317]
100. Shi X, Hu Y, Zhang Y, Li W, Hao Y, Alelaiwi A, et al. Multiple disease risk assessment with uniform model based on medical clinical notes. IEEE Access 2016;4:7074-7083. [doi: 10.1109/access.2016.2614541]
101. Geraci J, Wilansky P, de Luca V, Roy A, Kennedy JL, Strauss J. Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression. Evid Based Ment Health 2017 Jul 24;20(3):83-87 [FREE Full text] [doi: 10.1136/eb-2017-102688] [Medline: 28739578]
102. Roque F, Jensen P, Schmock H, Dalgaard M, Andreatta M, Hansen T, et al. Using electronic patient records to discover disease correlations and stratify patient cohorts. PLoS Comput Biol 2011 Aug;7(8):e1002141 [FREE Full text] [doi: 10.1371/journal.pcbi.1002141] [Medline: 21901084]
103. Ananda-Rajah MR, Martinez D, Slavin MA, Cavedon L, Dooley M, Cheng A, et al. Facilitating surveillance of pulmonary invasive mold diseases in patients with haematological malignancies by screening computed tomography reports using natural language processing. PLoS One 2014;9(9):e107797 [FREE Full text] [doi: 10.1371/journal.pone.0107797] [Medline: 25250675]
104. Baechle C, Agarwal A, Zhu X. Big data driven co-occurring evidence discovery in chronic obstructive pulmonary disease patients. J Big Data 2017 Apr 4;4(1). [doi: 10.1186/s40537-017-0067-6]
105. Gaskin GL, Pershing S, Cole TS, Shah NH. Predictive modeling of risk factors and complications of cataract surgery. Eur J Ophthalmol 2016 Jun 10;26(4):328-337 [FREE Full text] [doi: 10.5301/ejo.5000706] [Medline: 26692059]
106. Chase HS, Mitrani LR, Lu GG, Fulgieri DJ. Early recognition of multiple sclerosis using natural language processing of the electronic health record. BMC Med Inform Decis Mak 2017 Feb 28;17(1):24 [FREE Full text] [doi: 10.1186/s12911-017-0418-4] [Medline: 28241760]
107. Abhyankar S, Demner-Fushman D, Callaghan FM, McDonald CJ. Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. J Am Med Inform Assoc 2014;21(5):801-807 [FREE Full text] [doi: 10.1136/amiajnl-2013-001915] [Medline: 24384230]
108. Pivovarov R, Elhadad N. A hybrid knowledge-based and data-driven approach to identifying semantically similar concepts. J Biomed Inform 2012 Jun;45(3):471-481 [FREE Full text] [doi: 10.1016/j.jbi.2012.01.002] [Medline: 22289420]
109. Cronin R, Fabbri D, Denny J, Rosenbloom S, Jackson G. A comparison of rule-based and machine learning approaches for classifying patient portal messages. Int J Med Inform 2017 Dec;105:110-120 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.06.004] [Medline: 28750904]
110. Ni Y, Wright J, Perentesis J, Lingren T, Deleger L, Kaiser M, et al. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. BMC Med Inform Decis Mak 2015 Apr 14;15:28 [FREE Full text] [doi: 10.1186/s12911-015-0149-3] [Medline: 25881112]
111. Small AM, Kiss DH, Zlatsin Y, Birtwell DL, Williams H, Guerraty MA, et al. Text mining applied to electronic cardiovascular procedure reports to identify patients with trileaflet aortic stenosis and coronary artery disease. J Biomed Inform 2017 Dec;72:77-84 [FREE Full text] [doi: 10.1016/j.jbi.2017.06.016] [Medline: 28624641]
112. Zhou L, Lu Y, Vitale CJ, Mar PL, Chang F, Dhopeshwarkar N, et al. Representation of information about family relatives as structured data in electronic health records. Appl Clin Inform 2014;5(2):349-367 [FREE Full text] [doi: 10.4338/ACI-2013-10-RA-0080] [Medline: 25024754]
113. Chapman BE, Lee S, Kang HP, Chapman WW. Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. J Biomed Inform 2011 Oct;44(5):728-737 [FREE Full text] [doi: 10.1016/j.jbi.2011.03.011] [Medline: 21459155]
114. Lin FP, Pokorny A, Teng C, Epstein RJ. TEPAPA: a novel in silico feature learning pipeline for mining prognostic and associative factors from text-based electronic medical records. Sci Rep 2017 Jul 31;7(1):6918 [FREE Full text] [doi: 10.1038/s41598-017-07111-0] [Medline: 28761061]
115. Nath, Albaghdadi MS, Jonnalagadda SR. A natural language processing tool for large-scale data extraction from echocardiography reports. PLoS One 2016;11(4):e0153749 [FREE Full text] [doi: 10.1371/journal.pone.0153749] [Medline: 27124000]
116. Alnazzawi N, Thompson P, Ananiadou S. Mapping phenotypic information in heterogeneous textual sources to a domain-specific terminological resource. PLoS One 2016;11(9):e0162287 [FREE Full text] [doi: 10.1371/journal.pone.0162287] [Medline: 27643689]

117. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. J Biomed Inform 2017 Dec;73:14-29 [FREE Full text] [doi: 10.1016/j.jbi.2017.07.012] [Medline: 28729030]
118. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. Int J Methods Psychiatr Res 2016 Dec;25(2):86-100. [doi: 10.1002/mpr.1481] [Medline: 26184780]
119. Spasić I, Livsey J, Keane JA, Nenadić G. Text mining of cancer-related information: review of current status and future directions. Int J Med Inform 2014 Sep;83(9):605-623 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.009] [Medline: 25008281]
120. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural language processing in radiology: a systematic review. Radiology 2016 May;279(2):329-343. [doi: 10.1148/radiol.16142770] [Medline: 27089187]
121. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 2018 Apr;15(141) [FREE Full text] [doi: 10.1098/rsif.2017.0387] [Medline: 29618526]
122. Chiticariu L, Li Y, Reiss F. Rule-based information extraction is dead! Long live rule-based information extraction systems!. 2013 Presented at: Proc Conf Empir Methods Nat Lang Process EMNLP; 2013; Seattle.
123. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision-support system. JAMA 1987 Jul 03;258(1):67-74. [Medline: 3295316]

Abbreviations

BI-RADS: Breast Imaging-Reporting and Data System
CHF: congestive heart failure
CRF: conditional random field
DeepPhe: Cancer Deep Phenotype Extraction
EHR: electronic health record
EMR: electronic medical record
HF: heart failure
i2b2: Informatics for Integrating Biology and the Bedside
ICD: International Classification of Diseases, 10th Revision
IE: information extraction
MIMIC II: Multiparameter Intelligent Monitoring in Intensive Care II
NLP: natural language processing
PAD: peripheral arterial disease
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SVM: support vector machine
THYME: Temporal Histories of Your Medical Event

Edited by G Eysenbach; submitted 17.09.18; peer-reviewed by C Shivade, M Torii; comments to author 07.01.19; revised version received 04.03.19; accepted 24.03.19; published 27.04.19.

Please cite as:
Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V
Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review
JMIR Med Inform 2019;7(2):e12239
URL: http://medinform.jmir.org/2019/2/e12239/
doi:10.2196/12239
PMID:31066697

Original Paper

Physician Use of Electronic Health Records: Survey Study Assessing Factors Associated With Provider Reported Satisfaction and Perceived Patient Impact

Daniel Clay Williams[1], MSCR, MD; Robert W Warren[1], MPH, MD, PhD; Myla Ebeling[1], RA; Annie L Andrews[1], MSCR, MD; Ronald J Teufel II[1], MSCR, MD

Department of Pediatrics, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Daniel Clay Williams, MSCR, MD
Department of Pediatrics
Medical University of South Carolina
135 Rutledge Avenue MSC 561
Charleston, SC,
United States
Phone: 1 843 876 0794
Fax: 1 843 876 8709
Email: Willidan@musc.edu

Abstract

Background: The effect electronic health record (EHR) implementation has on physician satisfaction and patient care remains unclear. A better understanding of physician perceptions of EHRs and factors that influence those perceptions is needed to improve the physician and patient experience when using EHRs.

Objective: The objective of this study was to determine provider and clinical practice factors associated with physician EHR satisfaction and perception of patient impact.

Methods: We surveyed a random sample of physicians, including residents and fellows, at a US quaternary care academic hospital from February to March 2016. The survey assessed provider demographics, clinical practice factors (ie, attending, fellow, or resident), and overall EHR experience. The primary outcomes assessed were provider satisfaction and provider perceptions of impact to patient care. Responses on the satisfaction and patient impact questions were recorded on a continuous scale initially anchored at neutral (scale range 0 to 100: 0 defined as “extremely negatively” and 100 as “extremely positively”). Independent variables assessed included demographic and clinical practice factors, including perceived efficiency in using the EHR. One-way analysis of variance or the Kruskal-Wallis test was used for bivariate comparisons, and linear regression was used for multivariable modeling.

Results: Of 157 physicians, 111 (70.7%) completed the survey; 51.4% (57/111) of the respondents were attending physicians, and of those, 71.9% (41/57) reported a >50% clinical full-time-equivalency and half reported supervising residents >50% of the time. A total of 50.5% (56/111) of the respondents were primary care practitioners, previous EHR experience was evenly distributed, and 12.6% (14/111) of the total sample were EHR super-users. Responses to how our current EHR affects satisfaction were rated above the neutral survey anchor point (mean 58 [SD 22]), as were their perceptions as to how the EHR impacts the patient (mean 61 [SD 18]). In bivariate comparisons, only physician age, clinical role (resident, fellow, or attending), and perceived efficiency were associated with EHR satisfaction. In the linear regression models, physicians with higher reported perceived efficiency reported higher overall satisfaction and patient impact after controlling for other variables in the model.

Conclusions: Physician satisfaction with EHRs and their perception of its impact on clinical care were generally positive, but physician characteristics, greater age, and attending level were associated with worse EHR satisfaction. Perceived efficiency is the factor most associated with physician satisfaction with EHRs when controlling for other factors. Understanding physician perceptions of EHRs may allow targeting of technology resources to ensure efficiency and satisfaction with EHR system use during clinical care.

(JMIR Med Inform 2019;7(2):e10949) doi:10.2196/10949

KEYWORDS

electronic health records; user satisfaction; efficiency; physician survey

Introduction

Electronic health record (EHR) systems have been widely adopted in US hospitals in part due to financial incentive programs as well as the anticipated benefits of cost savings and improvements in safety and quality through a comprehensive approach to patient care [1-5]. Studies evaluating cost savings among hospitals after EHR adoption have had inconsistent findings [2,6]. In addition to potential cost savings, EHRs feature designs to improve patient safety through a variety of mechanisms such as real-time prompts during patient care encounters on drug dosing and potential medication error alerts [7-11]. While studies have shown that these prompts can prevent errors, they have also resulted in unintended consequences among providers such as alert fatigue [9-14].

Public perception of EHRs is generally positive but work by Emani et al [15,16] suggests physician skepticism might exist regarding the effect of meaningful use of EHRs on quality of care, patient-centeredness of care, and the patient care they personally provided. Although other studies regarding physician perceptions suggest EHRs may improve billing and quality, they also demonstrated skepticism regarding the impact on physician job satisfaction [17,18]. Decreased physician job satisfaction can lead to burnout, physician turnover, increased cost of physician recruitment, and potentially declines in quality of care [19-22]. While these studies have added to what is known, the lower response rates and focus on anticipated experience with meaningful use of EHR limits their generalizability and applicability to current provider practice.

The objective of this study was to determine provider and clinical practice factors, including perceived efficiency using the EHR, associated with physician EHR satisfaction and the perception of the EHR's impact on patient care. Ultimately, gaining insight into this complex issue may inform future efforts to improve physician efficiency and satisfaction with EHRs and optimize the positive impacts on patient safety and quality of care.

Methods

Setting

This descriptive survey study took place at the Medical University of South Carolina (MUSC) from February to March 2016. MUSC is a 700-bed quaternary care academic hospital that includes all adult and pediatric primary care and subspecialty services. MUSC Health manages over 1,000,000 outpatient encounters and 40,000 admissions annually through the employment of approximately 1200 physicians and 700 residents and fellows in 25 clinical departments.

MUSC currently uses Epic EHR software (Epic Systems Corporation), having adopted Epic outpatient systems in May 2012 and full Epic Enterprise, including all components of the fully integrated Epic health system, in June 2014. Initially developed in 1979, Epic is currently one of the most widely used EHR software systems worldwide. Epic is a fully integrated and encompassing EHR through which all health-related information is shared at MUSC. At MUSC, all physicians (attendings, residents, and fellows) are required to complete 8 hours of Epic training in a simulated practice environment prior to initial credentialing. During the implementation phase of the EHR, each clinical area had defined physician super-users, who were engaged in ongoing monthly interactive meetings with the Epic build team to stay up to date on relevant changes and new training updates, to help with immediate clinical and EHR needs of their respective areas.

Survey Assessment Tool

A team of EHR, clinical, and research experts (DW, RW, RJT) developed the survey content after a review of pertinent literature and key informant interviews with local stakeholders. The team piloted the survey for question clarity among a group of hospitalist physicians (n=8) and information technology medical directors that included physicians from a variety of pediatric and adult subspecialties (n=10). There were no content changes resulting from piloting, but several questions were clarified based on feedback (see Multimedia Appendix 1).

Respondent Sampling

We used a random number generator to identify a random sample of 157 physicians from a master list of all MUSC providers. The quantitative data analyzed for this project was part of a larger EHR satisfaction assessment project at the university that also included qualitative analysis of physician interviews. The qualitative interview data is not included in this analysis. Ten information technology medical directors were tasked with the entire project; thus, the final sample size was selected based on the ability of these 10 physicians to collect the data including completion of a face-to-face interview about the current EHR product. We used the qualitative data from the face-to-face interviews to drive improvement processes at the institution; we did not use the qualitative data for this analysis. We used Research Electronic Data Capture (REDCap) (REDCap Consortium) software for survey administration and data collection. We distributed surveys by email through REDCap, and the responses remained anonymous. Nonrespondents received email reminders from area specific medical directors. We did not incentivize or distribute reimbursements for survey completion. Our institution's institutional review board considered this project quality improvement.

Primary Outcomes

The primary outcomes assessed were provider satisfaction and provider perceived impact to patient care. We assessed provider satisfaction through the question, "How does Epic affect you overall?" We assessed impact to patient care through the question, "How does Epic affect your patients overall?" We recorded both question responses on a continuous scale ranging from 0 to 100 with 0 labeled as "extremely negatively" and 100 labeled as "extremely positively." We anchored the slide for

the response initially at a neutral value (50), and the survey respondents modified the answer from there.

Independent Variables

We assessed independent variables including physician demographics (age) and clinical practice factors. Clinical practice factors included clinical role (attending, resident, fellow), specialty department, percentage of clinical effort (reported clinical full-time-equivalent or cFTE), and percentage of attending providers whose encounters involved working with a trainee.

We also assessed perceived efficiency in using the EHR. We evaluated perceived provider efficiency through the statement, "Please rate your efficiency using Epic." Responses were recorded on a continuous scale from 0 to 100 with 0 labeled as "extremely inefficient" and 100 labeled as "extremely efficient." We anchored the slide for the response initially at a neutral value (50), and the survey respondents modified the answer from there.

The survey also evaluated physician EHR use experience (any EHR experience, any Epic experience, and Epic experience at MUSC in years of use), number of applications used in Epic, and Epic training above the standard eight hours (training as an Epic super-user).

Analysis Plan

We completed bivariate comparisons using one-way analysis of variance, and we calculated Pearson correlation coefficients for the continuous independent variables (perceived efficiency) for both outcome variables (satisfaction and perceived patient impact). The team also created linear regression models to predict reported provider satisfaction and perceived patient impact. A secondary analysis of factors associated with perceived efficiency was completed using one-way analysis of variance. All analyses were completed using SAS statistical software (SAS Institute).

Results

Of 157 randomly selected physicians, 111 (70.7%) completed the survey. An initial sample size of 160 was selected as described; however, 3 of the physicians randomly selected from the database were unable to respond due to temporary leave of absence (n=1) and recent retirement (n=2). A total of 51.3% (57/111) of the respondents were attending physicians, 32.4% (36/111) were residents, and 16.2% (18/111) were fellows. Mean age was 40.9 years (range 26 to 75 years) (Table 1). The mean age of the sample was similar to the mean age of all physicians at MUSC (39.8 years). A total of 50.5% (56/111) of the respondents were primary care practitioners.

Table 1. Survey respondent demographics (n=111).

Categorical variables	Value, n (%)
Age in years	
20-29	19 (17.1)
30-39	47 (42.3)
40-49	18 (16.2)
>50	27 (24.3)
Clinical role	
Attending	57 (51.4)
Clinical full-time-equivalent	
<0.5	16 (28.1)
0.5-0.99	13 (22.8)
1	28 (49.1)
Percentage of time supervising residents	
<50	29 (50.9)
51-99	18 (31.6)
100	10 (17.5)
Super-user	
No	45 (78.9)
Yes	12 (21.1)
Fellow	18 (16.2)
Resident	36 (32.4)
Postgraduate year (fellows and residents only; n=54)	
1	8 (14.8)
2	11 (20.3)
3	6 (11.1)
4	12 (22.2)
5	9 (16.7)
6	5 (9.3)
7	3 (5.6)
Clinical department	
Anesthesia	5 (4.5)
General medicine	34 (30.6)
Pediatrics	22 (19.8)
Psychiatry	11 (9.9)
Radiology	7 (6.3)
Medical subspecialty	21 (18.9)
Surgical subspecialty	11 (9.9)
Electronic health record experience in years	
<1	2 (1.8)
1-5	38 (34.2)
5-10	42 (37.8)
>10	29 (26.1)

Categorical variables	Value, n (%)
Epic experience in years	
<1	21 (18.9)
1-5	74 (66.7)
5-10	16 (14.4)
Epic experience at MUSC[a] in years	
<1	19 (17.1)
1-5	86 (77.5)
5-10	6 (5.4)
Number of systems used	
1	22 (19.8)
2	52 (46.8)
3	29 (26.1)
4	8 (7.2)

[a]MUSC: Medical University of South Carolina.

The overall mean response to the question assessing physician satisfaction demonstrated that physicians were generally satisfied (mean 58 [SD 22]), especially in light of the question being anchored at a neutral response of 50. Overall, physicians also felt that the EHR has a positive impact on the patient experience (mean 61 [SD 18]).

In the bivariate comparisons assessing categorical independent variables, only physician age and clinical role (resident, fellow, attending) were associated with satisfaction, with older age and attending role reporting lower satisfaction scores (both $P<.05$; Table 2).

Table 2. Bivariate analysis for the primary outcome variables provider satisfaction (EHR affects you) and patient impact (EHR affects patients). Numerical value represents the mean score for each group.

Categorical variables	Satisfaction mean	*P* value	Patient impact mean	*P* value
Total population score	57.8		60.6	
Age in years		**.05**		**.27**
20-29	67.5		65.8	
30-39	59.3		60.7	
40-49	49.1		54.4	
>50	54.2		60.8	
Clinical role		**.01**		**.21**
Attending (n=57)	52.1		57.9	
Clinical full-time-equivalent		**.88**		**.57**
<0.5	54.5		61.7	
0.5-0.99	50.9		58.3	
1.0	51.2		55.6	
Percentage of time supervising residents		**.07**		**.24**
<50	**52.9**		**61.4**	
51-99	44.0		52.2	
100	64.3		57.9	
Super-user		**.66**		**.39**
No	52.4		55.9	
Yes	50.8		65.3	
Fellow	66.7		65.6	
Resident	62.5		62.3	
Postgraduate year (fellows and residents only)		**.18**		**.37**
1	63.0		57.5	
2	76.1		70.7	
3	59.2		62.7	
4	57.4		61.0	
5	57.1		57.9	
6	63.8		64.4	
7	77.0		78.0	
Clinical department		**.22**		**.26**
Anesthesia	70.2		63.8	
General medicine	58.1		63.5	
Pediatrics	58.7		60.0	
Psychiatry	62.0		59.2	
Radiology	62.4		59.6	
Medical subspecialty	47.1		52.7	
Surgical subspecialty	62.8		68.5	
Electronic health record experience in years		**.11**		**.36**
<1	84.0		82.0	
1-5	61.7		59.3	
5-10	56.5		61.2	

Categorical variables	Satisfaction mean	*P* value	Patient impact mean	*P* value
>10	52.8		59.9	
Epic experience in years		**.12**		**.42**
<1	66.5		64.6	
1-5	55.8		59.1	
5-10	55.7		62.2	
Epic experience at MUSC[a] in years		**.27**		**.47**
<1	64.1		60.4	
1-5	57.0		61.2	
5-10	49.2		52.0	
Number of systems used		**.51**		**.15**
1	60.4		59.3	
2	57.6		62.6	
3	54.0		55.4	
4	65.8		69.4	

[a]MUSC: Medical University of South Carolina.

None of the assessed categorical variables were associated with perceived patient impact; however, physician reported perceived efficiency was correlated with both provider satisfaction (r=0.68) and perceived patient impact (r=0.6; Figures 1 and 2), indicating that physicians reporting higher perceived efficiency also reported higher overall satisfaction and EHRs having an overall more positive impact on the patient.

The regression model for physician satisfaction with predictors including clinical role, experience, and efficiency produced an R^2 of 0.5. As seen in Table 3, only perceived efficiency had a significant positive regression weight indicating that physicians with higher reported efficiency also reported higher overall satisfaction after controlling for other variables in the model. For every 1-point increase in efficiency, satisfaction scores increase by 0.74.

Figure 1. Scatter plots demonstrating the positive association between provider efficiency and satisfaction (r=0.68).

Figure 2. Scatter plot demonstrating the positive association between provider efficiency and provider reported patient impact (r=0.6).

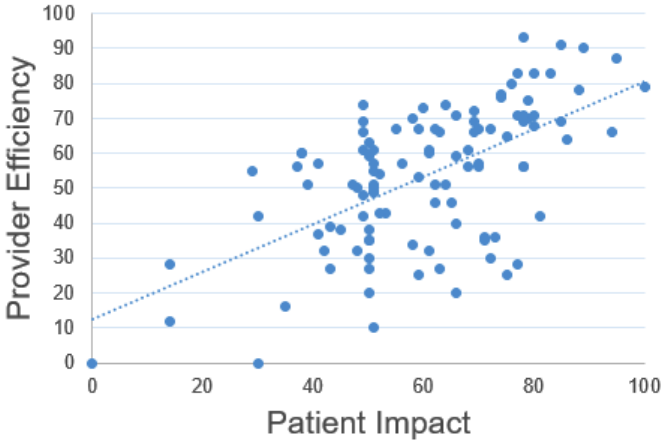


Table 3. Linear regression table predicting physician satisfaction: R^2=0.5; P<.001.

Variable	Parameter estimate	Standard error	*t* value	*P* value
Intercept	19.9	6.6	3.0	.003
Clinical role	1.7	1.8	1.0	.34
Epic experience at MUSC[a]	–0.7	0.8	0.8	.39
Efficiency	0.7	0.1	9.1	<.001

[a]MUSC: Medical University of South Carolina.

The regression model for physician-reported perceived patient impact, which also included clinical role, experience, and perceived efficiency as predictors, produced an R^2 of 0.4. As seen in Table 4, only perceived efficiency had a significant positive regression weight indicating that physicians with higher reported efficiency also reported higher positive perceived patient impact after controlling for other variables in the model. For every 1-point increase in efficiency, perception of patient impact increases by 0.53.

Because reported efficiency was the factor most predictive of both physician satisfaction and perceived patient impact, we conducted a secondary analysis of physician reported perceived efficiency. Overall, physician responses in rating their personal efficiency using EHR were positive (mean 54 [SD 20]). In bivariate comparisons, only clinical role was associated with perceived efficiency, with attending physicians reporting lower efficiency (Table 5).

Table 4. Linear regression table predicting perceived patient impact: R^2=0.4; P<.001.

Variable	Parameter estimate	Standard error	*t value*	*P value*
Intercept	32.00	5.90	5.40	<.001
Clinical role	–0.30	1.60	–0.20	.87
Epic experience at MUSC[a]	0.05	0.70	0.07	.94
Efficiency	0.53	0.07	7.70	<.001

[a]MUSC: Medical University of South Carolina.

Table 5. Bivariate analysis for factors associated with perceived efficiency.

Categorical variables	Efficiency, mean	*P* value
Total population perceived efficiency	53.9	
Age in years		.11
20-29	60.7	
30-39	56.2	
40-49	50.1	
>50	47.8	
Clinical role		.01
Attending	48.5	
Clinical full-time-equivalent		.28
<0.5	55.8	
0.5-0.99	46.6	
1.0	45.1	
Percentage of time supervising residents		.08
<50	48.6	
51-99	41.2	
100	60.7	
Super-user		.37
No	46.9	
Yes	54.3	
Fellow	63.8	
Resident	57.6	
Postgraduate year (fellows and residents only)		.05
1	53.1	
2	66.6	
3	56.5	
4	57.8	
5	52.7	
6	61.0	
7	85.0	
Clinical department		.86
Anesthesia	54.2	
General medicine	53.0	
Pediatrics	58.0	
Psychiatry	56.8	
Radiology	58.0	
Medical subspecialty	49.7	
Surgical subspecialty	50.9	
Electronic health record experience in years		.12
<1	83.0	
1-5	56.0	
5-10	53.6	
>10	49.8	

Categorical variables	Efficiency, mean	*P* value
Epic experience in years		.17
<1	60.1	
1-5	51.4	
5-10	57.3	
Epic experience at MUSC[a] in years		.47
<1	58.7	
1-5	53.2	
5-10	48.8	
Number of systems used		.67
1	49.7	
2	55.2	
3	55.7	
4	50.9	

[a]MUSC: Medical University of South Carolina.

Discussion

Principal Findings

This survey of physicians practicing at all levels of training and experience at a large academic medical center with a fully integrated and established EHR reveals that the EHR has had an overall positive influence on physician satisfaction with the EHR and perceived positive influence on patient care. Previous studies have reported physician concerns and challenges with provider use of an EHR system, such as those found in reports from Emani et al [15,16] and Shanafelt et al [20]. A 2012 study by Heyward et al [23] surveyed community-based clinicians before and after EHR implementation and found decreasing rates of overall job satisfaction among its providers. Our findings, in contrast to these reports, showed clinicians in a variety of clinical settings and practice types rated satisfaction with our EHR system positively and felt it has a positive impact on the care they provide.

Although older and attending-level physicians appeared more likely to report decreased satisfaction with EHR in our bivariate comparisons, it was their own perceived efficiency in using the EHR that was predictive of both satisfaction and positive impact for patients in adjusted analysis. We assessed perceived efficiency in the survey with the question, “Please rate your efficiency using Epic.” With perceived efficiency demonstrating the strongest association with physician satisfaction, we felt a more in-depth assessment of factors that influence perceived efficiency was needed. In a second bivariate analysis using efficiency as an outcome, only clinical role was associated with perceived efficiency. Attending physicians, when compared to residents and fellows, had the lowest overall perceived efficiency. This difference in perception of efficiency is likely multifactorial but could represent a true difference in how efficient attending-level providers are in using the EHR compared to their peers. To our knowledge, no validated measure exists to compare actual use efficiency.

Our diverse, randomly selected sample enables insight into perceptions of providers across disciplines and at various levels of training and EHR experience. For example, attending physicians and those practicing for a longer period of time may have had more experience with non-EHR systems and therefore may have a different perspective on EHR impact on patient care. Additionally, younger providers may have more general experience with technology. This experience may enable them to adapt easily to EHR adoptions, updates, or modifications.

We identified a trend (*P*=.06) in bivariate analysis toward increased time spent supervising residents with higher reported EHR satisfaction, but this was not significant. Attending physicians who supervise residents have less direct EHR use and responsibility. For example, physicians responsible for supervising residents more commonly cosign documentation as compared to writing notes and entering orders directly. Future studies may be needed to further evaluate the impact of resident supervision on attending use and satisfaction with EHRs.

Limitations

Our study has limitations. Although our response rate to the survey was above a commonly accepted benchmark (60%), it was only distributed at a single center. Although the EHR system used at this center is one of the most commonly adopted EHRs in the United States, it does have some degree of customizability and therefore our results may not be generalizable. Additionally, we chose our sample based on the number of surveys felt to be feasible to perform (N=160), and some of our associations that were close to significant may have become significant with a larger sample. This impact may have been greater when investigating subgroups such as attending only. Furthermore, our study assesses association and not causations. Due to the desire to keep the survey brief, other EHR-specific factors influencing satisfaction and clinician perception of the EHR’s impact on patient care were not assessed. This represents a potential area for further research. It is helpful that EHR implementation occurred between May 2012 and June 2014 as

many of the surveyed providers did have experience caring for patients without an EHR, but we did not measure important variables like efficiency before and after implementation as would be required to test causation. Last, further work is also needed to determine if perception of poor efficiency is correlated with actual efficiency. Unfortunately, no standard measure for efficiency exists with EHRs.

Conclusions

In our diverse sample of providers, perceived efficiency in using the institution's EHR was the factor most associated with both satisfaction and perceived impact to patient care. Targeting at-risk groups for training, efficiency improvement efforts, and continued monitoring especially during major upgrades may be needed to improve efficiency as a way to increase physician satisfaction and ensure high-quality patient care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Clinician survey tool.

[PDF File (Adobe PDF File), 38KB - medinform_v7i2e10949_app1.pdf]

References

1. Silow-Carroll S, Edwards JN, Rodin D. Using electronic health records to improve quality and efficiency: the experiences of leading hospitals. Issue Brief (Commonw Fund) 2012 Jul;17:1-40. [Medline: 22826903]
2. Kazley AS, Simpson AN, Simpson KN, Teufel R. Association of electronic health records with cost savings in a national sample. Am J Manag Care 2014 Jun 01;20(6):e183-e190 [FREE Full text] [Medline: 25180501]
3. Jamoom EW, Patel V, Furukawa MF, King J. EHR adopters vs. non-adopters: impacts of, barriers to, and federal initiatives for EHR adoption. Healthc (Amst) 2014 Mar;2(1):33-39 [FREE Full text] [doi: 10.1016/j.hjdsi.2013.12.004] [Medline: 26250087]
4. Teufel RJ, Kazley AS, Andrews AL, Ebeling MD, Basco WT. Electronic medical record adoption in hospitals that care for children. Acad Pediatr 2013;13(3):259-263. [doi: 10.1016/j.acap.2013.01.010] [Medline: 23680343]
5. Zlabek JA, Wickus JW, Mathiason MA. Early cost and safety benefits of an inpatient electronic health record. J Am Med Inform Assoc 2011;18(2):169-172 [FREE Full text] [doi: 10.1136/jamia.2010.007229] [Medline: 21292703]
6. Teufel RJ, Kazley AS, Ebeling MD, Basco WT. Hospital electronic medical record use and cost of inpatient pediatric care. Acad Pediatr 2012;12(5):429-435. [doi: 10.1016/j.acap.2012.06.004] [Medline: 22819201]
7. Rand CM, Blumkin A, Szilagyi PG. Electronic health record use and preventive counseling for US children and adolescents. J Am Med Inform Assoc 2014 Feb;21(e1):e152-e156 [FREE Full text] [doi: 10.1136/amiajnl-2013-002260] [Medline: 24013091]
8. Speedie SM, Park Y, Du J, Theera-Ampornpunt N, Bershow BA, Gensinger RA, et al. The impact of electronic health records on people with diabetes in three different emergency departments. J Am Med Inform Assoc 2014 Feb;21(e1):e71-e77 [FREE Full text] [doi: 10.1136/amiajnl-2013-001804] [Medline: 23842938]
9. Page N, Baysari MT, Westbrook JI. A systematic review of the effectiveness of interruptive medication prescribing alerts in hospital CPOE systems to change prescriber behavior and improve patient safety. Int J Med Inform 2017 Dec;105:22-30. [doi: 10.1016/j.ijmedinf.2017.05.011] [Medline: 28750908]
10. Galanter WL, Didomenico RJ, Polikaitis A. A trial of automated decision support alerts for contraindicated medications using computerized physician order entry. J Am Med Inform Assoc 2005;12(3):269-274 [FREE Full text] [doi: 10.1197/jamia.M1727] [Medline: 15684124]
11. Sorita A, Bos JM, Morlan BW, Tarrell RF, Ackerman MJ, Caraballo PJ. Impact of clinical decision support preventing the use of QT-prolonging medications for patients at risk for torsade de pointes. J Am Med Inform Assoc 2015 Apr;22(e1):e21-e27 [FREE Full text] [doi: 10.1136/amiajnl-2014-002896] [Medline: 25324555]
12. Stultz JS, Nahata MC. Appropriateness of commercially available and partially customized medication dosing alerts among pediatric patients. J Am Med Inform Assoc 2014 Feb;21(e1):e35-e42 [FREE Full text] [doi: 10.1136/amiajnl-2013-001725] [Medline: 23813540]
13. Strom BL, Schinnar R, Aberra F, Bilker W, Hennessy S, Leonard CE, et al. Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial. Arch Intern Med 2010 Sep 27;170(17):1578-1583. [doi: 10.1001/archinternmed.2010.324] [Medline: 20876410]
14. Zhou L, Soran CS, Jenter CA, Volk LA, Orav EJ, Bates DW, et al. The relationship between electronic health record use and quality of care over time. J Am Med Inform Assoc 2009;16(4):457-464 [FREE Full text] [doi: 10.1197/jamia.M3128] [Medline: 19390094]

15. Emani S, Ting DY, Healey M, Lipsitz SR, Karson AS, Einbinder JS, et al. Physician beliefs about the impact of meaningful use of the EHR: a cross-sectional study. Appl Clin Inform 2014;5(3):789-801 [FREE Full text] [doi: 10.4338/ACI-2014-05-RA-0050] [Medline: 25298817]
16. Emani S, Ting DY, Healey M, Lipsitz SR, Karson AS, Bates DW. Physician beliefs about the meaningful use of the electronic health record: a follow-up study. Appl Clin Inform 2017 Dec;8(4):1044-1053. [doi: 10.4338/ACI-2017-05-RA-0079] [Medline: 29241244]
17. Wylie MC, Baier RR, Gardner RL. Perceptions of electronic health record implementation: a statewide survey of physicians in Rhode Island. Am J Med 2014 Oct;127(10):e21-e27. [doi: 10.1016/j.amjmed.2014.06.011] [Medline: 24945882]
18. Williams C, Asi Y, Raffenaud A, Bagwell M, Zeini I. The effect of information technology on hospital performance. Health Care Manag Sci 2016 Dec;19(4):338-346. [doi: 10.1007/s10729-015-9329-z] [Medline: 26018176]
19. Ariely D, Lanier WL. Disturbing trends in physician burnout and satisfaction with work-life balance: dealing with malady among the nation's healers. Mayo Clin Proc 2015 Dec;90(12):1593-1596. [doi: 10.1016/j.mayocp.2015.10.004] [Medline: 26653295]
20. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clin Proc 2016 Jul;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
21. Shanafelt TD, Mungo M, Schmitgen J, Storz KA, Reeves D, Hayes SN, et al. Longitudinal study evaluating the association between physician burnout and changes in professional work effort. Mayo Clin Proc 2016 Apr;91(4):422-431. [doi: 10.1016/j.mayocp.2016.02.001] [Medline: 27046522]
22. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Arch Intern Med 2012 Oct 08;172(18):1377-1385. [doi: 10.1001/archinternmed.2012.3199] [Medline: 22911330]
23. Heyworth L, Zhang F, Jenter CA, Kell R, Volk LA, Tripathi M, et al. Physician satisfaction following electronic health record adoption in three Massachusetts communities. Interact J Med Res 2012 Nov 08;1(2):e12 [FREE Full text] [doi: 10.2196/ijmr.2064] [Medline: 23611987]

Abbreviations

cFTE: clinical full-time-equivalent
EHR: electronic health record
MUSC: Medical University of South Carolina
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 02.05.18; peer-reviewed by R Rubin, A Nguyen, Z Reis; comments to author 30.08.18; revised version received 10.12.18; accepted 31.12.18; published 04.04.19.

Please cite as:
Williams DC, Warren RW, Ebeling M, Andrews AL, Teufel II RJ
Physician Use of Electronic Health Records: Survey Study Assessing Factors Associated With Provider Reported Satisfaction and Perceived Patient Impact
JMIR Med Inform 2019;7(2):e10949
URL: https://medinform.jmir.org/2019/2/e10949/
doi:10.2196/10949
PMID:30946023

Original Paper

Clinical Requirements of Future Patient Monitoring in the Intensive Care Unit: Qualitative Study

Akira-Sebastian Poncette[1,2], MD; Claudia Spies[1], MD; Lina Mosch[1]; Monique Schieler[1]; Steffen Weber-Carstens[1], MD; Henning Krampe[1], PhD; Felix Balzer[1,2], MD, MSc, PhD

[1]Department of Anesthesiology and Intensive Care Medicine, Charité – Universitätsmedizin Berlin (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health), Berlin, Germany

[2]Einstein Center Digital Future, Berlin, Germany

Corresponding Author:
Felix Balzer, MD, MSc, PhD
Department of Anesthesiology and Intensive Care Medicine
Charité – Universitätsmedizin Berlin (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health)
Charitéplatz 1
Berlin,
Germany
Phone: 49 30450 ext 651166
Email: felix.balzer@charite.de

Abstract

Background: In the intensive care unit (ICU), continuous patient monitoring is essential to detect critical changes in patients' health statuses and to guide therapy. The implementation of digital health technologies for patient monitoring may further improve patient safety. However, most monitoring devices today are still based on technologies from the 1970s.

Objective: The aim of this study was to evaluate statements by ICU staff on the current patient monitoring systems and their expectations for future technological developments in order to investigate clinical requirements and barriers to the implementation of future patient monitoring.

Methods: This prospective study was conducted at three intensive care units of a German university hospital. Guideline-based interviews with ICU staff—5 physicians, 6 nurses, and 4 respiratory therapists—were recorded, transcribed, and analyzed using the grounded theory approach.

Results: Evaluating the current monitoring system, ICU staff put high emphasis on usability factors such as intuitiveness and visualization. Trend analysis was rarely used; inadequate alarm management as well as the entanglement of monitoring cables were rated as potential patient safety issues. For a future system, the importance of high usability was again emphasized; wireless, noninvasive, and interoperable monitoring sensors were desired; mobile phones for remote patient monitoring and alarm management optimization were needed; and clinical decision support systems based on artificial intelligence were considered useful. Among perceived barriers to implementation of novel technology were lack of trust, fear of losing clinical skills, fear of increasing workload, and lack of awareness of available digital technologies.

Conclusions: This qualitative study on patient monitoring involves core statements from ICU staff. To promote a rapid and sustainable implementation of digital health solutions in the ICU, all health care stakeholders must focus more on user-derived findings. Results on alarm management or mobile devices may be used to prepare ICU staff to use novel technology, to reduce alarm fatigue, to improve medical device usability, and to advance interoperability standards in intensive care medicine. For digital transformation in health care, increasing the trust and awareness of ICU staff in digital health technology may be an essential prerequisite.

Trial Registration: ClinicalTrials.gov NCT03514173; https://clinicaltrials.gov/ct2/show/NCT03514173 (Archived by WebCite at http://www.webcitation.org/77T1HwOzk)

(JMIR Med Inform 2019;7(2):e13064) doi:10.2196/13064

KEYWORDS

patient monitoring; digital health; qualitative research; intensive care unit; intensive care medicine; multidisciplinary; user-centered design; design thinking; digital literacy; grounded theory

Introduction

Background

In decades to come, demographic developments and an increasing number of comorbidities will lead to an ever-rising number of chronically ill patients in need of intensive care treatment [1]. Moreover, health care institutions are highly challenged with rising workloads, due to a shortage of medical staff and an increasing financial burden [2]. Within this context, rapid and sustainable implementation of advanced digital technologies could mitigate this development.

Continuous monitoring of patients is one of the most essential components in intensive care medicine: first, to notice critical changes of patients' health statuses, and second, to guide daily intensive care therapy [3]. Its implementation led to significant improvements in patient safety in the intensive care unit (ICU) [4]. Notably, in comparison with other medical devices, patient monitoring is used by a multidisciplinary team of physicians, nurses, and respiratory therapists.

With advances in information and communication technologies (ICTs) and medical device technologies, new options for patient monitoring are being introduced that may potentially improve patient safety [5]. However, most of the monitoring devices used today, such as the electrocardiogram (ECG) or invasive blood pressure measurement, were already available in the 1970s, using alarm thresholds for single sensors [6,7]. Nowadays, technologies to remotely monitor patients are available, such as wireless monitoring sensors (eg, ECG, pulse oximetry [8,9], and hemoglobin [10]), noninvasive measurement of hemodynamic parameters (eg, blood pressure and cardiac output [11]), as well as mobile communication devices (eg, mobile phones and tablets) [12-14]. Furthermore, clinical decision support systems (CDSS) based on artificial intelligence can assist physicians by analyzing multiple parameters to detect early indications of sepsis, respiratory failure, or bleeding [15,16].

Despite these technological developments, the introduction of novel patient monitoring applications in the ICU remains a lagging process compared to other industry sectors [17,18]. The manifold reasons for this could be rooted in a mismatch of expectations and assumptions by clinical users and manufacturers about novel patient monitoring [19,20].

Aim

This qualitative study evaluated statements by ICU staff—physicians, nurses, and respiratory therapists—on current patient monitoring. This study also evaluated the staff's expectations for future technological developments to explore clinical requirements and barriers to the implementation of a novel monitoring system. We aimed to explore desires, concerns, and perceived challenges of ICU staff on patient monitoring that may stimulate rapid and sustainable technological adaption in the ICU.

Methods

Ethics Approval and Consent to Participate

Ethical approval for this study was provided by the ethics committee of the Charite´—Universita¨tsmedizin Berlin, Germany (EA1/031/18). All participants gave their consent prior to the study.

Setting

This study was conducted at three ICUs of a German university hospital as a preliminary study of the implementation of the Vital Sync virtual patient monitoring platform 2.4, developed by Medtronic plc. This new system was installed in one of the three ICUs to monitor patients remotely and was utilized after completion of data collection for this study. In all three ICUs, the Philips IntelliVue patient monitoring system was installed at the time of the study (MX800 software version M.00.03; MMS X2 software version H.15.41-M.00.04). The COPRA 5 patient data management system (PDMS), developed by COPRA System GmbH, was used in all ICUs.

Research Team and Study Design

The research team consisted of a postdoctoral researcher with a background in anesthesiology, geriatrics, intensive care medicine, and digital health (ASP); a senior medical student with a strong affinity for digital health (LM); a professor for digital health, who is a consultant anesthesiologist and a computer scientist (FB); a psychologist (HK); a head nurse (MS); the ICU senior consultant (SWC); and the department's head of staff (CS). To maintain reflexivity, the research team challenged established assumptions in discussions and shared diaries throughout the study.

We chose an inductive, exploratory, qualitative research approach using semistructured interviews as described elsewhere [21-23]. The inductive approach allowed us to simultaneously collect and analyze data to see if any patterns emerged that would influence the study design.

Data Collection

Between April and May 2018, ASP and LM conducted face-to-face semistructured interviews with 5 physicians (4 women, 80%), 6 nurses (2 women, 33%), and 4 respiratory therapists (1 woman, 25%) from the ICU. The median of ICU experience was 4 years (range 2-15) for physicians, 6 years (range 1-14) for nurses, and 9 years (range 2-18) for respiratory therapists. Purposive sampling was employed to ensure an evenly distributed variety of professional staff.

The interview design was based on the research question and developed by the research team through consultation of further experts from intensive care medicine and psychology. Pilot interviews did not alter the questions. The developed questions were used as a guide for the interviews, giving the interviewers the freedom to change their weight or phrasing (see Textbox 1). Additionally, the order of the first three questions could be

changed. The interviews were conducted during breaks between patient care in the ICU, were recorded and transcribed verbatim by the interviewers, and were reviewed by the researcher who had not done the transcription. Median interview length was 13 minutes (range 8-26).

Data Analysis

After the completion of five interviews, we began analyzing the data through an inductive approach by means of the grounded theory [24]. Codes that were generated through line-by-line coding of three particularly different interviews resulted in a category system (see Multimedia Appendix 1) that was adjusted and extended by analyzing further interview transcripts (see Multimedia Appendix 2). All coding was performed using the MaxQDA 2018 qualitative data analysis software. The first five interviews were coded twice by two independent researchers (ASP and LM). Inconsistencies between coders were discussed in meetings among the research team until a mutual agreement was achieved. All following transcripts were coded by one researcher and the codes validated by another researcher.

After completion of coding, the research team reviewed and summarized each core statement to extract themes that were relevant to the study objective. Throughout the process of data analysis, the weight and phrasing of all questions and the order of the first three questions asked during the interviews were adapted using a feedback loop as previously described [25] (see Figure 1). Data collection was finalized when no new codes were identifiable from new interviews [26]. Out of each category, representative statements were selected and translated into English.

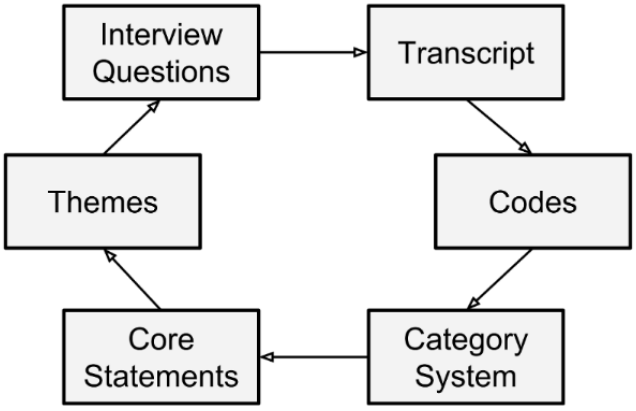
The datasets generated and analyzed during this study are not publicly available due to reasons of data privacy; however, they are available from the corresponding author (FB) upon reasonable request.

Textbox 1. Guide for intensive care staff interviews.

Interview questions:

- How often do you interact with the current patient monitoring system and which features do you use?
- Regarding the current patient monitoring system, is there anything that you find particularly useful? What suggestions for improvement do you have?
- Given endless financial and technical resources, what would your future patient monitoring system look like?
- Would you consider using a tablet for your clinical work regarding remote patient monitoring? In which situations would you use it?
- Would you consider using a clinical decision support system for your clinical routine?
- In your clinical workflow, is it important to have a graphical visualization of patients' vital parameters and their trends? Do you consider trend graphics of the patient monitoring system useful for shift handovers?
- What is more important to you: usability or number of features?

Figure 1. A feedback loop adapted the weight and order of the interview questions through parallel data collection and evaluation as previously described [25].



Results

Summary

This qualitative study was constructed based on 15 interviews with ICU staff regarding the complexity of patient monitoring in the ICU. According to our study objectives, resulting codes were classified into three main categories: (1) current patient monitoring, (2) future patient monitoring, and (3) barriers to implementation of novel patient monitoring. In the sunburst diagram (see Figure 2), the 12 most-relevant themes (middle ring) within the three categories (inner ring) are visualized and specified (outer ring).

Figure 2. Within three categories (inner ring), 12 themes (middle ring) were identified and specified (outer ring) to reflect the requirements of a novel patient monitoring technology from the view of intensive care staff. CDSS: clinical decision support system.

Most participants saw a need for improvement of patient monitoring in the ICU through novel technology, not only for enhanced efficiency in routine processes, but also to improve patient safety, quality of care, staff satisfaction, and quality of life for patients in the ICU as well as after discharge. Self-evaluation by participants regarding technological savviness using a Likert scale, with scores ranging from 1 (*no affinity for technology*) to 5 (*high affinity for technology*), resulted in the following median scores: physicians, 3.5 (range 2.0-5.0); nurses, 2.8 (range 1.5-4.0); and respiratory therapists, 3.8 (range 3.5-5.0).

Current Patient Monitoring

The interviewed ICU staff rated the software usability of the current patient monitoring as *good* with special emphasis on intuitiveness and uniformity. Standard features such as display of vital parameters and configuration of alarm thresholds were easy to use, however, advanced settings were considered difficult to set up without training.

> *It's sometimes very difficult to get all the parameters that I actually want on a monitor...Partly it's very complicated to be able to adjust the monitor quickly and effectively. So I often have the situation that I am called in by the nursing staff because they don't manage to display the parameters on the monitor that I would like to see. And then it costs me 20 minutes of work that is wasted during the day.* [Interview 13, physician]

For the visualization of single parameters, a graphical curve was stated to be essential for faster clinical interpretations and to ensure the validity of sensor measurements. All professional groups stated that they rarely use trend analysis on the patient monitoring device. Instead, the PDMS was used, as it provides other clinical data along with trends of vital parameters.

Concerning patient monitoring features used by ICU staff, alarm management was mentioned most frequently. Nurses and respiratory therapists would regularly adjust alarm thresholds

to current patient conditions. However, alarm fatigue or "cry wolf" situations (ie, multiple alarms going off at the same time) were considered as a major deficit of the current system, leading to stress in patients and staff and, potentially, reduced patient safety. Reasons for this were stated as (1) technical: difficult to distinguish between false and critical alarm, and susceptibility to error of the ECG, peripheral capillary *oxygen* saturation (SpO_2), and end-tidal carbon dioxide ($etCO_2$) sensors; (2) patient related: interference of artefacts related to delirium (ie, movement), sepsis (ie, centralized circulation), or high perspiration; and (3) ICU staff related: inadequate alarm hygiene due to lack of staff training with patient monitoring and lack of staff resources.

> *Alarm management is rather a big problem in the intensive care unit; some people set the alarm limits very tightly, which often leads to false alarms. I think it's important to work on the alarm management within the team...especially at night, also the sound for the patients. When the patient is supposed to sleep and then the monitor beeps all the time...* [Interview p02, nurse]

> *Too little alarm hygiene is being done. This is not due to the laziness of the people, but simply due to the staff situation; there are too few nurses, too few doctors. Therefore, it just beeps very often. And the monitor can't distinguish; is this critical or not? It gets its limits set, and if you've had an alarm five times because the patient is moving, and therefore the heart rate is supposedly elevated, you won't look at it the sixth time, but maybe there is something else. Yes, that's a bit of a problem, because one or the other critical situation is only recognized very late.* [Interview 11, respiratory therapist]

Long distances and an angled architecture of the ICU along with missing additional patient monitoring displays at strategic positions (eg, corridor and doctor's office) were indicated to possibly lead to critical situations. Furthermore, all interviewees criticized the entanglement of cables, especially in situations such as bedding and transport, posing a major patient safety issue.

Future Patient Monitoring

Participants from all professional groups emphasized the importance of intuitiveness and usability of a future patient monitoring system, especially in an emergency, with options to add more advanced and individual settings.

> *So if you want to use something like that, it would be good to have more functions and individualize it...Because, I think to myself, it is precisely because of the fact that there are so many different professional groups on the move here, that a senior physician in the department may also have completely different things that he finds important than perhaps a respiratory therapist or another specialist.* [Interview 12, respiratory therapist]

> *It all has to be self-explanatory in my eyes because we have too many devices that are complicated, so it would be nice if it was very user-friendly.* [Interview 7, respiratory therapist]

Future conceptions were more accurate in measurements, while at the same time less invasive, wireless, and with better interoperability between medical devices; for example, access to PDMS through patient monitoring.

> *How do you imagine the monitoring system of the future?* [Interview 11, interviewer]

> *Capture more values with less effort. So less invasive and a little more accurate, yes.* [Interview 11, respiratory therapist]

> *In any case, a wireless transmission of the monitor would be great. Because this would of course have a clear advantage for the patient in terms of mobility.* [Interview 12, respiratory therapist]

Participants from all interviewed professional groups believed that using mobile communication technology, such as tablet computers or mobile phones, as remote patient monitoring devices could increase patient safety, reduce the length of stay in the ICU, and improve job satisfaction.

> *I absolutely believe it [remote patient monitoring] is a step in the right direction. It benefits the patients, after all. And in the best case, it makes the work easier.* [Interview p02, nurse]

A reduction of stress through remote patient monitoring, in both ICU staff and patients, was pointed out and justified by optimized alarm management (ie, the possibility to cancel false alarms from a mobile device and, thus, less noise pollution).

> *And if I also had the option of canceling [false] alarms while sitting at the PC without having to run to the central system, I think that would make life easier for me. And above all, it would protect the patient. You do not ignore false alarms, or other alarms, which you interpret as false alarms—which can be life-threatening—and that the patient is perhaps less stressed, if he does not hear these alarms constantly at his own bed...I think I'm also preventing delirium.* [Interview 13, physician]

To reduce distractions of doctors by false alarms, interviewees also proposed an alarm filtering system by the nursing staff and critical alarm transmission to the doctor's mobile device.

> *If you get distracted by other things again and again...I think you accomplish less in the time you have. And, therefore, related to your question, of course it is important that you get alerted, but in the end, I see the nursing staff as a certain filter.* [Interview 2, physician]

> *For [external staff and new staff members], I actually don't find that bad at all. That they can just say, "Ok, I press a button and know...when the alarm comes, that goes to the doctor..." And that this makes them more relaxed and they don't have to search for him.* [Interview 8, nurse]

A point of criticism of remote patient monitoring was the fear of less interprofessional communication and less patient contact

when the physician is informed via a mobile device and the alarms are canceled remotely. To achieve better teamwork regarding alarm management, training in interprofessional communication was considered necessary.

> *I also find that a bit difficult, because then the communication just breaks down a bit. Because I like to go to the doc and say, "Hey, here, I noticed that, should I do something now?"* [Interview 8, nurse]

Staff expectations regarding the implementation of a CDSS, including artificial intelligence in monitoring, were ambivalent; however, an automatic adjustment of alarm thresholds through trend analysis and the CDSS was suggested. Critical attitudes resulted from lack of trust in the CDSS: the interviewees stressed plausibility to estimate the validity of CDSS recommendations in their clinical work routine.

> *And if I don't understand the physiology behind it, also in humans, and only stick to these theoretically calculated values there, then I think mistakes will occur...So a basic education in the basic understanding of physiology and also of technology, how these limits and parameters and recommendations arise, should be absolutely there.* [Interview 13, physician]

In terms of hardware design for remote patient monitoring, several interviewees of all professions agreed that a large tablet was applicable for stationary use because it would provide a better overview. However, most of the interviewed staff said they would prefer using a small device, even their own mobile phone, which would offer greater mobility since the pockets of the scrubs are too small for larger devices.

> *If I had to carry it [the tablet] with me all the time, then it would have to be the size of a scrubs pocket.* [Interview 3, nurse]

> *If it is stationary, then rather large [display] to provide a good overview.* [Interview 8, nurse]

Barriers to Implementation of Novel Patient Monitoring

We identified a lack of trust in technology as the greatest barrier to the implementation of novel patient monitoring devices in the ICU.

> *I think it's important to be at the patient's bedside, look at the patient, and not just rely on some kind of monitoring.* [Interview 10, physician]

ICU staff feared the implementation of new technology in the ICU that would increase workloads in a setting where time and resources are already scarce.

> *We have a lot of leasing staff [external staff], and we are a newly assembled team—I think it [new technology] would still be difficult to implement here at the moment.* [Interview p02, nurse]

They demanded more time for using advanced features and for training in new medical devices.

> *If I had more time, then I would like to have more functions [in patient monitoring] and we must be trained more intensively for using the new [medical] devices.* [Interview p02, nurse]

While satisfied with the current system, ICU staff reported that new technology seems very complex and they often did not foresee its benefit. By using new technology, they were afraid to lose their clinical skills and have less direct contact with the patient.

> *I think that we should use our brain, and that it makes sense to be able to rely on your own senses in case of a power failure, darkness, or whatever.* [Interview 10, physician]

> *Well, I think that the more you get taken off [by technology], the more you stop thinking. And then an ECG electrode falls off, and people think the patient is asystolic and start to resuscitate.* [Interview 4, nurse]

Additionally, lack of awareness and education of ICU staff about current technological developments was identified as a potential barrier to implementation.

Discussion

Principal Findings

This qualitative interview study provides valuable insights into the understanding of the complexity of patient monitoring in the ICU. For the ICU staff, the current patient monitoring system was intuitive to use for vital sign monitoring, but other features were difficult to set up due to lack of training and staff shortage. Further, ICU staff rated alarm fatigue and entanglement of cables as major threats to patient safety.

For future developments, a more interoperable, intuitive patient monitoring system was demanded with options to add advanced and individual features depending on the patients' or users' needs. Vital parameter measurements and alarms should be more specific, while being noninvasive and less obtrusive (eg, wireless). Interestingly, interviewees recognized mobile phones with a large screen as a potential remote patient monitoring device, which could reduce noise pollution, increase patient safety, and lead to enhanced job satisfaction. Additionally, a CDSS based on artificial intelligence could optimize alarm management if plausible for the ICU staff. For a more rapid introduction of novel patient monitoring solutions in the ICU, participants demanded more training in new medical devices.

As a major barrier to the implementation of novel patient monitoring, lack of both trust and awareness for novel, innovative technology was identified. Interviewees also admitted to being afraid to lose their clinical skills as a result of having less interprofessional communication and less contact with the patient due to novel patient monitoring technology.

False Alarms Endanger Patient Safety

Whereas alarm management is the main feature of patient monitoring used at the study sites, currently neither regular staff training nor a framework for alarm management is established. In the context where "cry wolf" situations with multiple alarms going off at the same time have become the standard environment in the ICU, this is an alarming insight [27]. Of all

auditory alarms, up to 99% have been described to be false alarms that do not change patient treatment [28]. These false alarms are a product of a complex interplay between the patient's condition, the users' competence, and the technical features of the patient monitoring system. False alarms desensitize clinical staff to critical alarms (ie, alarm fatigue) and pose a major patient safety issue, leading to alarm-related patient deaths every year [29]. According to our study results, patient safety might also be compromised through the constant noise pollution that induces interruptions, stress, and concentration difficulties among the ICU staff.

Although several strategies have been developed to reduce false alarms in the ICU [12,28-31], implementation into a clinical routine is still lacking. Notably, the reduction of alarms due to alarm management strategies ranges from 24% to 88.5% per ICU, indicating the effectiveness of such strategies, including staff training for any ward that uses patient monitoring devices [32-34].

Interoperability and Usability of Devices in Intensive Care

Today, most acute care medical devices are not designed to interoperate [18]. Remarkably, our results indicate that requirements for future patient monitoring are steadily increasing to more than just monitoring the vital parameters. ICU staff demand a patient monitoring device to interoperate with other medical devices for detailed comparisons of vital parameters and trend analysis in the context of medication, ventilation, fluid balance, and more, as recently suggested by Flohr et al [35]. This could optimize workflow and reduce redundant documentation in the ICU.

In terms of usability, ICU staff expressed their demand for intuitive and reactive systems for clinical use. Although the implementation of electronic applications in health care dates back more than a decade, usability—referring to the efficient, effective, and safe use of technology—is still not fully optimized for clinical use [36,37]. In the ICU, digital applications should not induce stress. Instead, their use should focus the user for efficient, effective, and safe work. In usability research, various simple and low-cost methods are available that should be applied by anyone working in medical device development [38].

For both interoperability and usability, regular adaptation and application of medical device communication (ie, Institute of Electrical and Electronics Engineers [IEEE] 11073) and technical standards (ie, International Electrotechnical Commission [IEC] 60601) to current developments might minimize use-related hazards and risks to patients and ICU staff [39,40].

Mobile Phones in Intensive Care Routine

The use of tablet computers with access to electronic medical records or multiparameter monitoring has been perceived as beneficial in inpatient settings [35,41]. However, for ICU staff, large tablets were too bulky to carry around due to the small pockets of their scrubs; they instead preferred small tablets that are portable [42] or larger mobile phones for remote patient monitoring in the ICU. This finding may influence further device developments for the ICU and the operating room where scrubs are worn. Recently released foldable mobile phones could be an approach to combine the advantages of pocket-size and large-screen devices [43]. As industry stakeholders are already developing apps for mobile devices in the ICU, more interdisciplinary studies are necessary to obtain early feedback from clinicians, developers, and engineers [12,14].

In the move toward a widespread implementation of telemedicine and remote patient monitoring technology into various health care sectors including the ICU, the mobile phone or tablet computer could easily be deployed for these tasks. ICU staff claimed that the length of stay in the ICU could be reduced through the utilization of remote patient monitoring, which is in line with several recent studies on telemedicine [44,45].

Clinical Decision Support Systems for Alarm Management

Integration of novel medical devices and technological advances result in a steadily growing amount of data that are being analyzed by ICU staff daily, thus making automated systems based on artificial intelligence a necessity for the future. Although various research projects are focusing on CDSS in the ICU, translation into the clinical routine is lagging far behind [15,46-49].

In our study, participating staff stated that they would utilize a CDSS only if it was plausible and underlying algorithms were readily understandable. A physician also indicated that appropriate training for the application would be useful to avoid misuse. Taking into account that most CDSS are based on complex machine learning methods, explaining the underlying mechanism to intensivists might be challenging. However, participants expressed the necessity to optimize detection of false alarms with a CDSS. Thus, a self-learning alarm system via machine learning might be practicable for the near future [50].

Furthermore, according to interviewees, trend-based alarms might be a useful complement to the traditional threshold-based alarms; this is consistent with a publication by Charbonnier et al, who was able to reduce 33% of false alarms by using a trend-based alarm system in the ICU [51].

Building Trust in Information and Communication Technology

The most disruptive implementation of ICT in intensive care medicine in the recent past has been the introduction of tele-ICUs, which has been accompanied by several staff acceptance studies [21,52,53]. With the implementation of tele-ICU technology in existing ICUs, ICU staff are not only confronted with novel ICTs, but also with changes in clinical processes, such as teamwork, communication, and staff structure. This is due to the fact that therapy decisions are influenced by external experts, who might be unfamiliar to the ICU staff on site. In this constellation, trust has to be formed first toward the new ICT and in a second step toward the external experts [21]. With respect to our study, similar concerns were reported: after trust in ICTs are established, ICU staff must also get familiar with the CDSS, in contrast to the external (ie, human) experts. Notably, our results did not show any influence

of prior experience with technology on the formation of trust [54].

We conclude that ICU staff are ready and willing to use more-advanced ICT devices in intensive care routine. Nevertheless, without adequate and regular training in novel technical and digital devices, even in alarm management, the full potential of digitization will not have been exhausted.

Digital Literacy

As suggested in recent publications, governments, health care institutions, and universities should include digital health care in the curriculum of high schools, as well as in medical and nursing schools, to ensure that future health care professionals acquire digital literacy [55,56]. Our finding of low tech-savviness among ICU staff indicates that regular staff training with novel medical devices, software, and mobile phone apps may be beneficial for successful implementation of future patient monitoring devices [20,57,58].

Innovation in health care derives from interdisciplinary teamwork with developers and medical engineers [59]. University hospitals, especially, should empower ICU staff to pursue academic research in the context of ICT implementation in the ICU.

Design Thinking in Health Care

In the context of digitization in health care, novel digital systems often fail after implementation as a result of a lack of user involvement [59]. The importance of validation of novel digital health solutions through early and continuous user involvement is often underestimated by the industry, hospitals, and governments [55]. Reasons for this include lack of financial resources, delays in time to market, or ignorance about how to validate a digital health product [59]. One way to mitigate this issue might be the design-thinking framework as a systematic process that prioritizes empathy for the users with the aim to develop a more comprehensive and effective solution [60]. In situations where the users cannot point out their needs, analyzing their behaviors through a more user-centered qualitative method such as design thinking can provide invaluable insights about their unmet desires [60].

Limitations

Through the use of a qualitative interview study design, we could identify several novel findings on the themes of patient monitoring from the perspective of ICU staff. However, as a descriptive approach, quantification of statements is not possible by design. When interpreting the results, it is crucial to take into account the small number of participants of a single hospital (ie, three ICUs) and possible biases due to the selection of participants. This makes the generalization to other hospital settings or countries difficult. A follow-up, quantitative, survey-based study with a larger cohort may be conducted on the basis of this study to further consolidate the results.

Moreover, it is not possible to draw conclusions about whether a novel patient monitoring system can improve patients' quality of life or quality of care in the ICU. Interdisciplinary investigations with patients, their relatives, health care providers, and technicians (ie, IT and engineering) might shed light on this question. Finally, a bias due to the implementation of the Vital Sync virtual patient monitoring platform cannot be excluded with certainty.

Conclusions

This qualitative study involves core statements by ICU staff in the analysis of current and novel patient monitoring applications in the ICU. In order to introduce more sustainable digital health solutions in the ICU, health care stakeholders might have to focus more on user-derived findings than top-down speculations. By valuing the opinions of health care providers, we may gain their trust to implement novel systems.

In particular, the results on alarm management and mobile devices in the ICU may be used (1) by health care organizations to prepare ICU staff for digital transformation, (2) by research institutes to reduce alarm fatigue, (3) by industry players to embrace medical device usability, and (4) by political stakeholders and decision makers to advance interoperability standards in intensive care medicine.

Our findings should motivate other researchers to conduct qualitative patient- and user-centered research in health care, especially before developing or implementing premature technological solutions.

Acknowledgments

The authors would like to thank Nicolas Fromm, Sabine Gehrke-Beck, and Niklas Keller for supporting the conception of the study design and the data analysis with their backgrounds in social sciences, general medicine, and psychology, respectively. The authors would also like to thank the staff from the intensive care units for their participation in this study. This study was partially supported by an unrestricted grant from Einstein Center Digital Future. We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité—Universitätsmedizin Berlin.

Conflicts of Interest

CS and FB report funding from Medtronic. The other authors do not declare a conflict of interest.

Authors' Contributions

CS had the idea for shared decision allocation and initiated the implementation of remote patient monitoring in the intensive care unit. The study was conceived by ASP, CS, and FB. ASP and LM conducted data acquisition and analyses, supported by MS and SWS, who provided perspectives from clinical routine and management. ASP wrote the manuscript with support from LM. HK

contributed to the study's methodology and interpretation of results from a psychologist's point of view. FB supervised all parts of the study. All authors critically reviewed and approved the manuscript.

Multimedia Appendix 1

Category system that was constructed through line-by-line coding of the interview transcripts.

[PDF File (Adobe PDF File), 184KB - medinform_v7i2e13064_app1.pdf]

Multimedia Appendix 2

Catalog with quotes from intensive care unit (ICU) staff regarding patient monitoring.

[PDF File (Adobe PDF File), 215KB - medinform_v7i2e13064_app2.pdf]

References

1. Zimmerman JE, Kramer AA, Knaus WA. Changes in hospital mortality for United States intensive care unit admissions from 1988 to 2012. Crit Care 2013 Apr 27;17(2):R81 [FREE Full text] [doi: 10.1186/cc12695] [Medline: 23622086]
2. Gulland A. Shortage of health workers is set to double, says WHO. BMJ 2013 Nov 12;347:f6804. [doi: 10.1136/bmj.f6804] [Medline: 24222676]
3. Kipnis E, Ramsingh D, Bhargava M, Dincer E, Cannesson M, Broccard A, et al. Monitoring in the intensive care. Crit Care Res Pract 2012;2012:473507 [FREE Full text] [doi: 10.1155/2012/473507] [Medline: 22970356]
4. Moreno RP, Rhodes A, Donchin Y, European Society of Intensive Care. Patient safety in intensive care medicine: The Declaration of Vienna. Intensive Care Med 2009 Oct;35(10):1667-1672 [FREE Full text] [doi: 10.1007/s00134-009-1621-2] [Medline: 19697007]
5. Michard F. Hemodynamic monitoring in the era of digital health. Ann Intensive Care 2016 Dec;6(1):15 [FREE Full text] [doi: 10.1186/s13613-016-0119-7] [Medline: 26885656]
6. Drews FA. Patient monitors in critical care: Lessons for improvement. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in Patient Safety: New Directions and Alternative Approaches. Volume 3: Performance and Tools. Rockville, MD: Agency for Healthcare Research and Quality; Aug 2008.
7. Thomas LJJ, Blaine GJ, Gerth VW, Hagen RW. Continuous Monitoring of Physiologic Variables with a Dedicated Minicomputer. Computer 1975 Jul;8(7):30-35 [FREE Full text] [doi: 10.1109/C-M.1975.219022]
8. Weller RS, Foard KL, Harwood TN. Evaluation of a wireless, portable, wearable multi-parameter vital signs monitor in hospitalized neurological and neurosurgical patients. J Clin Monit Comput 2018 Oct;32(5):945-951. [doi: 10.1007/s10877-017-0085-0] [Medline: 29214598]
9. Breteler MJM, Huizinga E, van Loon K, Leenen LPH, Dohmen DAJ, Kalkman CJ, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: A clinical validation study. BMJ Open 2018 Dec 27;8(2):e020162 [FREE Full text] [doi: 10.1136/bmjopen-2017-020162] [Medline: 29487076]
10. Baulig W, Seifert B, Spahn DR, Theusinger OM. Accuracy of non-invasive continuous total hemoglobin measurement by Pulse CO-Oximetry in severe traumatized and surgical bleeding patients. J Clin Monit Comput 2017 Feb;31(1):177-185. [doi: 10.1007/s10877-015-9816-2] [Medline: 26686689]
11. Balzer F, Habicher M, Sander M, Sterr J, Scholz S, Feldheiser A, et al. Comparison of the non-invasive Nexfin® monitor with conventional methods for the measurement of arterial blood pressure in moderate risk orthopaedic surgery patients. J Int Med Res 2016 Aug;44(4):832-843 [FREE Full text] [doi: 10.1177/0300060516635383] [Medline: 27142436]
12. Hoffmann R, Michaelsen J, Langenbrink L, Kastrati M, Piatkowski M, Hengemuhle G, et al. A novel ECG ward telemetry system with smartphone based alarm escalation. Int J Cardiovasc Res 2018;7(2):1-4. [doi: 10.4172/2324-8602.1000345]
13. Michard F, Pinsky MR, Vincent J. Intensive care medicine in 2050: NEWS for hemodynamic monitoring. Intensive Care Med 2017 Mar;43(3):440-442. [doi: 10.1007/s00134-016-4674-z] [Medline: 28124086]
14. Michard F, Barrachina B, Schoettker P. Is your smartphone the future of physiologic monitoring? Intensive Care Med 2018 Oct 19:1-3. [doi: 10.1007/s00134-018-5419-y] [Medline: 30341565]
15. Meyer A, Zverinski D, Pfahringer B, Kempfert J, Kuehne T, Sündermann SH, et al. Machine learning for real-time prediction of complications in critical care: A retrospective study. Lancet Respir Med 2018 Dec;6(12):905-914. [doi: 10.1016/S2213-2600(18)30300-X] [Medline: 30274956]
16. Komorowski M, Celi L, Badawi O, Gordon A, Faisal A. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018 Nov;24(11):1716-1720. [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
17. Hüsers J, Hübner U, Esdar M, Ammenwerth E, Hackl WO, Naumann L, et al. Innovative power of health care organisations affects IT adoption: A bi-national health IT benchmark comparing Austria and Germany. J Med Syst 2017 Feb;41(2):33. [doi: 10.1007/s10916-016-0671-6] [Medline: 28054195]

18. De Georgia MA, Kaffashi F, Jacono F, Loparo K. Information technology in critical care: Review of monitoring and data acquisition systems for patient care and research. ScientificWorldJournal 2015;2015:727694 [FREE Full text] [doi: 10.1155/2015/727694] [Medline: 25734185]
19. Heitmueller A, Bull A, Oh S. Looking in the wrong places: Why traditional solutions to the diffusion of innovation will not work. BMJ Innov 2016 Mar 25;2(2):41-47. [doi: 10.1136/bmjinnov-2015-000106]
20. Christodoulakis C, Asgarian A, Easterbrook S. Barriers to adoption of information technology in healthcare. In: Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering (CASCON '17). Riverton, NJ: IBM Corp; 2017 Presented at: 27th Annual International Conference on Computer Science and Software Engineering (CASCON '17); November 6-8, 2017; Markham, ON p. 66-75 URL: http://dl.acm.org/citation.cfm?id=3172795.3172804
21. Moeckli J, Cram P, Cunningham C, Reisinger HS. Staff acceptance of a telemedicine intensive care unit program: A qualitative study. J Crit Care 2013 Dec;28(6):890-901. [doi: 10.1016/j.jcrc.2013.05.008] [Medline: 23906904]
22. Wilson ME, Rhudy LM, Ballinger BA, Tescher AN, Pickering BW, Gajic O. Factors that contribute to physician variability in decisions to limit life support in the ICU: A qualitative study. Intensive Care Med 2013 Jun;39(6):1009-1018. [doi: 10.1007/s00134-013-2896-x] [Medline: 23559079]
23. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
24. Strauss A, Corbin J. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. 2nd edition. Thousand Oaks, CA: SAGE Publications; 1990.
25. Kean S, Salisbury L, Rattray J, Walsh T, Huby G, Ramsay P. 'Intensive care unit survivorship': A constructivist grounded theory of surviving critical illness. J Clin Nurs 2017 Oct;26(19-20):3111-3124. [doi: 10.1111/jocn.13659] [Medline: 27875013]
26. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: Guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
27. Donchin Y, Seagull F. The hostile environment of the intensive care unit. Curr Opin Crit Care 2002 Aug;8(4):316-320. [Medline: 12386492]
28. Paine C, Goel V, Ely E, Stave C, Stemler S, Zander M, et al. Systematic review of physiologic monitor alarm characteristics and pragmatic interventions to reduce alarm frequency. J Hosp Med 2016 Feb;11(2):136-144 [FREE Full text] [doi: 10.1002/jhm.2520] [Medline: 26663904]
29. Ruskin KJ, Hueske-Kraus D. Alarm fatigue: Impacts on patient safety. Curr Opin Anaesthesiol 2015 Dec;28(6):685-690. [doi: 10.1097/ACO.0000000000000260] [Medline: 26539788]
30. Sowan AK, Tarriela AF, Gomez TM, Reed CC, Rapp KM. Nurses' perceptions and practices toward clinical alarms in a transplant cardiac intensive care unit: Exploring key issues leading to alarm fatigue. JMIR Hum Factors 2015 Mar 16;2(1):e3 [FREE Full text] [doi: 10.2196/humanfactors.4196] [Medline: 27025940]
31. ClinicalTrials.gov. 2018 Apr 09. Outcomes and user acceptance of the IntelliVue alarm advisor software (Europe) URL: https://clinicaltrials.gov/ct2/show/NCT03182452 [accessed 2018-11-29] [WebCite Cache ID 74IIblp5w]
32. Turmell JW, Coke L, Catinella R, Hosford T, Majeski A. Alarm fatigue: Use of an evidence-based alarm management strategy. J Nurs Care Qual 2017;32(1):47-54. [doi: 10.1097/NCQ.0000000000000223] [Medline: 27500694]
33. Sendelbach S, Wahl S, Anthony A, Shotts P. Stop the noise: A quality improvement project to decrease electrocardiographic nuisance alarms. Crit Care Nurse 2015 Aug;35(4):15-22; quiz 1p following 22 [FREE Full text] [doi: 10.4037/ccn2015858] [Medline: 26232798]
34. Sowan AK, Gomez TM, Tarriela AF, Reed CC, Paper BM. Changes in default alarm settings and standard in-service are insufficient to improve alarm fatigue in an intensive care unit: A pilot project. JMIR Hum Factors 2016 Jan 11;3(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.5098] [Medline: 27036170]
35. Flohr L, Beaudry S, Johnson KT, West N, Burns CM, Ansermino JM, et al. Clinician-driven design of VitalPAD: An intelligent monitoring and communication device to improve patient safety in the intensive care unit. IEEE J Transl Eng Health Med 2018;6:3000114 [FREE Full text] [doi: 10.1109/JTEHM.2018.2812162] [Medline: 29552425]
36. Fidler R, Bond R, Finlay D, Guldenring D, Gallagher A, Pelter M, et al. Human factors approach to evaluate the user interface of physiologic monitoring. J Electrocardiol 2015;48(6):982-987. [doi: 10.1016/j.jelectrocard.2015.08.032] [Medline: 26375330]
37. Ratwani RM, Hodgkins M, Bates DW. Improving electronic health record usability and safety requires transparency. JAMA 2018 Nov 29:1. [doi: 10.1001/jama.2018.14079] [Medline: 30489619]
38. Peischl B, Ferk M, Holzinger A. The fine art of user-centered software development. Softw Qual J 2014 May 27;23(3):509-536. [doi: 10.1007/s11219-014-9239-1]
39. International Organization for Standardization (ISO). 2015 Jan. IEC 60601-1-11:2015. Medical electrical equipment -- Part 1-11: General requirements for basic safety and essential performance -- Collateral standard: Requirements for medical electrical equipment and medical electrical systems used in the home healthcare environment URL: https://www.iso.org/standard/65529.html [accessed 2019-02-26] [WebCite Cache ID 76TEr4I8Z]

40. International Organization for Standardization (ISO). 2018 Sep. ISO/IEEE 11073-20702:2018. Health informatics -- Point-of-care medical device communication -- Part 20702: Medical devices communication profile for web services URL: https://www.iso.org/standard/74703.html [accessed 2019-02-26] [WebCite Cache ID 76TFWoDks]
41. Duhm J, Fleischmann R, Schmidt S, Hupperts H, Brandt SA. Mobile electronic medical records promote workflow: Physicians' perspective from a survey. JMIR Mhealth Uhealth 2016 Jun 06;4(2):e70 [FREE Full text] [doi: 10.2196/mhealth.5464] [Medline: 27268720]
42. Mobasheri MH, King D, Judge S, Arshad F, Larsen M, Safarfashandi Z, et al. Communication aid requirements of intensive care unit patients with transient speech loss. Augment Altern Commun 2016 Dec;32(4):261-271. [doi: 10.1080/07434618.2016.1235610] [Medline: 27868434]
43. Dolcourt J. CNET. 2019 Mar 20. Galaxy Fold vs Mate X: Battle of the foldable phones URL: https://www.cnet.com/news/galaxy-fold-vs-huawei-mate-x-battle-of-the-foldable-phones-comparison-mwc/ [accessed 2019-02-26] [WebCite Cache ID 76TEaP34T]
44. Armaignac DL, Saxena A, Rubens M, Valle CA, Williams LS, Veledar E, et al. Impact of telemedicine on mortality, length of stay, and cost among patients in progressive care units: Experience from a large healthcare system. Crit Care Med 2018 May;46(5):728-735 [FREE Full text] [doi: 10.1097/CCM.0000000000002994] [Medline: 29384782]
45. Hawkins HA, Lilly CM, Kaster DA, Groves RH, Khurana H. ICU telemedicine comanagement methods and length of stay. Chest 2016 Aug;150(2):314-319. [doi: 10.1016/j.chest.2016.03.030] [Medline: 27048869]
46. Fritz BA, Chen Y, Murray-Torres TM, Gregory S, Ben Abdallah A, Kronzer A, et al. Using machine learning techniques to develop forecasting algorithms for postoperative complications: Protocol for a retrospective study. BMJ Open 2018 Dec 10;8(4):e020124 [FREE Full text] [doi: 10.1136/bmjopen-2017-020124] [Medline: 29643160]
47. Lee CK, Hofer I, Gabel E, Baldi P, Cannesson M. Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality. Anesthesiology 2018 Oct;129(4):649-662. [doi: 10.1097/ALN.0000000000002186] [Medline: 29664888]
48. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med 2018 Apr;46(4):547-553 [FREE Full text] [doi: 10.1097/CCM.0000000000002936] [Medline: 29286945]
49. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
50. Eerikäinen LM, Vanschoren J, Rooijakkers MJ, Vullings R, Aarts RM. Reduction of false arrhythmia alarms using signal selection and machine learning. Physiol Meas 2016 Dec;37(8):1204-1216. [doi: 10.1088/0967-3334/37/8/1204] [Medline: 27454128]
51. Charbonnier S, Gentil S. A trend-based alarm system to improve patient monitoring in intensive care units. Control Eng Pract 2007 Sep;15(9):1039-1050. [doi: 10.1016/j.conengprac.2006.12.005]
52. Goedken CC, Moeckli J, Cram PM, Reisinger HS. Introduction of tele-ICU in rural hospitals: Changing organisational culture to harness benefits. Intensive Crit Care Nurs 2017 Jun;40:51-56. [doi: 10.1016/j.iccn.2016.10.001] [Medline: 28216177]
53. Larinkari S, Liisanantti JH, Ala-Lääkkölä T, Meriläinen M, Kyngäs H, Ala-Kokko T. Identification of tele-ICU system requirements using a content validity assessment. Int J Med Inform 2016 Feb;86:30-36. [doi: 10.1016/j.ijmedinf.2015.11.012] [Medline: 26725692]
54. Van Velsen L, Wildevuur S, Flierman I, Van Schooten B, Tabak M, Hermens H. Trust in telemedicine portals for rehabilitation care: An exploratory focus group study with patients and healthcare professionals. BMC Med Inform Decis Mak 2016 Jan 27;16:11 [FREE Full text] [doi: 10.1186/s12911-016-0250-2] [Medline: 26818611]
55. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: Lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
56. The Topol Review: Preparing the Healthcare Workforce to Deliver the Digital Future. Leeds, UK: Health Education England; 2019 Feb. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2019-02-26] [WebCite Cache ID 76THfTwlX]
57. Cresswell KM, Lee L, Mozaffar H, Williams R, Sheikh A, NIHR ePrescribing Programme Team. Sustained user engagement in health information technology: The long road from implementation to system optimization of computerized physician order entry and clinical decision support systems for prescribing in hospitals in England. Health Serv Res 2017 Dec;52(5):1928-1957 [FREE Full text] [doi: 10.1111/1475-6773.12581] [Medline: 27714800]
58. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
59. Zajicek H, Meyers A. Digital health entrepreneurship. In: Rivas H, Wac K, editors. Digital Health: Scaling Healthcare to the World. Cham, Switzerland: Springer International Publishing; 2018:271-287.
60. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14. [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]

Abbreviations

BioCog: Biomarker Development for Postoperative Cognitive Impairment in the Elderly
CDSS: clinical decision support system
DFG: Deutsche Forschungsgemeinschaft
DLR: Deutsches Zentrum für Luft- und Raumfahrt eV
ECG: electrocardiogram
$etCO_2$: end-tidal carbon dioxide
ICT: information and communication technology
ICU: intensive care unit
IEC: International Electrotechnical Commission
IEEE: Institute of Electrical and Electronics Engineers
PDMS: patient data management system
SpO_2: peripheral capillary oxygen saturation

Edited by G Eysenbach; submitted 23.12.18; peer-reviewed by C Auschra, D Tscholl, M van Mourik, M Görges, G Bhutkar; comments to author 07.02.19; revised version received 05.03.19; accepted 30.03.19; published 30.04.19.

Please cite as:
Poncette AS, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, Balzer F
Clinical Requirements of Future Patient Monitoring in the Intensive Care Unit: Qualitative Study
JMIR Med Inform 2019;7(2):e13064
URL: http://medinform.jmir.org/2019/2/e13064/
doi:10.2196/13064
PMID:31038467

Original Paper

Facility and Regional Factors Associated With the New Adoption of Electronic Medical Records in Japan: Nationwide Longitudinal Observational Study

Hideaki Kawaguchi[1], MD, PhD; Soichi Koike[2], MD, PhD; Kazuhiko Ohe[1], MD, PhD

[1]Department of Biomedical Informatics, The University of Tokyo, Tokyo, Japan

[2]Division of Health Policy and Management, Center for Community Medicine, Jichi Medical University, Tochigi, Japan

Corresponding Author:
Hideaki Kawaguchi, MD, PhD
Department of Biomedical Informatics
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, 113-0033
Japan
Phone: 81 358006427
Email: hikawaguchi-tky@umin.ac.jp

Abstract

Background: The rate of adoption of electronic medical record (EMR) systems has increased internationally, and new EMR adoption is currently a major topic in Japan. However, no study has performed a detailed analysis of longitudinal data to evaluate the changes in the EMR adoption status over time.

Objective: This study aimed to evaluate the changes in the EMR adoption status over time in hospitals and clinics in Japan and to examine the facility and regional factors associated with these changes.

Methods: Secondary longitudinal data were created by matching data in fiscal year (FY) 2011 and FY 2014 using reference numbers. EMR adoption status was defined as "EMR adoption," "specified adoption schedule," or "no adoption schedule." Data were obtained for hospitals (n=4410) and clinics (n=67,329) that had no adoption schedule in FY 2011 and for hospitals (n=1068) and clinics (n=3132) with a specified adoption schedule in FY 2011. The EMR adoption statuses of medical institutions in FY 2014 were also examined. A multinomial logistic model was used to investigate the associations between EMR adoption status in FY 2014 and facility and regional factors in FY 2011. Considering the regional variations of these models, multilevel analyses with second levels were conducted. These models were constructed separately for hospitals and clinics, resulting in four multinomial logistic models. The odds ratio (OR) and 95% Bayesian credible interval (CI) were estimated for each variable.

Results: A total of 6.9% of hospitals and 14.82% of clinics with no EMR adoption schedules in FY 2011 had adopted EMR by FY 2014, while 10.49% of hospitals and 33.65% of clinics with specified adoption schedules in FY 2011 had cancelled the scheduled adoption by FY 2014. For hospitals with no adoption schedules in FY 2011, EMR adoption/scheduled adoption was associated with practice size characteristics, such as number of outpatients (from quantile 4 to quantile 1: OR 1.67, 95% CI 1.005-2.84 and OR 2.40, 95% CI 1.80-3.21, respectively), and number of doctors (from quantile 4 to quantile 1: OR 4.20, 95% CI 2.39-7.31 and OR 2.02, 95% CI 1.52-2.64, respectively). For clinics with specified EMR adoption schedules in FY 2011, the factors negatively associated with EMR adoption/cancellation of scheduled EMR adoption were the presence of beds (quantile 4 to quantile 1: OR 0.57, 95% CI 0.45-0.72 and OR 0.74, 95% CI 0.58-0.96, respectively) and having a private establisher (quantile 4 to quantile 1: OR 0.27, 95% CI 0.13-0.55 and OR 0.43, 95% CI 0.19-0.91, respectively). No regional factors were significantly associated with the EMR adoption status of hospitals with no EMR adoption schedules; population density was positively associated with EMR adoption in clinics with no EMR adoption schedule (quantile 4 to quantile 1: OR 1.49, 95% CI 1.32-1.69).

Conclusions: Different approaches are needed to promote new adoption of EMR systems in hospitals as compared to clinics. It is important to induce decision making in small- and medium-sized hospitals, and regional postdecision technical support is important to avoid cancellation of scheduled EMR adoption in clinics.

(JMIR Med Inform 2019;7(2):e14026) doi:10.2196/14026

KEYWORDS

electronic health records; health services research; health policy; Bayes theorem

Introduction

With the increasing focus on electronic health, the secondary utilization of electronic medical records (EMRs) is becoming important as a tool for collecting patient clinical information. Data mining methods for generating new knowledge from large datasets and machine learning methods (such as deep learning) are developing rapidly, and the use of medical information has accordingly attracted more attention [1].

Although the rate of adoption of the EMR system has increased internationally, the adoption rate in Japan is lower than that in other countries [2-5]. Furthermore, the Survey of Medical Institutions conducted by the Ministry of Health, Labour and Welfare in fiscal year (FY) 2014 revealed that 45.5% of hospitals and 60.8% of clinics are not planning to adopt EMRs in the future [6], indicating that the lack of new adoption of EMRs is a major issue in Japan.

Resistance to new adoption of EMRs is not specific to Japan. For example, while financial incentives such as the Meaningful Use program have contributed to the growing EMR adoption rate in the United States [7], the growth rate is slowing down [8] and the adoption rates in underserved and small hospitals are lower than those in other hospitals [9,10]. To promote further EMR adoption, it is necessary to analyze the characteristics of medical facilities that have newly adopted the EMR system.

Although previous studies have revealed the characteristics of medical institutions that have adopted EMR systems [11-15], to the best of our knowledge, no study has used time series data to analyze the factors directly related to new adoption of EMRs. To predict whether medical institutions will adopt the EMR system in the near future, it is necessary to analyze longitudinal data rather than cross-sectional data. Furthermore, the EMR adoption process consists of several stages, and therefore, decision making for EMR adoption should be considered separately from management after the decision to adopt EMRs has been made [16]. Therefore, research should evaluate the changes over time in the EMR adoption status and determine the factors associated with these changes. In addition to facility factors, geographical or regional factors affecting the new adoption of EMR should be considered. Previous studies have reported regional variation in EMR adoption and associations between EMR adoption and health care professional shortage area and metropolitan statuses [17,18]. Furthermore, the regional factors associated with EMR adoption differ between hospitals and clinics in Japan [19]. Thus, we hypothesized that the changes in EMR adoption status over time and the factors associated with these changes would differ between hospitals and clinics. To test this hypothesis, this study aimed to evaluate the changes in the EMR adoption status of hospitals and clinics in Japan over time and to evaluate the facility and regional factors associated with these changes.

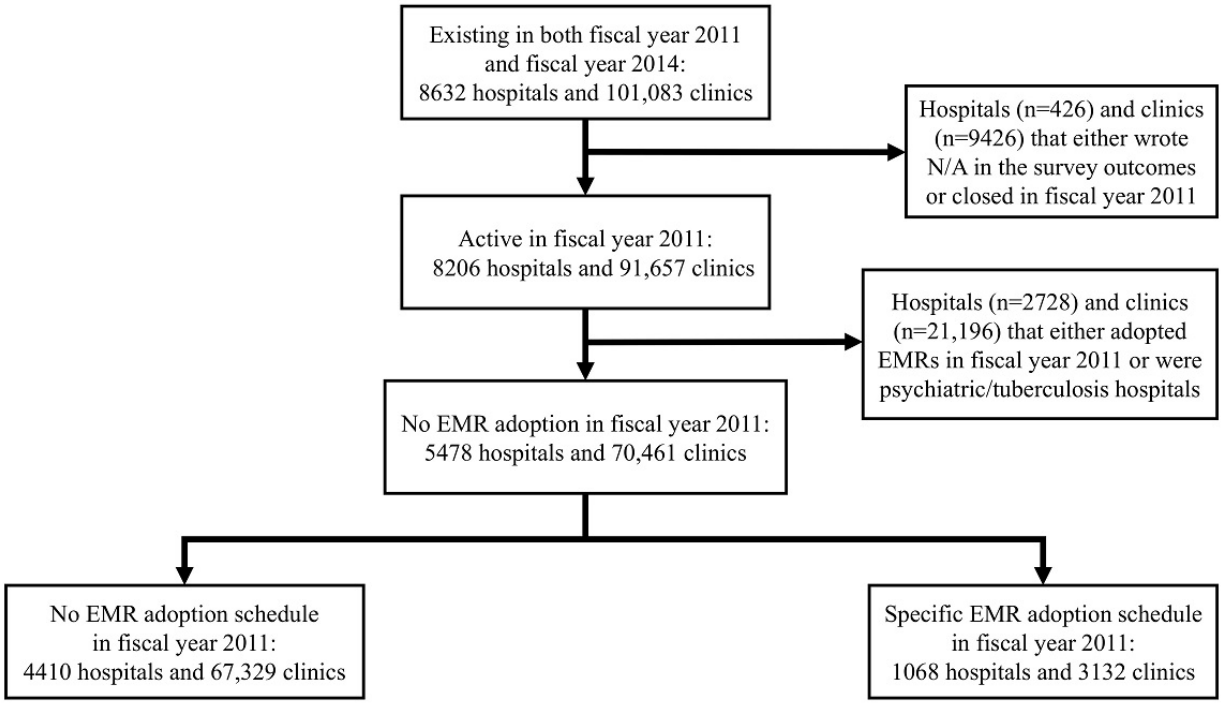
Methods

Study Design

This study was a nationwide longitudinal observational study that secondarily analyzed existing survey data. Data for FY 2011 and FY 2014 were matched using reference numbers, creating longitudinal data from the whole of Japan.

As in a previous study [16], the two steps involved in EMR adoption were assumed to be (1) having no adoption schedule to deciding to adopt EMRs and (2) deciding to adopt an EMR system for successful EMR adoption. The EMR adoption status in FY 2014 was examined in medical institutions that had no adoption schedule in FY 2011 and in medical institutions that had scheduled EMR adoption in FY 2011. This investigation included all the hospitals and clinics in Japan that had not adopted EMRs in FY 2011. A “clinic” in Japan was defined as a medical institution with fewer than 20 beds. First, these hospitals and clinics were divided into those that had not scheduled EMR adoption in FY 2011 and those that had scheduled EMR adoption in FY 2011. Second, the medical institutions were assessed in accordance with the inclusion/exclusion criteria in Figure 1. This study finally analyzed 4410 hospitals and 67,329 clinics that had not scheduled EMR adoption in FY 2011 and 1068 hospitals and 3132 clinics that had scheduled EMR adoption in FY 2011.

Figure 1. Inclusion/exclusion criteria of hospitals and clinics. EMR: electronic medical record; N/A: not applicable.



Data Sources

Data on EMR adoption and other characteristics of medical facilities were obtained from the Survey of Medical Institutions, which is a detailed triennial survey of all medical institutions conducted by the Japan Ministry of Health, Labour and Welfare [6]. Permission was obtained from the Japan Ministry of Health, Labour and Welfare to analyze the survey data of individual medical institutions. The EMR adoption status of each medical facility was defined as a response of 1/2, 3, or 4 to the survey item "Electronic medical record system adoption status," where 1=adopted in the entire hospital/clinic, 2=adopted in part of the hospital/clinic, 3=specified adoption schedule, and 4=no adoption schedule.

Geographical or regional information, such as municipality boundary data, was obtained from the Municipality Map Maker for Web [20]. Japan comprises 47 prefectures, and the Japanese Government established subprefectural medical regions called secondary medical service areas (SMSAs) [19]. An SMSA is defined as a medical unit that evaluates the demand and supply of health resources. To determine the SMSA data, ArcGIS version 10.2.1 (ESRI Japan Inc, Tokyo, Japan) was used to combine municipality-level parameters, as each SMSA consists of several municipalities. The assessed regional factors included available socioeconomic and macro health-environment factors identified in previous studies [17-19]; these data were collected from e-Stat, the national Japanese government database [21].

Statistical Analysis

Multinomial Logistic Regression Analysis

A multinomial logistic model was used to investigate the associations between EMR adoption status in FY 2014 and facility and regional factors in FY 2011. Multinomial logistic regression analysis is a statistical model that deals with more than three categorical variables. Three EMR adoption statuses (ie, EMR adoption, specified adoption schedule, and no adoption schedule) were set as outcome variables. It is also necessary to determine a reference in advance regarding outcome variables in the multinomial logistic regression model. "No adoption schedule" was set in advance as a reference regarding outcome variables in the model targeting medical institutions that had not scheduled EMR adoption in FY 2011 (model 1), and "specified adoption schedule" was set as a reference in the model targeting medical institutions that had scheduled EMR adoption in FY 2011 (model 2). These models were constructed separately for hospitals and clinics, resulting in four multinomial logistic models.

Multilevel Analysis

To take regional variations of these models into account, multilevel analyses with second levels were conducted. Random variations in intercepts at the SMSA level were set as the second level. Four multilevel multinomial logistic regression models were constructed.

Explanatory Variables

The facility variables used in this study comprised facility factors identified in previous studies [11-15], which were collected from the Survey of Medical Institutions in FY 2011. These factors were advocating internal medicine, advocating surgery, emergency medical institution design, number of outpatients, number of doctors, presence of interns, implementation of home medical care, classification of the establisher, and number of beds. The presence of interns was not included as an explanatory variable in the models targeting clinics, as clinics rarely have interns in Japan. As the numbers of outpatients were extremely skewed, medical institutions were categorized in accordance with the number of outpatients from quantile 1 (lowest) to quantile 4 (highest). Regarding the

classification of the establisher, "national," "public medical institution," and "social insurance affiliated organization" were defined as public, while "medical corporation," "private," and "others" were defined as private. Hospitals were categorized in accordance with the number of doctors from quantile 1 (lowest) to quantile 4 (highest), while clinics were categorized into those with more than one doctor or those with less than one doctor. In accordance with the format of the Survey of Medical Institutions, the number of doctors in hospitals refers to the number of working doctors, while the number of doctors in clinics refers to those employed on a full-time basis. Hospitals were categorized into those with less than 200 beds, 200-399 beds, and ≥400 beds, while clinics were categorized into those with or without beds.

The following regional factors were analyzed: population density (people per km^2), average per capita income (million JPY), number of working doctors per 1000 people (separately for hospitals and clinics), and proportion of interns to all working doctors. As the population density distribution was extremely skewed, medical institutions were categorized in accordance with the population density from quantile 1 (lowest density) to quantile 4 (highest density). As the two surveys were conducted in different years, the data for population density and average per capita income were obtained for FY 2010. Data with missing values in explanatory variables were deleted.

Parameter Estimations

Markov chain Monte Carlo simulations with 1000 iterations and a burn-in period of 500 iterations were used to estimate the parameters of the multilevel multinomial logistic models. R-hat diagnostic was used to check Markov chain Monte Carlo convergence, with 1.1 set as the cut-off value [22]. The odds ratio (OR) and 95% Bayesian credible intervals (CI) were calculated for each variable, and an association was considered nonsignificant if the 95% CI of the OR included 1.

The multicollinearity of covariates was evaluated using the generalized variance inflation factor (GVIF) [23]. Although the average per capita income was an important factor, this factor was removed because it had the greatest GVIF of >2.5. Furthermore, the number of doctors was removed from the model targeting hospitals with a specified EMR adoption schedule because this factor had a high GVIF of >2.5. All other variables had a GVIF of <2.5 and were entered into the multilevel multinomial logistic model.

The widely applicable information criterion (WAIC) was used as a measure of the goodness-of-fit of the Bayesian statistical model; the model with the lowest WAIC was considered the best-fit model [24]. The WAIC was used to compare the multilevel multinomial logistic regression model with a normal multinomial logistic regression model without consideration of the SMSA-level effects.

All analyses were conducted using R V.3.4.1 (R Core Team, Vienna, Austria) [25].

Results

Time Series Changes in the Electronic Medical Records Adoption Status of Hospitals and Clinics

Table 1 shows the status of EMR adoption in FY 2014. Only 6.9% of the hospitals with no EMR adoption schedule in FY 2011 had newly adopted EMRs by FY 2014, while 14.82% of clinics with no EMR adoption schedule in FY 2011 had newly adopted EMRs by FY 2014. However, 10.49% of hospitals with a specified adoption schedule in FY 2011 had cancelled the scheduled adoption by FY 2014, while 33.65% of clinics with a specified adoption schedule in FY 2011 had cancelled the scheduled adoption by FY 2014.

Multilevel Multinomial Logistic Regression Targeting Hospitals

Table 2 shows the associations between the EMR adoption status and each explanatory variable for hospitals. After removing the hospitals with missing data, the model included 4278 hospitals with no adoption schedule and 1051 hospitals with a specified adoption schedule in FY 2011.

For hospitals with no adoption schedule in FY 2011, the factors associated with EMR adoption and specified adoption schedules were the number of doctors, number of outpatients, and presence of interns. The number of outpatients was more strongly associated with EMR adoption, while the number of doctors and presence of interns were more strongly associated with specified adoption schedules.

For hospitals with specified adoption schedules in FY 2011, the number of outpatients, number of beds, presence of interns, and population density were associated with EMR adoption, while advocating surgery was associated with the cancellation of scheduled EMR adoption.

The WAICs of the multilevel models with consideration of regional effects targeting hospitals with no adoption schedule in FY 2011 and hospitals with specified adoption schedules in FY 2011 were 6538.6 and 1859.7, respectively; those of the regression models without consideration of regional effects were 6548.4 and 1859.9, respectively. This indicates that the multilevel models did not produce a much better fit than the normal regression models without consideration of the regional effects.

Table 1. Electronic medical record adoption status in fiscal year 2014.

Facility	No adoption schedule in fiscal year 2011			Specified adoption schedule in fiscal year 2011		
	Adoption, n (%)	Specified adoption schedule, n (%)	No adoption schedule, n (%)	Adoption, n (%)	Specified adoption schedule, n (%)	No adoption schedule, n (%)
Hospitals	303 (6.87)	1212 (27.48)	2895 (65.65)	563 (52.72)	393 (36.80)	112 (10.49)
Clinics	9981 (14.82)	3045 (4.52)	54303 (80.65)	1360 (43.42)	718 (22.92)	1054 (33.65)

Table 2. Results of multilevel multinomial logistic regression targeting hospitals. Significant variables are presented as italics.

Target	Hospitals with no adoption schedule in fiscal year 2011 (n=4278)		Hospitals with a specified adoption schedule in fiscal year 2011 (n=1051)	
	Adoption, OR[a] (95% CI[b])	Specified adoption schedule, OR (95% CI)	Adoption, OR (95% CI)	No adoption schedule, OR (95% CI)
Intercept	0.06 (0.03-0.14)	0.19 (0.13-0.31)	0.96 (0.42-2.19)	0.57 (0.16-1.86)
Advocating internal medicine	0.76 (0.46-1.29)	0.85 (0.63-1.12)	0.94 (0.51-1.78)	1.41 (0.60-3.83)
Advocating surgery	1.11 (0.76-1.58)	0.97 (0.81-1.17)	1.03 (0.68-1.58)	*0.51 (0.28-0.90)*
Designed as an emergency hospital	1.38 (0.999-1.94)	1.20 (0.99-1.45)	1.06 (0.74-1.56)	1.36 (0.77-2.47)
Number of outpatients				
Quantile 1	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Quantile 2	0.77 (0.47-1.23)	1.18 (0.94-1.49)	*1.83 (1.21-2.83)*	0.98 (0.53-1.72)
Quantile 3	1.04 (0.63-1.66)	*1.68 (1.31-2.14)*	*2.84 (1.82-4.40)*	*0.43 (0.20-0.88)*
Quantile 4	*1.67 (1.005-2.84)*	*2.40 (1.80-3.21)*	*2.79 (1.60-5.14)*	0.58 (0.21-1.44)
Number of doctors[c]				
Quantile 1	1.00 (reference)	1.00 (reference)	N/A[d]	N/A
Quantile 2	1.58 (0.99-2.55)	1.13 (0.90-1.40)	N/A	N/A
Quantile 3	*1.86 (1.13-3.08)*	*1.54 (1.23-1.95)*	N/A	N/A
Quantile 4	*4.20 (2.39-7.31)*	*2.02 (1.52-2.64)*	N/A	N/A
Presence of interns	*2.08 (1.34-3.16)*	*1.45 (1.07-1.95)*	*1.87 (1.28-2.86)*	0.88 (0.38-2.02)
Implementation of home medical care	0.96 (0.72-1.25)	*1.18 (1.01-1.39)*	0.93 (0.69-1.26)	0.70 (0.44-1.13)
Private establisher	0.73 (0.51-1.08)	0.85 (0.67-1.07)	0.73 (0.50-1.08)	0.74 (0.37-1.50)
Number of beds				
<200	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
200-399	1.19 (0.82-1.71)	1.12 (0.88-1.42)	1.18 (0.80-1.76)	0.94 (0.48-1.75)
≥400	1.38 (0.72-2.58)	1.34 (0.83-2.17)	*2.10 (1.12-3.86)*	0.81 (0.24-2.37)
Population density per km^2				
Quantile 1	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Quantile 2	0.84 (0.56-1.23)	1.13 (0.91-1.41)	0.78 (0.49-1.25)	0.71 (0.35-1.56)
Quantile 3	1.02 (0.68-1.57)	1.25 (0.98-1.58)	0.75 (0.47-1.21)	0.79 (0.37-1.80)
Quantile 4	0.80 (0.50-1.27)	0.97 (0.76-1.25)	*0.44 (0.25-0.78)*	1.09 (0.49-2.54)
Working doctors per 1000 population	1.08 (0.87-1.35)	1.07 (0.94-1.21)	1.05 (0.87-1.28)	0.84 (0.56-1.18)
Proportion of interns to all working doctors	0.94 (0.88-1.002)	0.98 (0.95-1.02)	0.99 (0.92-1.07)	1.03 (0.90-1.16)

[a]OR: odds ratio.

[b]CI: credible interval.

[c]The factor “number of doctors” was removed from the model targeting hospitals with a specified electronic medical record adoption schedule, as it had a high generalized variance inflation factor of >2.5

[d]N/A: not applicable.

Multilevel Multinomial Logistic Regression Targeting Clinics

Table 3 shows the associations between EMR adoption status and each explanatory variable for clinics. After removing the clinics with missing data, the model included 55,815 clinics with no adoption schedule and 3030 clinics with a specified adoption schedule in FY 2011.

Table 3. Results of multilevel multinomial logistic regression targeting clinics. Significant variables are presented as italics.

Target	Clinics with no adoption schedule in fiscal year 2011 (n=55,815)		Clinics with a specified adoption schedule in fiscal year 2011 (n=3030)	
	Adoption, OR[a] (95% CI[b])	Specified adoption schedule, OR (95% CI)	Adoption, OR (95% CI)	No adoption schedule, OR (95% CI)
Intercept	0.07 (0.06-0.08)	0.02 (0.01-0.02)	7.05 (3.34-15.19)	6.87 (3.21-15.46)
Advocating internal medicine	*1.17 (1.11-1.23)*	*1.26 (1.15-1.38)*	0.99 (0.79-1.22)	0.84 (0.66-1.05)
Advocating surgery	1.06 (0.99-1.14)	1.10 (0.998-1.22)	1.07 (0.84-1.35)	0.98 (0.76-1.25)
Designed as an emergency clinic	0.85 (0.55-1.29)	1.14 (0.71-1.81)	0.72 (0.26-2.01)	1.23 (0.46-3.34)
Number of outpatients				
Quantile 1	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Quantile 2	*1.26 (1.17-1.35)*	*1.67 (1.47-1.90)*	0.94 (0.72-1.25)	*0.68 (0.52-0.91)*
Quantile 3	*1.33 (1.24-1.44)*	*2.05 (1.81-2.33)*	0.88 (0.67-1.16)	*0.51 (0.39-0.68)*
Quantile 4	*1.52 (1.41-1.64)*	*2.61 (2.29-2.95)*	0.86 (0.65-1.14)	*0.47 (0.34-0.63)*
More than one doctor	*1.30 (1.24-1.37)*	*1.94 (1.79-2.10)*	1.21 (0.98-1.48)	0.88 (0.71-1.08)
Implementation of home medical care	*1.15 (1.09-1.21)*	*1.48 (1.35-1.62)*	0.93 (0.75-1.15)	0.87 (0.70-1.08)
Private establisher	*1.18 (1.04-1.35)*	0.90 (0.75-1.11)	*0.27 (0.13-0.55)*	*0.43 (0.19-0.91)*
Presence of beds	*0.89 (0.82-0.97)*	*1.20 (1.08-1.33)*	*0.57 (0.45-0.72)*	*0.74 (0.58-0.96)*
Population density per km^2				
Quantile 1	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Quantile 2	*1.14 (1.04-1.26)*	1.07 (0.94-1.19)	0.92 (0.71-1.21)	0.84 (0.63-1.11)
Quantile 3	*1.35 (1.22-1.50)*	1.05 (0.92-1.20)	1.07 (0.79-1.45)	1.18 (0.89-1.60)
Quantile 4	*1.49 (1.32-1.69)*	1.14 (0.99-1.30)	1.11 (0.79-1.54)	1.26 (0.92-1.76)
Working doctors per 1000 population	0.98 (0.93-1.03)	1.03 (0.98-1.08)	0.92 (0.83-1.03)	0.93 (0.84-1.02)
Proportion of interns to all working doctors	1.01 (0.995-1.03)	1.00 (0.98-1.02)	1.04 (0.99-1.09)	1.02 (0.97-1.07)

[a]OR: odds ratio.

[b]CI: credible interval.

For clinics with no adoption schedule in FY 2011, a wider range of factors (such as private establisher and population density) were associated with EMR adoption than with specified adoption schedules.

For clinics with specified adoption schedules in FY 2011, the presence of beds and having a private establisher were associated with both EMR adoption and the cancellation of scheduled EMR adoption. In contrast, the number of outpatients was negatively associated with the cancellation of scheduled EMR adoption.

The WAICs of the multilevel models with consideration of regional effects targeting clinics with no adoption schedule in FY 2011 and clinics with specified adoption schedules in FY 2011 were 65477.2 and 6411.3, respectively; those of the regression models without consideration of regional effects were 65615.6 and 6416.3, respectively. This indicates that the multilevel model targeting clinics with no adoption schedule in FY 2011 had a slightly better fit than the normal regression model without consideration of regional effects.

Discussion

Time Series Changes in the Electronic Medical Records Adoption Status in Hospitals Versus Clinics

Time series data were used to precisely analyze the changes over time in the EMR adoption status and the factors affecting new EMR adoption in hospitals and clinics in Japan. To the best of our knowledge, this is the first study to detail the changes in the EMR adoption status over time.

Fewer hospitals with no EMR adoption schedule in FY 2011 had adopted EMR within 3 years compared with clinics with no EMR adoption schedules in FY 2011. However, more hospitals with no EMR adoption schedule in FY 2011 had planned to adopt EMR within 3 years compared with clinics without an EMR adoption schedule in FY 2011. This shows that more hospitals than clinics planned to adopt EMR, but that clinics took less time to adopt EMR than hospitals. As 37,876 or more clinics were staffed by a maximum of one physician, such clinics would be able to implement EMR more quickly than hospitals, and this difference in implementation speed may have influenced the increased incidence of new EMR adoption

in clinics compared with hospitals. Furthermore, only few clinics planned to but did not adopt EMR within 3 years (4.52%). Although such clinics can make decisions relatively easily, it is likely that their EMR adoption capabilities are lacking; thus, follow-up on the implementation of EMR is necessary.

About half of the hospitals with specified EMR adoption schedules in FY 2011 had actually adopted EMR within 3 years, while about 10% had cancelled the scheduled EMR adoption within 3 years. In contrast, a greater proportion of clinics with specified EMR adoption schedules in FY 2011 had cancelled the scheduled EMR adoption within 3 years (33.65%). The reason why more clinics than hospitals cancelled the scheduled EMR adoption might be that clinics are also quicker to make decisions to cancel scheduled adoption than hospitals.

In summary, once the decision to adopt an EMR system has been made, hospitals tend to follow through with scheduled EMR adoption more often than clinics; therefore, the decision-making process itself seems to be important for hospitals. However, compared with hospitals, more clinics decided to adopt EMR and then cancelled the scheduled EMR adoption; therefore, adequate EMR adoption support after decision making seems to be important for clinics.

Facility Factors Associated With the Electronic Medical Records Adoption Status of Hospitals Versus Clinics

For hospitals with no EMR adoption scheduled in FY 2011, the facility factors associated with actual EMR adoption were also associated with the scheduling of EMR adoption. In particular, the EMR system was adopted more often by hospitals with large numbers of medical staff, which is consistent with previous studies [11,26,27]. In addition, multilevel multinomial logistic regression targeting hospitals with specified EMR adoption schedules in FY 2011 revealed that medium-sized hospitals rarely cancelled scheduled EMR adoption. Therefore, it is important to encourage small- or medium-sized hospitals to adopt the EMR system. For example, the implementation of financial incentives such as the Meaningful Use program might effectively increase the decision to adopt the EMR system, as in the United States [7,28]. In Japan, financial incentives for EMR adoption have been offered to large hospitals [5], but these also need to be offered to small- and medium-sized hospitals.

Regarding clinics with no EMR adoption schedule in FY 2011, large clinics with a large number of outpatients and more than one doctor, similar to hospitals, were more likely to adopt and plan to adopt the EMR system compared with small clinics. Furthermore, the implementation of home medical care was a significant factor influencing EMR adoption, and the use of EMRs might be expected to prompt sharing of medical information, as in the United States [8,29]. Of the clinics with no EMR adoption schedule in FY 2011, those that had beds were more likely to plan to adopt EMR, while those without beds were more likely to actually adopt EMR. In addition, multilevel multinomial logistic regression targeting clinics with a specified EMR adoption schedule in FY 2011 revealed that the clinics that had beds tended not to adopt the EMR system and not to cancel the EMR adoption schedule. In other words, clinics that had beds tended to postpone the scheduled EMR adoption, despite being more likely to adopt the EMR system. Therefore, postdecision support might be particularly useful for clinics with beds.

Regional Factors Associated With the Electronic Medical Records Adoption Status of Hospitals Versus Clinics

Regional factors were not associated with the EMR adoption status of hospitals with no EMR adoption schedule in FY 2011. In addition, the WAIC of the model that considered the SMSA-level effects was close to that of the model that did not consider SMSA-level effects, indicating that regionality did not have a large influence on the EMR adoption status of hospitals. This is consistent with our previous study [19]. For hospitals, the new adoption of EMR was mainly influenced by facility factors rather than regional factors.

Regarding clinics, population density was positively associated with EMR adoption in clinics with no EMR adoption schedule in FY 2011, and the WAIC of the model that considered the SMSA-level effects was less than that of the model that did not consider SMSA-level effects, indicating that regionality influences EMR adoption in clinics. These results are also consistent with our previous research [19]. An example of EMR adoption support on a regional basis is the Regional Extension Centers program in the United States [18,30], which provides technical support for EMR implementation, mainly in rural areas. Development of the Regional Extension Centers program might lead to the expansion of regional health care networks in Japan.

Trends in Electronic Medical Records Adoption After 2015

Although this study used the most recent available data (from FY 2014), the status of EMR usage is progressing rapidly. Therefore, the trends regarding EMR adoption after 2015 are described here and compared with the results of this study.

In the United States, the EMR adoption rate has rapidly grown, and the Health Information Technology for Economic and Clinical Health act has received a certain appreciation [7]. Similarly, EMR adoption has advanced via the distribution of financial incentives in other countries. For example, financial incentives are considered important for the adoption of national EMR in France [31] and Canada [32]. This suggests that our findings are consistent with the EMR adoption trends in other countries after 2015.

Although it was not possible to include the data of individual medical institutions from the Survey of Medical Institutions conducted in FY 2017 in Japan, in this study, we compared the aggregated public data from FY 2017 with the aggregated data from FY 2011 and FY 2014 [6] (Multimedia Appendix 1). Although the changes from FY 2011 to 2014 and from FY 2014 to 2017 regarding clinics showed similar trends, the trends regarding hospitals differed between time periods. Specifically, the decrease in the proportion of hospitals with no EMR adoption schedule in FY 2014-2017 was smaller than that in FY 2011-2014, and the proportion of hospitals with no EMR adoption schedule in FY 2017 was smaller than that in FY 2014.

Although a certain number of hospitals with specific EMR adoption schedules in FY 2014 could have implemented EMR adoption by FY 2017, only a small number of hospitals with no EMR adoption schedules in FY 2014 could have adopted EMRs by FY 2017. Therefore, the results of this study regarding hospitals with no EMR adoption schedule in FY 2011 might not be directly applicable to the period from FY 2014 to 2017.

Limitations

This study has several limitations. First, our study used secondary data, and several data were unavailable. For example, it was not possible to consider important factors such as the characteristics and attitude toward EMR of working physicians, and the profits of the medical institutions. Second, as this study used secondary data sources, the time period was set as 3 years, and the change in EMR adoption status within 3 years was not evaluated. However, using this 3-year period to evaluate the change in EMR adoption status might be too short for hospitals and too long for clinics. In the future, cohort data should be prepared to analyze shorter time periods. Third, although data were acquired from all medical institutions in Japan and generalizability was secured, the number of samples accordingly increased and the regression coefficients tended to be significant; therefore, it might be difficult to interpret the regression coefficients. In particular, there were 10 times more clinics than hospitals. Our results require validation via comparison with other survey data. Fourth, our study used data from up to FY 2014, and therefore, these results cannot explain the EMR adoption situation after FY 2015. Although the recent trends in EMR adoption were described and compared with the results, it is necessary to conduct ongoing research using data from FY 2017.

Despite these limitations, to our knowledge, this is the first study to use longitudinal and spatial data to perform a detailed analysis of the facility and regional factors related to new adoption of EMR in Japan. Although many other countries do not have data available on the EMR adoption status in each hospital [2], this study has the advantage of using time series data obtained from almost all medical institutions in Japan. Our findings will help effectively promote new adoption of the EMR system.

Conclusions

As the characteristics of time series changes in EMR adoption differ between hospitals and clinics, different approaches are important for the promotion of new adoption of EMRs in hospitals versus clinics in Japan. For hospitals, it is important to induce decision making; for clinics, in addition to inducing decision making, it is important to provide postdecision technical support. In addition, facility factors affecting EMR adoption should mainly be considered for hospitals, while both regional and facility factors should be considered for clinics.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science (grant number JP16K00432). A draft of this manuscript was edited by Convention Linkage, Inc, Tokyo, Japan.

Conflicts of Interest

HK is employed by Mitsubishi Tanabe Pharma as a principal research scientist.

Multimedia Appendix 1

Aggregated data regarding the electronic medical record adoption status in fiscal years 2011, 2014, and 2017. Values are given as n (%).

[PDF File (Adobe PDF File), 143KB - medinform_v7i2e14026_app1.pdf]

References

1. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
2. Jha AK, Doolan D, Grandt D, Scott T, Bates DW. The use of health information technology in seven nations. Int J Med Inform 2008 Dec;77(12):848-854. [doi: 10.1016/j.ijmedinf.2008.06.007] [Medline: 18657471]
3. Schoen C, Osborn R, Squires D, Doty M, Rasmussen P, Pierson R, et al. A survey of primary care doctors in ten countries shows progress in use of health information technology, less in other areas. Health Aff (Millwood) 2012 Dec;31(12):2805-2816 [FREE Full text] [doi: 10.1377/hlthaff.2012.0884] [Medline: 23154997]
4. Protti D. Comparison of information technology in general practice in 10 countries. Healthc Q 2007;10(2):107-116. [Medline: 17491575]
5. Yoshida Y, Imai T, Ohe K. The trends in EMR and CPOE adoption in Japan under the national strategy. Int J Med Inform 2013 Oct;82(10):1004-1011. [doi: 10.1016/j.ijmedinf.2013.07.004] [Medline: 23932755]
6. Ministry of Health, Labour and Welfare: Survey of medical institutions (e-Stat) (in Japanese). URL: https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00450021&tstat=000001030908 [accessed 2019-06-10] [WebCite Cache ID 76JknWpFg]
7. Adler-Milstein J, Jha AK. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]

8. 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]
9. Kharrazi H, Gonzalez CP, Lowe KB, Huerta TR, Ford EW. Forecasting the Maturation of Electronic Health Record Functions Among US Hospitals: Retrospective Analysis and Predictive Model. J Med Internet Res 2018 Aug 07;20(8):e10458 [FREE Full text] [doi: 10.2196/10458] [Medline: 30087090]
10. 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. [doi: 10.1377/hlthaff.2013.0308] [Medline: 23840052]
11. Kruse CS, Kothman K, Anerobi K, Abanaka L. Adoption Factors of the Electronic Health Record: A Systematic Review. JMIR Med Inform 2016;4(2):e19 [FREE Full text] [doi: 10.2196/medinform.5525] [Medline: 27251559]
12. Cheung CS, Tong EL, Cheung NT, Chan WM, Wang HH, Kwan MW, et al. Factors associated with adoption of the electronic health record system among primary care physicians. JMIR Med Inform 2013;1(1):e1 [FREE Full text] [doi: 10.2196/medinform.2766] [Medline: 25599989]
13. Yasunaga H, Imamura T, Yamaki S, Endo H. Computerizing medical records in Japan. Int J Med Inform 2008 Oct;77(10):708-713. [doi: 10.1016/j.ijmedinf.2008.03.005] [Medline: 18448385]
14. Archer N, Cocosila M. A comparison of physician pre-adoption and adoption views on electronic health records in Canadian medical practices. J Med Internet Res 2011 Aug 12;13(3):e57 [FREE Full text] [doi: 10.2196/jmir.1726] [Medline: 21840835]
15. Abdekhoda M, Ahmadi M, Gohari M, Noruzi A. The effects of organizational contextual factors on physicians' attitude toward adoption of Electronic Medical Records. J Biomed Inform 2015 Feb;53:174-179 [FREE Full text] [doi: 10.1016/j.jbi.2014.10.008] [Medline: 25445481]
16. Merhi MI. A process model leading to successful implementation of electronic health record systems. IJEH 2015;8(2/3/4):185. [doi: 10.1504/IJEH.2015.075355]
17. King J, Furukawa MF, Buntin MB. Geographic variation in ambulatory electronic health record adoption: implications for underserved communities. Health Serv Res 2013 Dec;48(6 Pt 1):2037-2059 [FREE Full text] [doi: 10.1111/1475-6773.12078] [Medline: 23800087]
18. Samuel CA. Area-level factors associated with electronic health record adoption and meaningful use in the Regional Extension Center Program. J Am Med Inform Assoc 2014;21(6):976-983 [FREE Full text] [doi: 10.1136/amiajnl-2013-002347] [Medline: 24798687]
19. Kawaguchi H, Koike S, Ohe K. Regional differences in electronic medical record adoption in Japan: A nationwide longitudinal ecological study. Int J Med Inform 2018 Dec;115:114-119. [doi: 10.1016/j.ijmedinf.2018.05.002] [Medline: 29779713]
20. Kirimura T, Nakaya T, Yano K. Building a Spatio-temporal GIS Database about Boundaries of Municipalities (in Japanese). Theory and Applications of GIS 2011;19(2):139-148. [doi: 10.5638/thagis.19.139]
21. General government statistics portal (e-Stat) (in Japanese). URL: https://www.e-stat.go.jp/ [accessed 2019-06-10] [WebCite Cache ID 76JkyMoLP]
22. Gelman A, Carlin J, Stern H, Rubin D. Bayesian Data Analysis. London, UK: Chapman & Hall; 2013.
23. Allison PD. Multiple Regression: A Primer (Research Methods and Statistics). Newbury Park, CA: Pine Forge Press; 1999.
24. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 2010;11:3571-3594 [FREE Full text]
25. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. URL: http://www.r-project.org/ [accessed 2019-06-10] [WebCite Cache ID 76Jl9ImxT]
26. Simon SR, Kaushal R, Cleary PD, Jenter CA, Volk LA, Poon EG, et al. Correlates of electronic health record adoption in office practices: a statewide survey. J Am Med Inform Assoc 2007;14(1):110-117 [FREE Full text] [doi: 10.1197/jamia.M2187] [Medline: 17068351]
27. DesRoches CM, Campbell EG, Rao SR, Donelan K, Ferris TG, Jha A, et al. Electronic health records in ambulatory care--a national survey of physicians. N Engl J Med 2008 Jul 3;359(1):50-60. [doi: 10.1056/NEJMsa0802005] [Medline: 18565855]
28. Blumenthal D. Stimulating the adoption of health information technology. N Engl J Med 2009 Apr 9;360(15):1477-1479. [doi: 10.1056/NEJMp0901592] [Medline: 19321856]
29. Gold M, McLaughlin C. Assessing HITECH Implementation and Lessons: 5 Years Later. Milbank Q 2016 Sep;94(3):654-687 [FREE Full text] [doi: 10.1111/1468-0009.12214] [Medline: 27620687]
30. Samuel CA, King J, Adetosoye F, Samy L, Furukawa MF. Engaging providers in underserved areas to adopt electronic health records. Am J Manag Care 2013;19(3):229-234 [FREE Full text] [Medline: 23544764]
31. Burnel P. The introduction of electronic medical records in France: More progress during the second attempt. Health Policy 2018 Sep;122(9):937-940 [FREE Full text] [doi: 10.1016/j.healthpol.2018.07.022] [Medline: 30100529]
32. Progress in electronic medical record adoption in Canada. Can Fam Physician 2015 Dec;61(12):1076-1084 [FREE Full text] [Medline: 27035020]

Abbreviations

CI: Bayesian credible interval

EMR: electronic medical record
FY: fiscal year
GVIF: generalized variance inflation factor
OR: odds ratio
SMSA: secondary medical service area
WAIC: widely applicable information criterion

Edited by G Eysenbach; submitted 15.03.19; peer-reviewed by G Strudwick, S Housbane; comments to author 03.04.19; revised version received 19.05.19; accepted 03.06.19; published 14.06.19.

Please cite as:
Kawaguchi H, Koike S, Ohe K
Facility and Regional Factors Associated With the New Adoption of Electronic Medical Records in Japan: Nationwide Longitudinal Observational Study
JMIR Med Inform 2019;7(2):e14026
URL: http://medinform.jmir.org/2019/2/e14026/
doi:10.2196/14026
PMID:31199307

Original Paper

Adapting State-of-the-Art Deep Language Models to Clinical Information Extraction Systems: Potentials, Challenges, and Solutions

Liyuan Zhou[1,2], BSc, MSc; Hanna Suominen[1,2,3,4], MSc, PhD; Tom Gedeon[1], BSc (Hons), PhD

[1]Research School of Computer Science, College of Engineering and Computer Science, The Australian National University, Canberra, Australia

[2]Machine Learning Group, Data61, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia

[3]Faculty of Science and Technology, University of Canberra, Canberra, Australia

[4]Department of Future Technologies, Faculty of Science and Engineering, University of Turku, Turku, Finland

Corresponding Author:
Liyuan Zhou, BSc, MSc
Research School of Computer Science
College of Engineering and Computer Science
The Australian National University
108 North Road
Canberra, 2600
Australia
Phone: 61 (02) 6125 5111
Email: annjouno@gmail.com

Abstract

Background: Deep learning (DL) has been widely used to solve problems with success in speech recognition, visual object recognition, and object detection for drug discovery and genomics. Natural language processing has achieved noticeable progress in artificial intelligence. This gives an opportunity to improve on the accuracy and human-computer interaction of clinical informatics. However, due to difference of vocabularies and context between a clinical environment and generic English, transplanting language models directly from up-to-date methods to real-world health care settings is not always satisfactory. Moreover, the legal restriction on using privacy-sensitive patient records hinders the progress in applying machine learning (ML) to clinical language processing.

Objective: The aim of this study was to investigate 2 ways to adapt state-of-the-art language models to extracting patient information from free-form clinical narratives to populate a handover form at a nursing shift change automatically for proofing and revising by hand: first, by using domain-specific word representations and second, by using transfer learning models to adapt knowledge from general to clinical English. We have described the practical problem, composed it as an ML task known as information extraction, proposed methods for solving the task, and evaluated their performance.

Methods: First, word representations trained from different domains served as the input of a DL system for information extraction. Second, the transfer learning model was applied as a way to adapt the knowledge learned from general text sources to the task domain. The goal was to gain improvements in the extraction performance, especially for the classes that were topically related but did not have a sufficient amount of model solutions available for ML directly from the target domain. A total of 3 independent datasets were generated for this task, and they were used as the training (101 patient reports), validation (100 patient reports), and test (100 patient reports) sets in our experiments.

Results: Our system is now the state-of-the-art in this task. Domain-specific word representations improved the macroaveraged F1 by 3.4%. Transferring the knowledge from general English corpora to the task-specific domain contributed a further 7.1% improvement. The best performance in populating the handover form with 37 headings was the macroaveraged F1 of 41.6% and F1 of 81.1% for filtering out irrelevant information. Performance differences between this system and its baseline were statistically significant (P<.001; Wilcoxon test).

Conclusions: To our knowledge, our study is the first attempt to transfer models from general deep models to specific tasks in health care and gain a significant improvement. As transfer learning shows its advantage over other methods, especially on classes with a limited amount of training data, less experts' time is needed to annotate data for ML, which may enable good results even in resource-poor domains.

(JMIR Med Inform 2019;7(2):e11499) doi:10.2196/11499

KEYWORDS

computer systems; artificial intelligence; deep learning; information storage and retrieval; medical informatics; nursing records; patient handoff

Introduction

Background

Machine learning (ML) is being studied and used in a variety of health informatics applications (eg, disease progression prediction, therapy planning, medical diagnostic reasoning, and automatic patient management) as a way to help clinical experts to improve the efficiency and quality of medical care [1,2].

A clear majority of these applications use supervised learning, which infers knowledge from labeled training data. However, because of stringent restrictions on the use of clinical data [3], data collections on real health care scenarios that are open for research and development are very limited [4]. Moreover, the few available sources have limitations such as research-only use [5], nondisclosure of data [6], or limited commercial licenses [7].

Zheng et al [8] proposed an information extraction (IE) framework called IDEAL-X, which uses online learning techniques to update the learning model based on user feedback. Although the performance of their system looks very impressive, the types of text this system is able to extract are limited to 5, and these types such as age, gender, and medicine can be easily retrieved with rule-based systems rather than ML systems. Leroy introduced a rule-based automated IE system that extracts diagnostic criteria from electronic health records for autism spectrum disorders [9]. As the rules are manually generated based on human observations of 1 specific data set, their system cannot be generalized to other tasks.

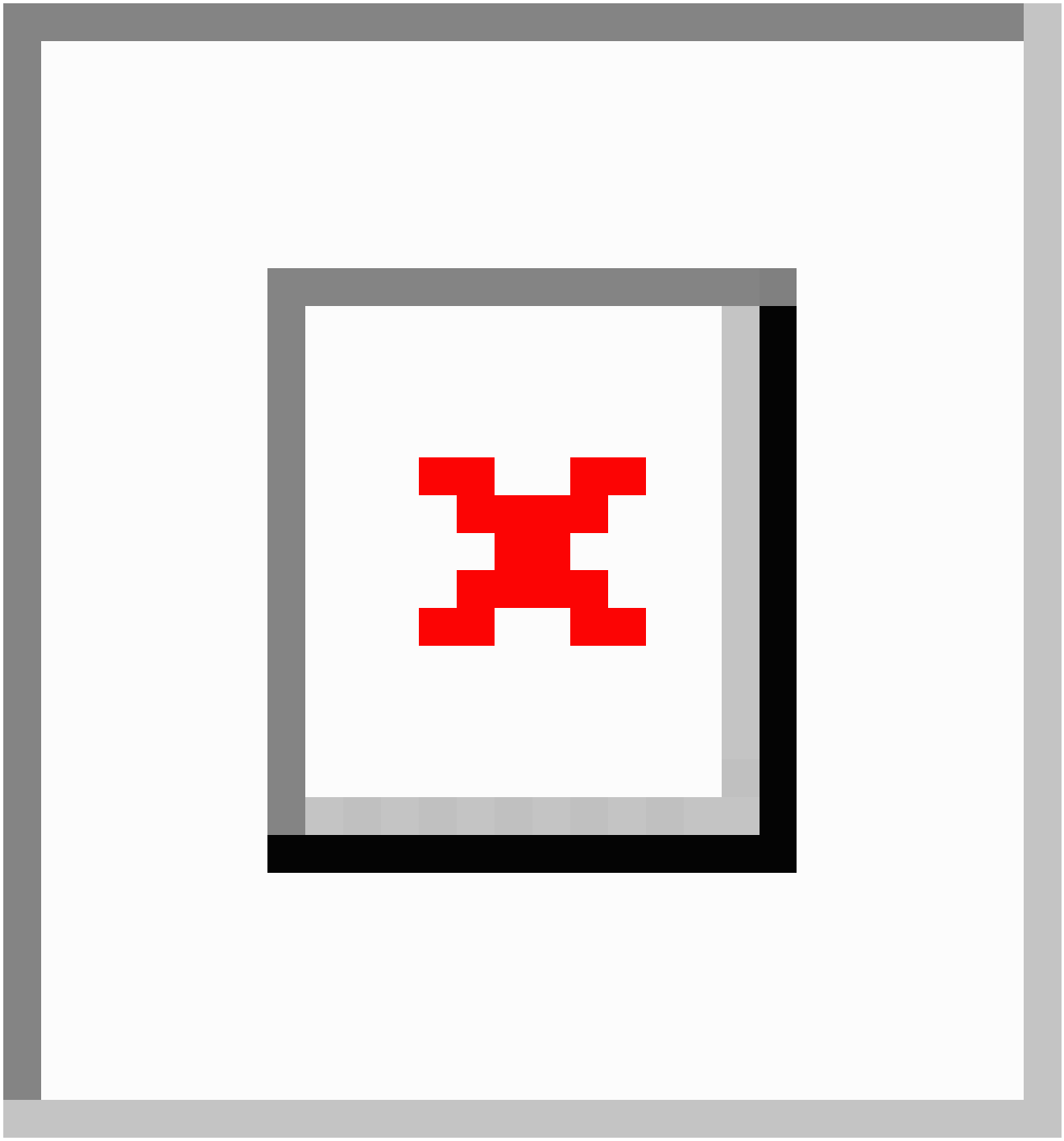
In our previous study [4], we have already (1) discussed the importance of comprehensive record keeping along with information flow in health care in general and clinical handover in particular, (2) developed and freely released a set of 101 synthetic clinical handover cases with verbatim conversations and associated audio recordings constructed by a nurse with over 12 years of experience in clinical practice to make sure the cases are closely matched with the typical data found in a nursing shift change, and (3) introduced and evaluated a cascaded system that uses speech recognition (SR) to recognize verbal clinical handover information and IE to fill in a handover form for clinical proofing and sign-off (Figure 1).

Objectives

In this study, we have released another 2 datasets that follow exactly the same format as our first release to supplement the original dataset called the National Information and Communications Technology Australia (NICTA) Synthetic Nursing Handover Data. These 3 independent datasets target researchers who are training, validating, and testing ML-based SR and IE methods for the handover record-keeping task. A description of our dataset is available in Multimedia Appendix 1.

More importantly, in this study, we have improved our IE performance by using an ML method, which learns from other data collections and transfers this knowledge to the handover task. Processing correctness is crucial in medical informatics applications; our benchmark results show that this task is very challenging [4], and the previous state-of-the-art result on this task was only 38.2% on macro F1 [10]. Even with our supplementary data, the size of the in-domain training set is still not adequate to train a traditional multilayer neural network (NN) model for our IE task composed as a 50-class classification.

Figure 1. A processing pipeline that transforms verbal clinical handover information into electronic structured records automatically.



Generating or getting access to a large manually labeled training corpus for this task is not easy. Fortunately, distributed word representations, which can be learned from unlabeled data, have recently been shown to have high utility in many natural language processing (NLP) applications [11-14]. In this study, we have investigated whether pretrained word embeddings generated from general Web text could improve our system performance in the IE task, even though it is relatively domain-specific. Furthermore, we are also interested to discover if training supplementary word embeddings based on a domain-related corpus is more helpful than from a general English corpus.

Transferring the knowledge learned from another domain to our task is another way to cope with the problem of lack of training instances. This method has shown its effectiveness in previous studies [15,16]. In this study, we have implemented a transfer learning–based approach to adapt weights of features and labels from different source data corpora to gain an improvement in the clinical handover IE task. More specifically, if we define our current task as the target domain, then the dataset that we want to adapt weights from is the source domain, so we first train sequence labeling models on a source domain training corpus and then learn the correlations between source labels and the labels in our task. After this, we use the model parameters of each related class in the source model to initialize our conditional random field (CRF) model in our clinical handover IE task. To extend the study, we have also explored whether models learned from a source corpus, which is close

to a clinical domain, are more helpful than models trained on a generic, large labeled corpus.

To summarize the contributions of this study, we have released the data to study SR and IE and introduced a state-of-the-art IE method for the handover task. The method is based on transfer learning and compares with both the most recent deep learning (DL) approaches and more traditional CRFs for sequence labeling.

Methods

NICTA Synthetic Nursing Handover Data

To fulfil the purpose of constructing systems to automatically generate structuring of the narrative documents from nursing shift change speech and handover, NICTA Synthetic Nursing Handover Data [17-19] was created at NICTA/Data61 from 2012 to 2016. Their main author was a professional Registered Nurse (RN), Maricel Angel, who has over 12 years of experience in clinical nursing, based on general practice in medical wards. Therefore, the text is very similar to real documents in typical clinical scenarios.

This data collection of 301 records in total contained 3 disjoint subsets for training (101 records), validation (100 records), and testing (100 records; Figure 2). All 3 subsets were created under a consistent practice with the same standards as used by Suominen et al [4]. Each record contains a patient profile; a written, free-form clinical handover for this profile; a voice speech record of the handover; and, finally, a written, structured document. To represent the most common chronic diseases and national health priority areas in Australia [20], 4 kinds of patients (ie, cardiovascular, neurological, renal, and respiratory patients) were introduced into each subset and independently followed a uniform distribution to provide a balanced demographic sample. The structured document includes annotation of 5 classes (PATIENT INTRODUCTION, MY SHIFT, APPOINTMENTS, MEDICATION, and FUTURE CARE), which were further divided into 37 subclasses, supplemented by the category of not applicable (NA) for irrelevant information.

Figure 2. Descriptive statistics of text snippets highlighted in the training, validation, and test set.

CATEGORY	No. of instances			CATEGORY	No. of instances		
	Training	Validation	Test		Training	Validation	Test
A. PATIENT INTRODUCTION	2064	2224	1611	**C. APPOINTMENTS**	393	373	499
Given Names/Initials	119	104	100	Status	159	91	79
Last Name	99	100	101	Description	157	222	313
Age in Years	246	281	281	Clinician Given names / initials	1	0	1
Gender	489	378	178	Clinician Last name	2	4	1
Current Room	54	54	100	Day	40	26	34
Current Bed	180	156	198	Time	28	19	39
Under Dr_GivenNames/Initials	15	60	53	city	1	0	0
UnderDr_Lastname	181	108	128	ward	2	0	0
Admission Reason/Diagnosis	414	544	153	Clinician Title	0	10	26
Allergy	14	3	38	Hospital	0	1	1
Chronic Condition	70	11	73	Appointment/Procedure_Ward	0	0	5
Disease/Problem History	147	266	206	**D. MEDICATION**	262	329	493
Care Plan	36	156	2	Medicine	157	177	296
B. MY SHIFT	1353	1153	1033	Dosage	37	117	52
Status	483	293	285	Status	68	35	145
Contraption	44	88	43	**E. FURURE CARE**	644	499	252
Input/Diet	101	79	107	Alert/Warning/Abnormal Result	59	24	28
Output/Diuresis/Bowel Movement	52	64	50	Goal/Task To Be Completed/Expect	496	386	103
Wounds/Skin	55	24	25	Discharge/Transfer Plan	89	89	121
Activities of Daily Living	245	182	295	**F. N.A.**	3771	3152	2652
Risk Management	12	94	40	**Total**	8487	7730	6540
Other Observation	361	329	188				

Word Representation

Word embedding is a vector matrix learned from an unlabeled text corpus that maps vocabulary to a dense vector space. It attempts to model the distributional hypothesis that words that occur in similar contexts tend to be semantically similar. It has been shown to contribute to a variety of NLP tasks even without using any other features [21].

To capture word vector representations from large amounts of unlabeled text, we adapted a skip-gram model [22] that uses each current word to predict words in the neighboring context. The training objective of the skip-gram model is to maximize the averaged log probability over all training cases (T) of appearance of the context word w_{t+j} given the current word w_t, where j is the offset of the context word from the current word in a context window size of c:

$$1/T\sum\_\{T\}\^\{t=1\}(\sum\_\{-c\leq j\leq c, j\neq 0\}(log\ p(w\_\{t+j\}|w\_t)))$$

Then, applies softmax to each context word w_O of a given occurrence of word w_I:

$$P(w\_O|w\_I)=exp(v'\^\{T\}\_\{w\_O\}v\_\{w\_I\})/\sum\^W\_\{w=1\} exp(v'\^T\_w\ v\_\{w\_I\})$$

...where v_w is the input and v'_w is the output word embedding of a word w, and W is the size of the training vocabulary.

Out of the 2 variations to optimize computational efficiency of the skip-gram model, we have used negative sampling rather than hierarchical softmax because, for sequence tagging tasks in NLP, it can maintain more semantic information during the training process [23] and obtain better results [24]. Rather than calculating *exp(v'^T_w v_{w_I})* for all w in the vocabulary when *calculating log p(w_{t+j}|w_t)*, negative sampling replaces

it with a logistic regression and distinguishes a context word *w_O* from noise (negative samples):

log σ(v'^T_{w_O} v_{w_I}) + ∑^k_{i=1} E_{w_i~U(w)} [log σ(-v'^T_{w_i} v_{w_I})],

...where *k* is the amount of negative samples for each data sample, and *U(w)* is a unigram distribution of words.

To integrate context information from a new domain-specific corpus *D_1* (eg, our clinical handover dataset), we need to update our existing word embeddings that were learned from a general large text collection *D_0*. However, this comes with a significant challenge in that the original word embeddings were trained on a very large corpus whereas our target domain training data are normally much smaller in size. More specifically, if we compare the vocabulary between *D_0* and *D_1*, *V_0* ∩ *V_1* is the intersection between 2 vocabularies, which is mainly composed of more general terms compared with *V_1/V_0*: the relative complement of *V_0* with regard to *V_1*. As word vectors in *V_0* have already been trained for several epochs and converged to our desired values, we do not want to change them significantly. However, new words that were just introduced to the vocabulary from *V_1/V_0* contain domain-specific, related terms about which we care the most. Owing to the limitations in the available amount of training samples, a large learning rate at the beginning is useful to adjust these vectors to the regions they belong. Using the original skip-gram algorithm will potentially either adjust the already converged vectors from *V_0* away from their optimized values or the new word vectors will not get close to words that are similar to them in the vector space. To cope with this problem, we used 2 strategies in our experiments: averaged initialization as well as different learning rates.

In averaged initialization, our assumption was that words in similar contexts would have similar meanings or have grammatical similarities. Although this is not always true, this strategy helps in learning the new words, as they start from an (averaged) optimized point rather than from scratch. To model this, whenever a new word appears, the initial value of the vector is set to be the averaged value in each dimension of words in the same sentence:

1/S∑^{S}_{i=1} **v**_i, where **v**_i= **v**, if **v** V_0,

= **v**_0, otherwise

...where *S* is the sentence length and **v**_*0* denotes the vector in *V_0*, which represents the vector of *unknown* words.

During the training procedure, we have also used different learning rates for words in *V_1/V_0* in contrast with *V_0∩V_1*:

α_t=α_{0_new}(1–t/n), t≠0, and **v** V_1/V_0

α_t=0.2, t=0, and **v** V_1/V_0

α_t=α_{0_old}(1–t/n), t≠0, and **v** V_1∩V_0

α_t=α_{0_new}(|V_1/V_0|/|V_0∪V_1|), t≠0, and **v** V_1/V_0

...where *n* is the amount of samples input to the network. For new words, the initial learning rate α*_{0_new}* is set to 0.2 and decreases over time. For words that are already in *V_0*, the initial learning rate is set to the portion of new words in the entire vocabulary: the more learning samples we have for new words, the larger the initial learning rate for old words.

Transfer Learning for Sequence Labeling

For sequence tagging tasks that use supervised ML, the amount and purity of training data is crucial to the performance of our system. As the complexity of learning a 40-class classifier is high, for some labels, there is only 1 case in the training set, which could not be generalized to infer good functions [25]. Therefore, introducing more training data could improve the final results. However, when more training instances are not available, or are extremely costly in human labor, transfer learning, which adapts weight matrixes from functions trained with another dataset and applied to the current task, is another way to gain knowledge of labels with limited training instances.

The underlying idea of transfer learning is simple: in deep NNs, there are several hidden layers between the input and output layers; as data feed forward from the input layer to the output layer, the composition of features is learned from earlier layers [26]. A typical sequence tagging structure can be demonstrated as the left block in Figure 3: neurons in lower layers tend to capture some common, nondomain, or task-specific concepts, and later layers would concatenate these features and generate higher-level concepts. Therefore, weights learned from other datasets or even other tasks can be potentially reused as long as the structure of the later layers are consistent between the source model and target model.

Several strategies of transfer learning on different NLP tasks and domains have been explored. A simple strategy is to copy all weight matrixes in the source model to the target model and fine-tune the target model with new data [27], which successfully outperformed the leading team in the Multilingual Emoji Prediction task [28] by 1.55% without any feature engineering procedure. However, this method requires the source and target model to have the exact same structure. An alternative strategy is to map annotations from different datasets into 3 consistent labels and use source domain model parameters directly as initializations for a target domain model in named entity recognition [16]. Finally, human adjustment of rules and features [29,30] or clustering labels from source domain and target domains to automatically generate label mappings from one dataset to the other [31] can be applied as transfer learning strategies. However, their productivity may be limited when adapting a general source model to multiple tasks and also the generated mappings might not be satisfactory, especially when 2 datasets have dramatic differences in terms of phraseology and grammar.

The method we have introduced in our study was able to transfer knowledge to a target domain that does not match labels in the source domain, does not depend on human integration during the label mapping process, and is able to map labels from very different datasets. This method follows 3 steps (Figure 3): the first step trains a CRF model on the source domain, the second step uses the weight matrix of the source model *W^s* to train a 2-layer CRF that predicts a target domain label given a source domain label, and, finally, the third step is to initialize the parameters of the target domain model using the product of *W_s* and the second layer weight matrix *W^t* obtained in step 2.

First, in the source model training step, a linear-chain CRF model is trained on a large labeled source dataset. For each word *x_i* in a sequence **x**, *y_i* Y is the label of *x_i*, where Y is the label set. Then (**x**; **y**) is a sequence of word label pairs; a linear chain CRF is a distribution *p* (**y**| **x**) that takes the form:

$$p(\mathbf{y}|\mathbf{x}) = 1/Z \prod^{L}_{l=1} (\mathbf{W}^{f} f(y_l, \mathbf{x}) + \mathbf{W}^{g} g(y_{l-1}, y_l))$$

...where *Z* is a normalization constant, *L* is the length of **x**, *f(y_l*, **x***)* is a real value feature function of **x**, *g(y_{l-1}, y_l)* is a feature function of current label *y_l* and previous label *y_{l-1}* in the sequence to capture the cooccurrence between adjunct labels, and contains the parameters of the feature functions.

For now, the source CRF model can be seen as an NN with 2 hidden layers: the lower layer, which is connected by **W**^*f*, and the upper layer connected by **W**^*g* (Figure 3). However, because the information that is captured by **W**^*g* is normally domain- and task-specific, label correlations do not always have similarities between different domains, which would thus not be suitable to be transferred to the other task. In contrast, the lower layer learned correlations between words and source labels, which is what we are interested in, are actually a logistic regression model:

$$\sigma(y^*, \mathbf{x}\_i, \mathbf{W}^{f}) = \exp(\mathbf{W}^{f}\_{.y^*} f(y^*\_i, \mathbf{x}\_i)) / \sum\_{y\ Y} \exp(\mathbf{W}^{f}\_{.y} f(y, \mathbf{x}\_i))$$

Second, in the source label and target label correlations step, we have propagated the label probabilities from the later layer of the source domain model to another logistic regression classifier to form a 2-layer linear-chain CRF model that predicts target domain labels and uses source domain labels to learn correlations between the source and target labels. More specifically, the linear layer from the source model can be defined as:

$$\mathbf{a}\_i = \mathbf{W}^{s} f(y, \mathbf{x}\_i)$$

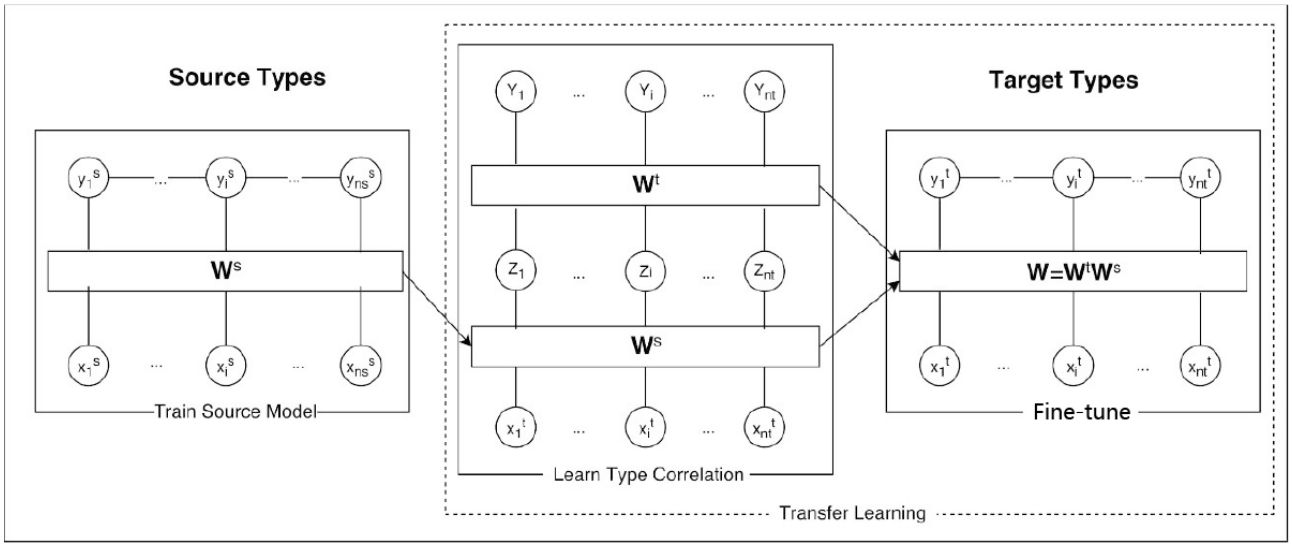
...where each **a**_*i* is the probability for each source label and **W**^*s* denotes the weight matrix from source domain. After this layer, a linear regression classifier takes the output from **a**_*i* to predict target labels:

$$p(y'|\mathbf{a}) = \sigma(y', \mathbf{a}\_i; \mathbf{W}^{t}),$$

...where *y'* is the target type. This is equal to:

$$p(y'|\mathbf{x}) = \sigma(y', \mathbf{x}\_i; \mathbf{W}^{t}\mathbf{W}^{s})$$

Figure 3. Transfer learning model structure.



After **W**^*s* and **W**^*t* are trained, we were able to initialize a CRF model to predict target labels with **W**^*f* = **W**^*t***W**^*s* as the third step:

$$p(\mathbf{y}|\mathbf{x}) = 1/Z \prod^{L}_{l=1} (\mathbf{W}^{t}\mathbf{W}^{s} f(y_l, \mathbf{x}))$$

During this procedure, the parameters of label NA, **W**^*t_{NA}* are reset to be zeros because the amount of instances of NA in the text corpus is much larger than that of other labels, which would cause the model to be biased to the dominant class.

Theoretically, the parameters of feature functions will converge to the same weights with a random initialization model when the number of iterations is large enough because the loss function is convex. However, our aim was to inherit knowledge from the source domain, updating **W**^*t***W**^*s* too often would cause the model to forget what it has learned so far. Therefore, early stopping and adaptive gradient algorithm (AdaGrad) [32] are applied to preserve the learned source domain knowledge.

Performance Evaluation and Experimental Design

To compare the performance of systems, we have measured the precision, recall, and F1 (harmonic mean of precision and recall) over all categories [33]. More specifically, microaveraged F1 and macroaveraged F1 are calculated. As our purpose was to emphasize on systems that perform well in all classes rather than only in the classes that have majority instances, macroaveraged F1 was selected as the main evaluation measurement.

The resources used in our experiments were derived from 7 different corpora (Table 1). Among them, 3 were general English text based, 3 were specific to the English health care domain, and 1 was the test set. Details of the test set are available in Multimedia Appendix 2.

1. General English Corpora:

- *English Wikipedia* is a freely available corpus from the September 2014 version of all pages from all Wikipedia wikis. It contains more than 3 million English pages, 100 million sentences, and 3.4 billion words in total after cleaning.
- *University of Maryland, Baltimore County (UMBC) WebBase corpus* is a dataset containing a collection of 100 million English Web pages from more than 50,000 websites with over 3 billion words processed from the February 2007 crawl by the Stanford WebBase project [34].
- *One Billion Words Benchmark for Language Modeling* is a freely available standard corpus of 4.2 GB (0.8 billion words) for building and testing language models [35].

2. Medical Domain–specific English Corpora:
   - *I2B2* is a collection of fully deidentified clinical records provided by the I2B2 National Centre for Biomedical Computing funded by U54LM008748 and was originally prepared for *Shared Tasks for Challenges in NLP for Clinical Data* organized by Uzuner, I2B2, and SUNY [36-39].
   - *PubMed* is a free resource containing over 27 million citations to the biomedical literature and publication abstracts derived from MEDLINE, life science journals, and Web books. It was developed and is maintained by the National Centre for Biotechnology Information at the US National Library of Medicine (NLM).
   - *PubMed Central (PMC) Open Access Subset* contains over 1 million biomedical articles from PMC, which is a free archive of biomedical and life sciences journal publications at the US National Institutes of Health's NLM.
   - *NICTA TRAIN* is the NICTA Synthetic nursing handover dataset, an open clinical dataset of 3 sets of nursing handover records, very similar to real documents in Australian English. Each record consists of a patient profile, spoken free-form text document, written free-form text document, and written structured document [40].

Table 1. Word embeddings training corpora.

Corpus	Size	Source
English Wikipedia	3.4 billion words	Wikimedia downloads [41]
UMBC[a]	>3 billion words	UMBC WebBase corpus [42]
One Billion	0.8 billion words	One Billion Word Benchmark for Measuring Progress in Statistical Language Modeling [43]
I2B2	18,082 unique words	I2B2 NLP[b] research data sets [6]
PubMed	27 million records	PubMed resources [44]
PubMed Central	1 million articles	PubMed resources [45]
National Information and Communications Technology Australia Train	101 records	Hospital handover forms [17]

[a]UMBC: University of Maryland, Baltimore County.

[b]NLP: natural language processing.

Table 2. Mapping of entity types between the source and target corpora.

In-domain source: Bolt, Beranek and Newman	Out-domain source: I2B2	Target: NICTA[a] Clinical Handover
PERSON	PATIENT	PATIENT_INTRODUCTION/Given_names
PERSON	PATIENT	PATIENT_INTRODUCTION/Last_name
PERSON	DOCTOR	PATIENT_INTRODUCTION/Under_Dr:Given_names
PERSON	DOCTOR	PATIENT_INTRODUCTION/Under_Dr:Last_name
PERSON	DOCTOR	APPOINTMENTS/Clinician:Last_name
PERSON	DOCTOR	APPOINTMENTS/Clinician: Given_names
ORGANIZATION:HOSPITAL	HOSPITAL	APPOINTMENTS/Hospital
DATE:AGE	__[b]	PATIENT_INTRODUCTION/Age_in_years
PER_DESC	—	PATIENT_INTRODUCTION/Gender
PER_DESC	—	APPOINTMENTS/Clinician Title
GPE:CITY	LOCATION	APPOINTMENTS/City
DATE:DATE	DATE	APPOINTMENTS/Day
TIME	—	APPOINTMENTS/Time
CARDINAL	ID	PATIENT_INTRODUCTION/Current_room
CARDINAL	ID	PATIENT_INTRODUCTION/Current_bed
PRODUCT:OTHER	—	Medication/Medicine
SUBSTANCE:DRUG	—	Medication/Medicine
QUANTITY:3D (volume)	—	Medication/Dosage
QUANTITY:OTHER	—	Medication/Dosage
QUANTITY:TEMPERATURE	—	My_shift/Status
QUANTITY:WEIGHT	—	My_shift/Status
SUBSTANCE:FOOD	—	My_shift/Input_diet
FACILITY	—	APPOINTMENTS/Ward
DISEASE	—	PATIENT_INTRODUCTION/Admission_reason/diagnosis
DISEASE	—	PATIENT_INTRODUCTION/Chronic_condition
DISEASE	—	PATIENT_INTRODUCTION/Disease/problem_history

[a]NICTA: National Information and Communications Technology Australia.

[b]Does not contain any matching label from the source domain.

For the source domain corpus, we used 1 domain-related dataset, which contained labels that were relevant but not exactly the same as in our target domain data, and 1 of the domain corpora, which contained many of the general labels, including some labels that were relevant to biometrics. With this setup, we were keen to find out whether the parameters learned from the same domain were more valuable than those from general English.

The related domain source corpus was the aforementioned I2B2. It included fully deidentified discharge summaries and progress notes from real hospital scenarios. All records had been manually annotated for concept, assertion, and relation information. The corpus contained entities of 7 different labels: PATIENT, DOCTOR, HOSPITAL, DATE, PHONE, LOCATION, and ID. They were potentially relevant to labels in our NICTA dataset.

The general domain source corpus was Bolt, Beranek and Newman (BBN), which has a 1 million-word Penn Treebank corpus of Wall Street Journal texts annotated by BBN with 28 main types of entities: 12 named entity types (Person, Facility, Organization, geographical entities (GPE), Location, Nationality, Product, Event, Work of Art, Law, Language, and Contact-Info), 9 nominal entity types (Person, Facility, Organization, GPE, Product, Plant, Animal, Substance, and Disease and Game), and 7 numeric types (Date, Time, Percent, Money, Quantity, Ordinal, and Cardinal). These types were further divided into 64 subtypes [46] (see Table 2 for the types related to labels in the target domain).

To examine what kind of word embeddings were most valuable to our task, we classified all the available datasets into 3 different groups: *Group 1 General (English Wikipedia+UMBC+One Billion)* was composed of general English materials, which do not contain many domain-specific words or sentences. *Group 2 Biomedical literature (PubMed+PMC)* was composed of biomedical literature and

abstracts. Words in this group could be similar to clinical words but would be used in different ways, considering that publication writing is different from authoring clinical documents. *Group 3 Clinical documents (I2B2+NICTA Train)* was composed of clinical handovers, discharge summaries, and progress notes that closely resemble our task data. All corpora were preprocessed with the Stanford CoreNLP sentence splitter and tokenizer [47]. Digits were replaced with NUM[Length] (eg, 08-08-1988 is replaced by NUM2-NUM2-NUM4), this method helps to capture some digit patterns such as date and phone numbers and will dramatically decrease the amount of words in vocabularies as well. To compute vector representations of word, word2vec [48] was used and modified with an extra option to incrementally train word embeddings based on existing models being given new text materials. We inherited the best parameter settings for named entity recognition from a previous study [24] with 200-word vector dimensions, 5 words in the context window, 10 negative samples, start with a 0.05 learning rate, and run over 20 iterations.

Besides using word vectors as features, we also used a collection of hand-crafted features that were identical to our previous NICTA IE system [25] for performance tracking. For each feature of 1-word instance, a unigram with a window size of 3 *(w_{i-1}, w_i, w_{i+1})*, and bigrams with a window size of 2 *(w_{i-1}w_i, w_iw_{i+1})* were used. Features used in our experiments include the lemma, part of speech tag, and parse tree, top 5 candidates and top mapping retrieved from the Unified Medical Language System (UMLS) [49], medication score—derived from the Anatomical Therapeutic Chemical List, location, and frequency.

To track the performance improvement on this task, the following 10 baselines were included for comparison, they are: 1) *Benchmark*, 2) *TUC-MI-A*, 3) *TUC-MI-B*, 4) *ECNU_ICA-A*, 5) *ECNU_ICA-B*, 6) *LQRZ-A*, 7) *LQRZ-B*, 8) *Unigram NN*, 9) *Random*, 10) *Majority*.

Benchmark

This was the initial NICTA benchmark system on this task using a single-layer linear-chain CRF [50] with L2 regulator with the handcrafted features mentioned before as input. A detailed description of this system can be found in the study by Suominen et al [4].

Participants of Conference and Labs of the Evaluation Forum (CLEF) eHealth Evaluation Lab 2016 Task 1: The CLEF eHealth 2016 Task 1 required the participants to implement systems that are able to identify relevant text snippets from free-text nursing handovers [51]. Participants were expected to train their systems using the given training set, optimize their performances using the validation set, and their final result was tested on a previous confidential test set. It should be noted that the benchmark NICTA IE system was provided to participants in the CLEF task as well as feature generators and intermediate processing results [51]. Participants could start their experiments from any point based on our previous work with very little effort. In fact, all systems except a and b were started from the NICTA benchmark IE system.

- *TUC-MI-A* was based on our benchmark system; rather than using our default features, this method constructed a 41-feature set based on Stanford CoreNLP, latent Dirichlet allocation, regular expressions, and the ontologies of WordNet and UMLS features [10].
- *TUC-MI-B* optimized TUC-MI-A; 19 features were selected from the whole feature set with forward and backward greedy search.
- *ECNU_ICA-A* was a rule-based IE system to recognize bed number, room number, age, and doctor's name and was combined with CRF results using the same feature collection with the organizers' benchmark system [52].
- *ECNU_ICA-B* has the same system architecture as ECNU_ICA-A, except for CRF training, and a subcollection of features was used for different label types [52].
- *LQRZ-A* was a feed-forward neural network with one hidden layer initialized with uniform distribution. Inputs to this NN model are pretrained word embeddings from GoogleNews. No handcrafted features were used in this model [53].
- *LQRZ-B* firstly used a random forest to predict a subset of the tags and the previous NN to further discriminate between the remaining labels [53].

Unigram NN

The unigram NN was an implementation of a 2-layer, first-order linear-chain graph transformer [21] with handcrafted features weighted by word vectors as the first layer and a linear-chain CRF on top of it. The model was trained using AdaGrad. This is a baseline to show separately, from the multilayer NN, what is the performance gain from using word embeddings and transfer learning.

Other Baselines

We evaluated the task difficulty of labeling each word with 1 out of 37 classes by comparing 2 baseline systems: First, we built a system that assigns classes randomly. Second, we implemented another system that always predicts the majority class (ie, the most common class in the training set): *Random* to randomly select 1 class label for each instance and *Majority* to assign the majority class of Future_Goal/TaskToBeCompleted/ExpectedOutcome for every instance.

Results

Researchers worldwide have contributed to achieve a significant improvement on the clinical handover task because of a shared computational task organized in 2016 [51]. In this study, we have reported the results from our experiments on the test set (Table 3) and have also taken this opportunity to overview performance improvements in the task, to summarize methods that have been used to solve the problems so far, and to inspire researchers to work further on this task. Overall, the state-of-the-art benchmark has been increased from 38.2% to 41.6% F1 (P<.001; Wilcoxon text [54]). Our transfer learning method using BBN as source domain (Trans_BBN) outperforms all other methods.

Table 3. Results of transfer learning compared with baseline systems.

Method	MacPrec[a]	MacRec[b]	MacF1[c]	MicPrec[d]	MicRec[e]	MicF1[f]
Trans_BBN	*0.498* [g]	*0.419*	*0.416*	0.547	*0.488*	0.516
Trans_I2B2	0.481	0.390	0.392	0.565	0.471	0.514
TUC-MI-B	0.493	0.369	0.382	0.500	0.505	0.503
ECNU_ICA-A	0.493	0.406	0.374	0.510	0.522	0.516
General+I2B2+train	0.477	0.361	0.354	*0.612*	0.483	*0.540*
I2B2+train	0.443	0.367	0.354	0.604	0.484	0.537
General	0.429	0.356	0.345	0.606	0.478	0.535
LQRZ-B	0.425	0.383	0.345	0.490	0.517	0.503
General+PubMed+PMC	0.409	0.346	0.334	0.606	0.474	0.532
Unigram	0.393	0.292	0.311	0.574	0.448	0.503
TUC-MI-A	0.423	0.300	0.311	0.503	0.443	0.471
LQRZ-A	0.411	0.307	0.308	0.563	0.472	0.514
ECNU_ICA-B	0.428	0.292	0.297	0.581	0.459	0.513
National Information and Communications Technology Australia	0.435	0.233	0.246	0.433	0.368	0.398
Random	0.018	0.028	0.019	0.018	0.030	0.022
Majority	0.000	0.029	0.001	0.016	0.027	0.020

[a]Macro averaged precision.

[b]Macro averaged recall.

[c]Macro averaged F1.

[d]Micro averaged precision.

[e]Micro averaged recall.

[f]Micro averaged F1.

[g]Italics indicate the best result over the column.

Transfer learning with I2B2 as a source model (Trans_I2B2) is also able to increase the overall macro F1 by 4.7% (P<.001) compared with models using a 2-layer NN with general word embeddings (General). When using the same collection of handcrafted features, a 2-layer NN model (Unigram) performs 6.5% better than a single-layer linear-chain CRF (NICTA). The same model (Unigram) gains 3.4% improvement of macro F1 (P<.001) by using word embeddings pretrained with a large text collection with general English (Wiki). Word embeddings trained with a domain-related corpus but different context (Wiki+PubMed+PMC) actually harm rather than help the result. This is possibly because although the domain-related corpus contains medical terms, which are also used in a clinical health care environment, the context of these terms is still very different from clinical handovers. On the contrary, documents used in a similar scenario (I2B2+train) show their advantage at this point. Finally, embeddings trained with a combination of I2B2+train with general English (Wiki_I2B2+train) do not help the system to increase the macro F1, but they yield the best result on micro F1.

Discussion

Principal Findings

It can be seen from the experiment results that the DL system using pretrained word representations as the input, and the proposed transfer learning technique, is able to achieve better performance.

When comparing the results of different system setups on different subclasses, we observed that word representations learned from different domains and the knowledge transferred from various sources affect the clinical IE system on certain subclasses.

Comparing with the best result of feature engineering methods used in TUC-MI [9], our transfer learning method performs 3.4% better without a labor-costing feature-selection procedure. Furthermore, in contrast with the rule-based methods used in ECNU_ICA [52], which require domain-specific experts to inspect data carefully and make the rules, our method is much more efficient and still able to achieve a 4.2% better macro F1. Finally, the best LQRZ [53] used a very similar architecture with our General model, and we can see their performance is very similar as well; the minor difference is caused by different materials to train the word embeddings. Our transfer learning

method is able to improve 7.1% macro F1 on top of the General model (P<.001).

In this section, we have analyzed these results and discussed these effects. The default measure will be the official macro F1 unless specifically mentioned otherwise.

Word Representations

Word embeddings trained from general English can improve the clinical IE performance. Our results show that the *General* model, which used exactly the same model structure and feature map as the *Unigram* model, except it used a combination of 3 large corpora (English Wiki, UMBC, and One Billion) to train general word embeddings, performed better on the overall task (34.5% vs 31.1%, respectively; P<.001). This indicates that word representations trained from unlabeled general English text are able to capture word features that contribute to classifying different annotations in clinical handovers.

Moreover, general word embeddings fine-tuned with a small task-relevant dataset can further increase the result. The model trained with I2B2 and NICTA training data *(I2B2+train)* outperforms the *Unigram* model by 4.3% and outperforms general embeddings when it is compared with the *General* model (35.4% vs 34.5%, respectively; P=.17).

However, no evidence was found to indicate that continuing training word embeddings with a relevant dataset based on pretrained general word embeddings contributes to the system performance when comparing the *I2B2+train* with *General+I2B2+train*. This might be because although the corpora of I2B2 and NICTA training data are significantly smaller than the general English corpus, the vocabulary is still enough to cover words that are present in the test set, and after several iterations of training, word embeddings in these 2 different settings eventually converged to similar values.

Word embeddings trained from domain relevant data do not show any evidence to contribute to improving the system result either. Our results showed that the *General+PubMed+PMC* model performed worse than the *General* model (33.4% vs 34.5%, respectively; P=.07). This might be because even though we considered clinical and biomedical areas as relevant, but because of having different scenarios, vocabulary and context could end up too different. This would introduce more noise to the word embeddings and so does not contribute to the IE performance.

Transfer Learning

Transfer learning shows its advantage in the clinical handover IE task. The top 2 systems were both transfer learning models. Transfer learning from BBN *(Trans_BBN)* was 3.4% higher than the previous best system *TUC-MI-B* (41.6% vs 38.2%, respectively; P<.001).

For the overall result, there is no strong evidence to show any advantage of transfer from domain-relevant source data *(Trans_I2B2)* over general annotations *(Trans_BBN)*. On the contrary, transfer learning from BBN with general annotations performed slightly better than I2B2, which contains more relevant entities with our target task on macro F1 (41.6% vs 39.2%, respectively; P<.001).

For subclasses, Table 4 shows the results of transfer learning compared with the baseline system when the performance is improved on subclasses. When referring to Table 2:

1. Some subclasses where the performance is improved by transfer learning *HAVE* a mapping annotation type from the source domain: for example, subclass PATIENT_INTRODUCTION: Age in years has a mapping annotation DATE:AGE in the source domain BBN, and Trans_BBN on this subclass performed better than the General model (96.5% vs 94.8%, respectively; P<.001). This indicates that when the target domain labels have mappings from the source domain annotations, transfer learning can improve the extraction results of these labels.
2. Some subclasses where the performance is improved by transfer learning *do not have* a mapping annotation type from the source domain: for subclass FUTURE_CARE: Alert/waring/abnormal result, the general model was not able to predict any instance correctly, whereas transfer learning did learn some knowledge from the training set but the performance was still not very high. This might be because these subclasses may have some underlying correlations with source domain labels that are automatically learned during the second process in our method, even though the correlations were not straightforward or obvious for human readers.
3. Some subclasses that have mappings from the source domains do not gain any improvement from transfer learning: for example, PATIENT_INTRODUCTION/ Given_names. These classes normally already have good performance from only using general models, so transfer learning, in this case, might introduce extra noise from other domains that potentially have different sentence structures to the target domain, and thus harm the results.

Table 4. Results of subclasses when transfer learning improved in the baseline system (F1 score).

Entity type	Instances (n)	General	Trans_I2B2	Trans_BBN
PATIENT_INTRODUCTION: Age (years)	246	0.948	0.879	*0.965* [a]
PATIENT_INTRODUCTION: Gender	88	0.826	0.896	*0.917*
PATIENT_INTRODUCTION: Admission reason	412	0.214	0.311	*0.344*
PATIENT_INTRODUCTION: Chronic condition	70	0.000	*0.105*	0.081
PATIENT_INTRODUCTION: Disease/problem history	147	0.016	*0.083*	0.044
PATIENT_INTRODUCTION: Care plan	36	0.069	0.129	*0.133*
PATIENT_INTRODUCTION: Allergy	14	0.267	0.500	*0.566*
APPOINTMENTS: Time	28	0.114	*0.431*	0.400
APPOINTMENTS: Place: Ward	3	0.000	0.000	*0.400*
APPOINTMENTS: Status	159	0.111	*0.175*	0.132
FUTURE_CARE: Alert/warning/abnormal result	59	0.000	0.087	*0.178*
FUTURE_CARE: Goal/task to be completed/expected outcome	496	0.000	0.068	*0.070*
FUTURE_CARE: Discharge/transfer place	89	0.327	0.288	*0.361*
MY_SHIFT: Status	481	0.570	*0.688*	0.638
MY_SHIFT: Input/diet	101	0.413	0.783	*0.804*
MY_SHIFT: Output/diuresis/bowel movement	52	0.286	0.396	*0.478*
MY_SHIFT: Wounds/skin	55	0.444	0.357	*0.457*
MY_SHIFT: Activities of daily living	245	0.579	*0.753*	0.748
MY_SHIFT: Other observation	361	0.177	*0.220*	0.202
MEDICATION: Medicine	156	0.450	*0.548*	0.495
MEDICATION: Status	68	0.034	*0.086*	0.085

[a]Italics indicate the best result over the column.

Conclusions

This study investigated adapting a DL method to extract patient information from clinical reports. Domain and task specification word representations have been used as inputs to a DL system to achieve better performance. In addition, a transfer learning model has been applied to adapt knowledge learned from general text sources to a domain-specific task. This method was able to further improve the overall result, especially in the classes related to the source domain. Domain-specific word representations improve the overall clinical IE system performance by 3.4% on macro-F1. Transferring the knowledge from a general English corpus to our task-specific domain gains a further 7.1% improvement. To our knowledge, our study is the first attempt to transfer knowledge from general deep models to specific tasks in health care and gain a significant improvement. The result of our system is state-of-the-art on this task. Our method and result point out the way toward adapting an advanced ML technique to professional informatics system tasks.

Acknowledgments

This work was supported by the Commonwealth Department of Education and Training (The Australian National University Australian Postgraduate Award). The authors express their gratitude to Maricel Angel, RN at NICTA, for helping them to create the dataset for SR and IE.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Dataset description.

[PDF File (Adobe PDF File), 77KB - medinform_v7i2e11499_app1.pdf]

Multimedia Appendix 2

The test set of our experiments in the format of each row contains the word with its features, and the last column is the human-assigned label.

[ZIP File (Zip Archive), 189KB - medinform_v7i2e11499_app2.zip]

References

1. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
2. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? Am J Med 2018 Feb;131(2):129-133. [doi: 10.1016/j.amjmed.2017.10.035] [Medline: 29126825]
3. Anderson W. 2007 National statement on ethical conduct in human research. Intern Med J 2011 Jul;41(7):581-582. [doi: 10.1111/j.1445-5994.2011.02528.x] [Medline: 21762341]
4. Suominen H, Zhou L, Hanlen L, Ferraro G. Benchmarking clinical speech recognition and information extraction: new data, methods, and evaluations. JMIR Med Inform 2015;3(2):e19 [FREE Full text] [doi: 10.2196/medinform.4321] [Medline: 25917752]
5. Boyle DI, Rafael N. BioGrid Australia and GRHANITE™: privacy-protecting subject matching. Stud Health Technol Inform 2011;168:24-34. [doi: 10.3233/978-1-60750-791-8-24] [Medline: 21893908]
6. I2B2. NLP research data sets URL: https://www.i2b2.org/NLP/DataSets/ [accessed 2018-03-07] [WebCite Cache ID 6xjXRGcPN]
7. Graff D, Cieri C. Linguistic Data Consortium. English gigaword, linguistic data consortium URL: https://catalog.ldc.upenn.edu/LDC2003T05 [accessed 2019-01-28] [WebCite Cache ID 75lT1QCRw]
8. Zheng S, Lu JJ, Ghasemzadeh N, Hayek S, Quyyumi AA, Wang F. Effective information extraction framework for heterogeneous clinical reports using online machine learning and controlled vocabularies. JMIR Med Inform 2017 May 9;5(2):e12 [FREE Full text] [doi: 10.2196/medinform.7235] [Medline: 28487265]
9. Leroy G, Gu Y, Pettygrove S, Galindo M, Arora A, Kurzius-Spencer M. Automated extraction of diagnostic criteria from electronic health records for autism spectrum disorders: development, evaluation, and application. J Med Internet Res 2018 Nov 7;20(11):e10497 [FREE Full text] [doi: 10.2196/10497] [Medline: 30404767]
10. Ebersbach M, Herms R, Lohr C, Eibl M. Wrappers for feature subset selection in CRF-based clinical information extraction. 2016 Presented at: Conference and Labs of the Evaluation Forum (CLEF); September 5-8, 2016; Evora, Portugal p. 69-80 URL: http://ceur-ws.org/Vol-1609/16090069.pdf
11. Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th international conference on Machine learning.: ACM; 2008 Presented at: ICML'08; July 5-9, 2008; Helsinki, Finland p. 160-167.
12. Turian J, Ratinov L, Bengio Y. Word representations: A simple and general method for semi-supervised learning. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics.: the Association for Computational Linguistics (ACL); 2010 Presented at: ACL'10; July 11-16, 2010; Uppsala, Sweden p. 384-394.
13. Baroni M, Dinu G, Kruszewski G. Don't count, predict! A systematic comparison of context-counting vs context-predicting semantic cectors. : Association for Computational Linguistics (ACL); 2014 Presented at: ACL'14; June 22-27, 2014; Baltimore, Maryland, USA p. 1:238-1:247 URL: http://www.aclweb.org/anthology/P14-1023
14. Andreas J, Klein D. How much do word embeddings encode about syntax? : Association for Computational Linguistics (ACL); 2014 Presented at: ACL'14; June 22-27, 2014; Baltimore, Maryland, USA p. 822-827.
15. Qu L, Ferraro G, Zhou L, Hou W, Baldwin T. Named entity recognition for novel types by transfer learning. : Association for Computational Linguistics (ACL); 2016 Presented at: 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP); November 1–5, 2016; Austin, Texas, USA p. 899-905.
16. Arnold A, Nallapati R, Cohen WW. Exploiting feature hierarchy for transfer learning in named entity recognition. : Association for Computational Linguistics (ACL); 2008 Presented at: 46th Annual Meeting of the Association for Computational Linguistics (ACL); June 15-20, 2008; Columbus, Ohio, USA p. 245-253 URL: http://aclweb.org/anthology/P08-1029
17. Suominen H. Kaggle. 2017. NICTA synthetic nursing handover data URL: https://www.kaggle.com/c/hospital-handover-forms [accessed 2018-05-06] [WebCite Cache ID 6zDe1dLzT]
18. Angel M, Suominen H, Zhou L, Hanlen L. CSIRO Data Access Portal. 2014. Synthetic nursing handover training and development data set - text files URL: https://data.csiro.au/collections/ [accessed 2019-01-28] [WebCite Cache ID 75lrDsAGg]
19. Angel M, Suominen H, Zhou L, Hanlen L. CSIRO Data Access Portal. 2014. Synthetic nursing handover training and development data set - audio files URL: https://data.csiro.au/collections/ [accessed 2019-01-28] [WebCite Cache ID 75lrAfyeb]
20. Australian Institute of Health and Welfare. 2016. Chronic diseases URL: https://www.aihw.gov.au/reports-data/health-conditions-disability-deaths/chronic-disease/overview [accessed 2019-01-28] [WebCite Cache ID 75lV7EA2r]

21. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. J Mach Learn Res 2011;12(2011):2493-2537 [FREE Full text]
22. Mikolov T, Chen K, Corrado G, Dean J. arXiv. 2013. Efficient estimation of word representations in vector space URL: https://arxiv.org/pdf/1301.3781.pdf [accessed 2019-03-20] [WebCite Cache ID 770Q1SOtq]
23. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. 2013 Presented at: NIPS'13; December 5-10, 2013; Stateline, NV, USA p. 3111-3119.
24. Qu L, Ferraro G, Zhou L, Hou W, Schneider N, Baldwin T. Big data small data, in domain out-of domain, known word unknown word: The impact of word representations on sequence labelling tasks. 2015 Presented at: 19th Conference on Computational Natural Language Learning (CoNLL); July 30-31, 2015; Beijing, China p. 83-93.
25. Suominen H, Johnson M, Zhou L, Sanchez P, Sirel R, Basilakis J, et al. Capturing patient information at nursing shift changes: methodological evaluation of speech recognition and information extraction. J Am Med Inform Assoc 2014 Oct 21. [doi: 10.1136/amiajnl-2014-002868] [Medline: 25336589]
26. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015 Jan;61:85-117. [doi: 10.1016/j.neunet.2014.09.003] [Medline: 25462637]
27. Zhou L, Xu Q, Suominen H, Gedeon T. EPUTION at SemEval-2018 task 2: Emoji prediction with user adaption. In: Proceedings of The 12th International Workshop on Semantic Evaluation. 2018 Presented at: SemEval 2018; June 5-6, 2018; New Orleans, Louisiana, USA p. 449-453. [doi: 10.18653/v1/S18-1071]
28. Barbieri F, Camacho-Collados J, Ronzano F, Espinosa-Anke L, Ballesteros M, Basile V, et al. SemEval 2018 task 2: Multilingual emoji prediction. In: Proceedings of The 12th International Workshop on Semantic Evaluation. 2018 Presented at: SemEval 2018; June 5-6, 2018; New Orleans, Louisiana, USA p. 24-33. [doi: 10.18653/v1/S18-1003]
29. Maynard D, Tablan V, Ursu C, Cunningham H, Wilks Y. Named entity recognition from diverse text types. In: Proceedings of the Recent Advances in Natural Language Processing 2001 Conference. 2001 Presented at: RANLP 2001; September 5-7, 2001; Tzigov Chark, Bulgaria p. 257-274.
30. Chiticariu L, Krishnamurthy R, Li Y, Reiss F, Vaithyanathan S. Domain adaptation of rule-based annotators for named-entity recognition tasks. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP).: Association for Computational Linguistics (ACL); 2010 Presented at: EMNLP'10; October 9-11, 2010; Cambridge, Massachusetts p. 1002-1012.
31. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. 2014 Presented at: NIPS'14; December 8-13, 2014; Montreal, Quebec, Canada p. 3320-3328.
32. Rizzo G, Troncy R. NERD: A framework for unifying named entity recognition and disambiguation extraction tools. In: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics.: Association for Computational Linguistics (ACL); 2012 Presented at: EACL'12; April 23-27, 2012; Avignon, France p. 73-76.
33. Suominen H, Pahikkala T, Salakoski T. Critical points in assessing learning performance via cross-validation. In: Proceedings of the 2nd International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning.: Helsinki University of Technology; 2008 Presented at: AKKR'08; September 17-19, 2008; Porvoo, Finland p. 9-22.
34. Han L, Kashyap AL, Finin T, Mayfield J, Weese J. UMBC_EBIQUITY-CORE: Semantic textual similarity systems. 2013 Presented at: 2nd Joint Conference on Lexical and Computational Semantics (*SEM); 2013; Atlanta, GA, USA p. 44-52.
35. Chelba C, Mikolov T, Schuster M, Ge Q, Brants T, Koehn P, et al. One billion word benchmark for measuring progress in statistical language modeling. In: Proceedings of the 15th Annual Conference of the International Speech Communication Association. 2014 Presented at: Interspeech 2014; September 14-18, 2014; Singapore p. 2635-2639.
36. Uzuner O, Sibanda TC, Luo Y, Szolovits P. A de-identifier for medical discharge summaries. Artif Intell Med 2008 Jan;42(1):13-35 [FREE Full text] [doi: 10.1016/j.artmed.2007.10.001] [Medline: 18053696]
37. Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. J Am Med Inform Assoc 2007;14(5):550-563 [FREE Full text] [doi: 10.1197/jamia.M2444] [Medline: 17600094]
38. Uzuner O, Solti I, Xia F, Cadag E. Community annotation experiment for ground truth generation for the I2B2 medication challenge. J Am Med Inform Assoc 2010;17(5):519-523 [FREE Full text] [doi: 10.1136/jamia.2010.004200] [Medline: 20819855]
39. Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. J Am Med Inform Assoc 2010;17(5):514-518 [FREE Full text] [doi: 10.1136/jamia.2010.003947] [Medline: 20819854]
40. Zhou L, Suominen H, Hanlen L. Evaluation data and benchmarks for cascaded speech recognition and entity extraction. In: Proceedings of the Third Edition Workshop on Speech, Language & Audio in Multimedia.: ACM; 2015 Presented at: SLAM'15; October 30, 2015; Brisbane, Australia p. 15-18.
41. Wikimedia Downloads. URL: https://dumps.wikimedia.org/ [accessed 2018-03-07] [WebCite Cache ID 6xjWnvCVm]
42. UMBC Ebiquity. UMBC webbase corpus URL: https://ebiquity.umbc.edu/resource/html/id/351 [accessed 2018-03-07] [WebCite Cache ID 6xjX5evo8]

43. Google Research. One billion word benchmark for measuring progress in statistical language modeling URL: https://research.google.com/pubs/pub41880.html [accessed 2018-03-07] [WebCite Cache ID 6xjXGBpnU]
44. NCBI. PubMed URL: https://www.ncbi.nlm.nih.gov/pubmed [accessed 2019-03-19] [WebCite Cache ID 6xjXaRZXU]
45. NCBI. PMC URL: https://www.ncbi.nlm.nih.gov/pmc/ [accessed 2018-03-07] [WebCite Cache ID 6xjXgELxS]
46. Weischedel R, Brunstein A. Linguistic Data Consortium. 2005. BBN pronoun conference and entity type corpus URL: https://catalog.ldc.upenn.edu/LDC2005T33
47. Manning C, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D. The Stanford CoreNLP natural language processing toolkit. 2014 Presented at: 52nd Annual Meeting of the Association for Computational Linguistic system Demonstrations; June 22-27, 2014; Baltimore, Maryland, USA p. 55-60.
48. Google Code. 2013. Tool for computing continuous distributed representations of words URL: https://code.google.com/archive/p/word2vec/ [WebCite Cache ID 748ORV86C]
49. UMLS Terminology Server (and API). 2018. URL: https://umls.terminology.tools/ [accessed 2018-11-23] [WebCite Cache ID 748Oh3vNG]
50. Lafferty JD, McCallum A, Pereira FC. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the Eighteenth International Conference on Machine Learning.: Morgan Kaufmann Publishers Inc; 2001 Presented at: ICML'01; July 1, 2001; Williamstown, MA, USA p. 282-289.
51. Suominen H, Zhou L, Goeuriot L, Kelly L. Task 1 of the CLEF eHealth evaluation lab 2016: Handover information extraction. 2016 Presented at: Conference and Labs of the Evaluation Forum (CLEF); September 5-8, 2016; Evora, Portugal p. 1-14.
52. Song Y, He Y, Liu H, Wang Y, Hu Q, He L, et al. ECNU at 2016 eHealth task 1: Handover information extraction. 2016 Presented at: Conference and Labs of the Evaluation Forum (CLEF); September 5-8, 2016; Evora, Portugal p. 147-156.
53. Quiroz L, Mennes L, Dehghani M. Distributional semantics for medical information extraction. 2016 Presented at: Conference and Labs of the Evaluation Forum (CLEF); September 5-8, 2016; Evora, Portugal p. 109-122.
54. Wilcoxon F. Individual comparisons by ranking methods. Biometrics 1945;1(6):80-83. [doi: 10.1007/978-1-4612-4380-9_16]
55. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
56. Suominen H, Lehtikunnas T, Hiissa M, Back B, Karsten H, Salakoski T, et al. Natural Language Processing for Nursing Documentation. 1945 Presented at: Proceddings of the Second International Conference on Computational Intelligence in Medicine and Health Care (CIMED), Lissabon, Portugal; 2005.
57. Suominen H, Lehtikunnas T, Back B, Karsten H, Salakoski T, Salantera S. Applying Language Technology to Nursing Documents: Pros and Cons with A Focus on Ethics. International Journal of Medical Informatics 2007;76:S293-S301. [Medline: 17604685]

Abbreviations

AdaGrad: adaptive gradient algorithm
BBN: Bolt, Beranek and Newman
CLEF: Conference and Labs of the Evaluation Forum
CRF: conditional random field
DL: deep learning
GPE: geographical entities
IE: information extraction
ML: machine learning
NICTA: National Information and Communications Technology Australia
NN: neural network
NLM: National Library of Medicine
NLP: natural language processing
PMC: PubMed Central
RN: registered nurse
SR: speech recognition
UMBC: University of Maryland, Baltimore County
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 06.07.18; peer-reviewed by Y Chu, M Torii; comments to author 07.10.18; revised version received 02.02.19; accepted 17.02.19; published 25.04.19.

Please cite as:
Zhou L, Suominen H, Gedeon T
Adapting State-of-the-Art Deep Language Models to Clinical Information Extraction Systems: Potentials, Challenges, and Solutions
JMIR Med Inform 2019;7(2):e11499
URL: http://medinform.jmir.org/2019/2/e11499/
doi:10.2196/11499
PMID:31021325

Original Paper

Natural Language Processing for the Identification of Silent Brain Infarcts From Neuroimaging Reports

Sunyang Fu[1], MHI; Lester Y Leung[2], MD; Yanshan Wang[1], PhD; Anne-Olivia Raulli[2], MD; David F Kallmes[3], MD; Kristin A Kinsman[3], MD; Kristoff B Nelson[2], MD; Michael S Clark[3], MD; Patrick H Luetmer[3], MD; Paul R Kingsbury[1], PhD; David M Kent[4], MD; Hongfang Liu[1], PhD

[1]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States

[2]Department of Neurology, Tufts Medical Center, Boston, MA, United States

[3]Department of Radiology, Mayo Clinic, Rochester, MN, United States

[4]Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, United States

Corresponding Author:
Hongfang Liu, PhD
Department of Health Sciences Research
Mayo Clinic
205 3rd Ave SW
Rochester, MN,
United States
Phone: 1 5077730057
Email: Liu.Hongfang@mayo.edu

Abstract

Background: Silent brain infarction (SBI) is defined as the presence of 1 or more brain lesions, presumed to be because of vascular occlusion, found by neuroimaging (magnetic resonance imaging or computed tomography) in patients without clinical manifestations of stroke. It is more common than stroke and can be detected in 20% of healthy elderly people. Early detection of SBI may mitigate the risk of stroke by offering preventative treatment plans. Natural language processing (NLP) techniques offer an opportunity to systematically identify SBI cases from electronic health records (EHRs) by extracting, normalizing, and classifying SBI-related incidental findings interpreted by radiologists from neuroimaging reports.

Objective: This study aimed to develop NLP systems to determine individuals with incidentally discovered SBIs from neuroimaging reports at 2 sites: Mayo Clinic and Tufts Medical Center.

Methods: Both rule-based and machine learning approaches were adopted in developing the NLP system. The rule-based system was implemented using the open source NLP pipeline MedTagger, developed by Mayo Clinic. Features for rule-based systems, including significant words and patterns related to SBI, were generated using pointwise mutual information. The machine learning models adopted convolutional neural network (CNN), random forest, support vector machine, and logistic regression. The performance of the NLP algorithm was compared with a manually created gold standard. The gold standard dataset includes 1000 radiology reports randomly retrieved from the 2 study sites (Mayo and Tufts) corresponding to patients with no prior or current diagnosis of stroke or dementia. 400 out of the 1000 reports were randomly sampled and double read to determine interannotator agreements. The gold standard dataset was equally split to 3 subsets for training, developing, and testing.

Results: Among the 400 reports selected to determine interannotator agreement, 5 reports were removed due to invalid scan types. The interannotator agreements across Mayo and Tufts neuroimaging reports were 0.87 and 0.91, respectively. The rule-based system yielded the best performance of predicting SBI with an accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of 0.991, 0.925, 1.000, 1.000, and 0.990, respectively. The CNN achieved the best score on predicting white matter disease (WMD) with an accuracy, sensitivity, specificity, PPV, and NPV of 0.994, 0.994, 0.994, 0.994, and 0.994, respectively.

Conclusions: We adopted a standardized data abstraction and modeling process to developed NLP techniques (rule-based and machine learning) to detect incidental SBIs and WMDs from annotated neuroimaging reports. Validation statistics suggested a high feasibility of detecting SBIs and WMDs from EHRs using NLP.

(JMIR Med Inform 2019;7(2):e12109) doi:10.2196/12109

KEYWORDS

natural language processing; neuroimaging; electronic health records

Introduction

Background

Silent brain infarction (SBI) is defined as the presence of 1 or more brain lesions, presumed to be because of vascular occlusion, found by neuroimaging (magnetic resonance imaging, MRI or computed tomography, CT) in patients without clinical manifestations of stroke. SBIs are more common than stroke and can be detected on MRI in 20% of healthy elderly [1-3]. Studies have shown that SBIs are associated with increased risk of subsequent stroke, cognitive decline, and deficiency in physical function [1,2]. Despite the high prevalence and serious consequences, there is no consensus on the management of SBI as routinely discovering SBIs is challenged by the absence of corresponding diagnosis codes and the lack of the knowledge about the characteristics of the affected population, treatment patterns, or the effectiveness of therapy [1]. Even though there is strong evidence shows that antiplatelet and statin therapies are effective in preventing recurrent stroke in patients with prior stroke, the degree to which these results might apply to patients with SBI is unclear. Although SBI is understood by some clinicians to be pathophysiologically identical to stroke (and thus similarly treated), others view SBI as an incidental neuroimaging finding of unclear significance. The American Heart Association/American Stroke Association has identified SBI as a major priority for new studies on stroke prevention because the population affected by SBI falls between primary and secondary stroke prevention [4].

In addition to SBI, white matter disease (WMD) or leukoaraiosis is another common finding in neuroimaging of elderly. Similar to SBI, WMD is usually detected incidentally on brain scans and is commonly believed to be a form of microvascular ischemic brain damage resulting from typical cardiovascular risk factors [5]. WMD is associated with subcortical infarcts due to small vessel disease and is predictive of functional disability, recurrent stroke, and dementia [6-8]. SBI and WMD are related, but it is unclear whether they result from the same, independent, or synergistic processes [9,10]. As with SBI, there are no proven preventive treatments or guidelines regarding the initiation of risk factor–modifying therapies when WMD is discovered.

Objectives

Identifying patients with SBI is challenged by the absence of corresponding diagnosis codes. One reason is that SBI-related incidental findings are not included in a patient's problem list or other structured fields of electronic health records (EHRs); instead, the findings are captured in neuroimaging reports. A neuroimaging report is a type of EHR data that contains the interpretation and finding from neuroimage such as CT and MRI in unstructured text. Incidental SBIs can be detected by the review of neuroradiology reports obtained in clinical practice, typically performed manually by radiologists or neurologists. However, manually extracting information from patient narratives is time-consuming, costly, and lacks robustness and standardization [11-14]. Natural language processing (NLP) has been leveraged to perform chart review for other medical conditions by automatically extracting important clinical concepts from unstructured text. Researchers have used NLP systems to identify clinical syndromes and biomedical concepts from clinical notes, radiology reports, and surgery operative notes [15]. An increasing amount of NLP-enabled clinical research has been reported, ranging from identifying patient safety occurrences [16] to facilitating pharmacogenomic studies [17]. Our study focuses on developing NLP algorithms to routinely detect incidental SBIs and WMDs.

Methods

Study Setting

This study was approved by the Mayo Clinic and Tufts Medical Center (TMC) institutional review boards. This work is part of the Effectiveness of Stroke PREvention in Silent StrOke project, which is to use NLP techniques to identify individuals with incidentally discovered SBIs from radiology reports, at 2 sites: Mayo Clinic and TMC.

Gold Standard

The detailed process of generating the gold standard is described in Multimedia Appendix 1. The gold standard annotation guideline was developed by 2 subject matter experts: a vascular neurologist (LYL) and a neuroradiologist (PHL), and the annotation task was performed by 2 third-year residents (KAK, MSC) from Mayo and 2 first-year residents (AOR, KN) from TMC. Each report was annotated with 1 of the 3 labels for SBI (positive SBI, indeterminate SBI, or negative SBI) and one of the 3 labels for WMD (positive WMD, indeterminate WMD, or negative WMD).

The gold standard dataset includes 1000 radiology reports randomly retrieved from the 2 study sites (500 from Mayo Clinic and 500 from TMC) corresponding to patients with no prior or current diagnosis of stroke or dementia. To calculate interannotator agreement (IAA), 400 out of the 1000 reports were randomly sampled and double read. The gold standard dataset was equally split to 3 subsets for training (334), developing (333), and testing (333).

Experimental Methods

We compared 2 NLP approaches. One was to define the task an information extraction (IE) task, where a rule-based IE system can be developed to extract SBI or WMD findings. The other was to define the task as a sentence classification task, where sentences can be classified to contain SBI or WMD findings.

Rule-Based Information Extraction

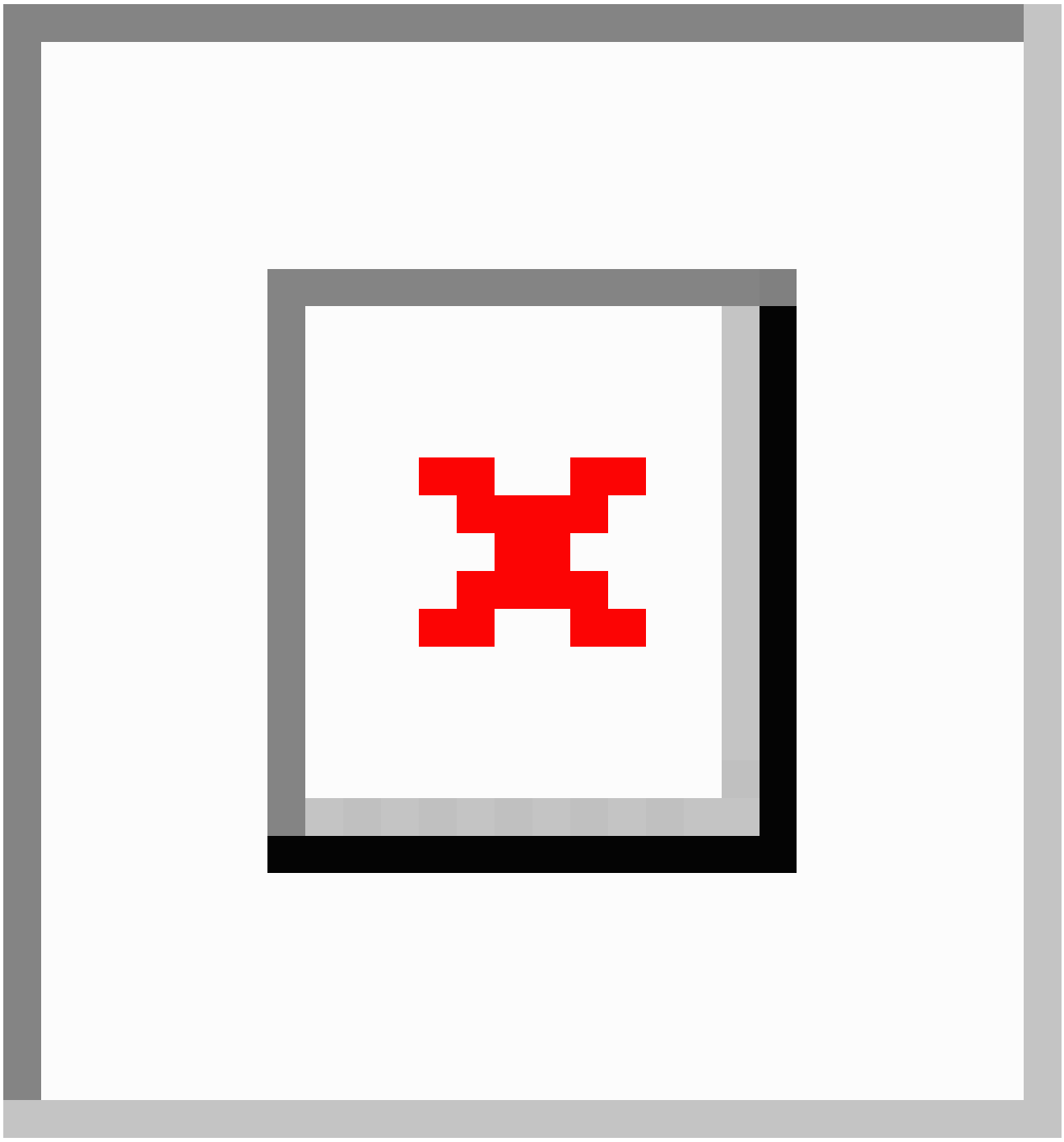
We adopted the open source NLP pipeline, MedTagger, as the infrastructure for the rule-based system implementation. MedTagger is a resource-driven, open source unstructured information management architecture–based IE framework [18]. The system separates task-specific NLP knowledge engineering

from the generic NLP process, which enables words and phrases containing clinical information to be directly coded by subject matter experts. The tool has been utilized in the eMERGE consortium to develop NLP-based phenotyping algorithms [19]. Figure 1 shows the process workflow. The generic NLP process includes sentence tokenization, text segmentation, and context detection. The task-specific NLP process includes the detection of concept mentions in the text using regular expressions and normalized to specific concepts. The summarization component applies heuristic rules for assigning the labels to the document.

For example, the sentence "probable right old frontal lobe subcortical infarct as described above," is processed as an SBI concept with the corresponding contextual information with status as "probable," temporality as "present," and experiencer as "patient."

The domain-specific NLP knowledge engineering was developed following 3 steps: (1) Prototype algorithm development, (2) Formative algorithm development using the training data, and (3) Final algorithm evaluation. We leveraged pointwise mutual information [20] to identify significant words and patterns associated with each condition for prototyping the algorithm (Multimedia Appendix 2). The algorithm was applied to the training data. False classified reports were manually reviewed by 2 domain experts (LYL, PHL). Keywords were manually curated through an iteratively refining process until all issues were resolved. The full list of concepts, keywords, modifiers, and diseases categories are listed in Textbox 1.

Figure 1. Rule system process flow. SBI: silent brain infarction; WMD: white matter disease.



Textbox 1. Silent brain infarction (SBI) and white matter disease (WMD) risk factor and indication keywords.

- Confirmation keywords—disease-finding SBI: infarct, infarcts, infarctions, infarction, lacune, lacunes
- Confirmation keywords—disease modifier SBI: acute, acute or subacute, recent, new, remote, old, chronic, prior, chronic foci of, benign, stable small, stable
- Confirmation keywords—disease location SBI: territorial, lacunar, cerebellar, cortical, frontal, caudate, right frontoparietal lobe, right frontal cortical, right frontal lobe, embolic, left basal ganglia lacunar, basal ganglia lacunar, left caudate and left putamen lacunar
- Confirmation keywords—disease-finding WMD: leukoaraiosis, white matter, microvascular ischemic, microvascular leukemic, microvascular degenerative
- Exclusion WMD: degenerative changes

Machine Learning

The machine learning (ML) approach allows the system to automatically learn robust decision rules from labeled training data. The task was defined as a sequential sentence classification task. We adopted Kim's convolutional neural network (CNN) [21] and implemented using TensorFlow 1.1.02 [22]. The model architecture, shown in Figure 2, is a variation of the CNN architecture of Collobert R [23].

We also adopted 3 traditional ML models—random forest [24], support vector machine [25] and logistic regression [26]—for baseline comparison. All models used word vector as input representation, where each word from the input sentence is represented as the k-dimensional word vector. The word vector is generated from word embedding, a learned representation for text where words that have the same meaning have a similar representation. Suppose $x_1, x_2, \ldots, x_n$ is the sequence of word representations in a sentence where

$$x_i = E_{xi}, I = 1,2, \ldots, n.$$

Here, E_{xi} is the word embedding representation for word x_i with the dimensionality d. In our ML experiment, we used Wang's word embedding trained from Mayo Clinic clinical notes where d=100 [27]. The embedding model is the skip-gram of word2vec, an architecture proposed by Mikolov T [28]. Let $x_{i:i+k-1}$ represent a window of size k in the sentence. Then the output sequence of the convolutional layer is

$$con_i = f(w_k x_{i:i+k-1} + b_k),$$

where f is a rectify linear unit function, w_k and b_k are the learning parameters. Max pooling was then performed to record the largest number from each feature map. By doing so, we obtained fixed length global features for the whole sentence, that is,

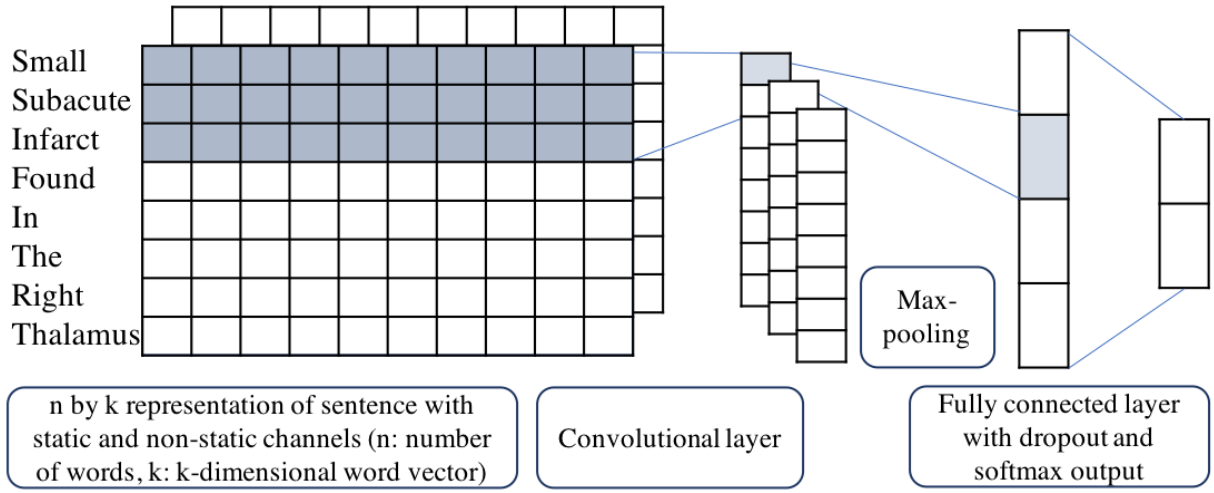
$$m_k = \max_{1 \le i \le n-k+1}(con_i).$$

Then the features are fit into a fully connected layer with the output being the final feature vector $O=wm_k + b$. Finally, a softmax function is utilized to make final classification decision, that is,

$$p(sbi \mid x,\theta) = e^{(O_{sbi})}/(e^{(O_{sbi})}+e^{(O_{other})}),$$

where θ is a vector of the hyper parameters of the model, such as w_k, b_k, w and b.

Figure 2. Convolutional neural network architecture with 2 channels for an example sentence.



Evaluation Metric

For evaluation of the quality of the annotated corpus, Cohen kappa was calculated to measure the IAA during all phases [29]. As the primary objective of the study is case ascertainment, we calculated the IAA at the report level.

A 2 x 2 confusion matrix was used to calculate performance score for model evaluation: positive predictive value (PPV), sensitivity, negative predictive value (NPV), specificity, and accuracy using manual annotation as the gold standard. The McNemar test was adopted to evaluate the performance difference between the rule-based and ML models [30,31]. To

have a better understanding of the potential variation between neuroimaging reports and neuroimages, we compared the model with the best performance (rule-based) with neuroimaging interpretation. A total of 12 CT images and 12 MRI images were stratified—randomly sampled from the test set. A total of 2 attending neurologists read all 24 images and assigned the SBI and WMD status. The cases with discrepancies were adjudicated by the neuroradiologist (PHL) The agreement was assessed using kappa and F-measure [32].

Results

Interannotator Agreements Across Neuroimaging Reports

Among the total 400 double-read reports, 5 reports were removed because of invalid scan types. The IAAs across Mayo and Tufts neuroimaging reports were 0.87 and 0.91. Overall, there is a high agreement between readers on both reports (Tables 1 and 2). Age-specific prevalence of SBI and WMD is provided in Multimedia Appendix 2.

Table 1. Interreader agreement across 207 Mayo neuroimaging reports.

Interannotator agreement	Computed tomography (n=63)		Magnetic resonance imaging (n=144)		Total (n=207)	
	% agree	kappa	% agree	kappa	% agree	kappa
Silent brain infarction	98.4	0.92	97.2	0.83	97.6	0.87
White matter disease	100.0	1.00	98.6	0.97	99.0	0.98

Table 2. Interreader agreement across 188 Tufts Medical Center neuroimaging reports.

Interannotator agreement	Computed tomography (n=80)		Magnetic resonance imaging (108)		Total (n=188)	
	% agree	kappa	% agree	kappa	% agree	kappa
Silent brain infarction	98.8	0.79	99.1	0.94	99.5	0.91
White matter disease	100.0	1.00	99.1	0.98	99.5	0.99

Natural Language Processing System Performance

Overall, the rule-based system yielded the best performance of predicting SBI with an accuracy of 0.991. The CNN achieved the best score on predicting WMD (0.994). Full results are provided in Table 3.

According to the McNemar test, we found the difference between rule-based system and CNN on SBI is considered to be statistically significant (*P* value=.03). We found no statistically significant difference between the rest of the models.

Table 4 lists the evaluation results of NLP and gold standard derived from reports against the neuroimaging interpretation for SBI and WMD. Both NLP and gold standard had moderate-high agreements with the neuroimaging interpretation, with kappa scores around .5. Our further analysis showed the practice graded findings (gold standard and NLP) achieved high precision and moderate recall scores compared with the neuroimaging interpretation. Through the confirmation with Mayo and TMC radiologists, we believed such discrepancy was because of the inconsistency in documentation standards related to clinical incidental findings, causing SBIs and WMDs underreported.

Table 3. Performance on test dataset against human annotation as gold standard.

Evaluation of natural language processing, model name	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Accuracy
Silent brain infarction (n=333)					
Rule-based system	0.925	1.000	1.000	0.990	0.991
CNN[a]	0.650	0.993	0.929	0.954	0.952
Logistic regression	0.775	0.983	0.861	0.970	0.958
SVM[b]	0.825	1.000	1.000	0.977	0.979
Random forest	0.875	1.000	1.000	0.983	0.986
White matter disease (n=333)					
Rule-based system	0.942	0.909	0.933	0.921	0.928
CNN	0.994	0.994	0.994	0.994	0.994
Logistic regression	0.906	0.865	0.896	0.877	0.888
SVM	0.864	0.894	0.917	0.830	0.877
Random forest	0.932	0.880	0.913	0.906	0.910

[a]CNN: convolutional neural network.

[b]SVM: support vector machine.

Table 4. Comparison of the neuroimaging interpretation with gold standard and natural language processing.

Evaluation of natural language processing against the neuroimaging interpretation	F-measure	kappa	Precision	Recall
Silent brain infarction (n=24)				
Gold standard	0.74	0.50	0.92	0.69
NLP[a]	0.74	0.50	0.92	0.69
White matter disease (n=24)				
Gold standard	0.78	0.56	0.86	0.80
NLP	0.74	0.49	0.85	0.73

[a]NLP: natural language processing.

Discussion

Machine Learning Versus Rule

In summary, the rule-based system achieved the best performance of predicting SBI, and the CNN model yielded the highest score of predicting WMD. When detecting SBI, the ML models were able to achieve high specificity, NPV, and PPV but moderate sensitivity because of the small number of positive cases. Oversampling is a technique to adjust the class distribution of training data to balance the ratio between positive and negative cases [33]. This technique was applied to the training data to help boost the signals of positive SBIs. The performance was slightly improved but was limited by the issue of overfitting, a situation when a model learns the training data too well. Due to that, unnecessary details and noises in the training data can create negative impact to the generalizability of the model. In our case, the Mayo reports have larger language variation (noise) because of a free style of documentation method, whereas TMC uses a template-based documentation method. According to the sublanguage analysis, Mayo had 212 unique expressions for describing no acute infarction, whereas TMC had only 12. Therefore, the model trained on oversampled data had a bias toward the expressions that only appeared in the training set. When predicting WMD, the ML model outperformed the rule-based model. The reason is because the dataset for WMD is more balanced than SBI (60% positive cases), which allows the system to equally learn from both classes (positive and negative). The overall performance on WMD is better than SBI because WMDs are often explicitly documented as important findings in the neuroimaging report.

False Prediction Analysis

Coreference resolution was the major challenge to the rule-based model for identifying SBIs. Coreference resolution is an NLP task to determine whether 2 mentioned concepts refer to the same real-world entity. For example, in Textbox 2, “The above findings” refers to “where there is an associated region of nonenhancing encephalomalacia and linear hemosiderin disposition.” To determine if a finding is SBI positive, the system needs to extract both concepts and detect their coreference relationship.

Textbox 2. Example of coreference resolution.

> "Scattered, nonspecific T2 foci, most prominently in the left parietal white matter *<Concept 1>where there is an associated region of nonenhancing encephalomalacia and linear hemosiderin disposition. <Concept 1/>* Linear hemosiderin deposition overlying the right temporal lobe (series 9, image 16) as well. No abnormal enhancement today. *<Concept 2>The above findings are nonspecific but the evolution, hemosiderin deposition, and gliosis suggest post ischemic change. <Concept 2>*"

For the ML system, the false positives from the identification of SBIs were commonly contributed by disease locations. As the keywords *foci*, *right occipital lobe*, *right parietal lobe*, *right subinsular region*, and *left frontal region* often coexisted with SBI expressions, the model assigned higher weights to these concepts when the model was trained. For example, the expression: "there are a bilateral intraparenchymal foci of susceptibility artifact in the right occipital lobe, right parietal lobe, right subinsular region and left frontal region" has 4 locations with no mention of "infarction" appearing in the sentence. The ML system still predicted it as SBI positive. Among all ML models, the CNN yielded the worse NPV, which suggested the CNN was more likely to receive false signals from disease locations. Our next step is to further refine the system by increasing the volume of training size through leveraging distant supervision to obtain additional SBI positive cases.

Limitations

Our study has several limitations. First, despite the high feasibility of detecting SBIs from neuroimaging reports, there is a variation between NLP-labeled neuroimaging reports and neuroimages. Second, the performances of the ML models are limited by the number of annotated datasets. Additional training data are required to have a comprehensive comparison between the rule-based and ML systems. Third, the systems were only evaluated using datasets from 2 sites; the generalizability of the systems may be limited.

Conclusions

We adopted a standardized data abstraction and modeling process to developed NLP techniques (rule-based and ML) to detect incidental SBIs and WMDs from annotated neuroimaging reports. Validation statistics suggested a high feasibility of detecting SBIs and WMDs from EHRs using NLP.

Acknowledgments

The authors would like to gratefully acknowledge National Institutes of Health grant 1R01NS102233 and Donna M Ihrke for case validation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Gold standard development.

[PDF File (Adobe PDF File), 221KB - medinform_v7i2e12109_app1.pdf]

Multimedia Appendix 2

Supplementary result.

[PDF File (Adobe PDF File), 465KB - medinform_v7i2e12109_app2.pdf]

References

1. Fanning JP, Wesley AJ, Wong AA, Fraser JF. Emerging spectra of silent brain infarction. Stroke 2014 Nov;45(11):3461-3471. [doi: 10.1161/STROKEAHA.114.005919] [Medline: 25293663]
2. Fanning J, Wong A, Fraser J. The epidemiology of silent brain infarction: a systematic review of population-based cohorts. BMC Med 2014 Jul 9;12:119 [FREE Full text] [doi: 10.1186/s12916-014-0119-0] [Medline: 25012298]
3. Vermeer S, Longstreth JW, Koudstaal P. Silent brain infarcts: a systematic review. Lancet Neurol 2007 Jul;6(7):611-619. [doi: 10.1016/S1474-4422(07)70170-9] [Medline: 17582361]
4. Furie K, Kasner S, Adams R, Albers G, Bush R, Fagan S, American Heart Association Stroke Council, Council on Cardiovascular Nursing, Council on Clinical Cardiology,Interdisciplinary Council on Quality of CareOutcomes Research. Guidelines for the prevention of stroke in patients with stroke or transient ischemic attack: a guideline for healthcare professionals from the american heart association/american stroke association. Stroke 2011 Jan;42(1):227-276. [doi: 10.1161/STR.0b013e3181f7d043] [Medline: 20966421]
5. Gouw A, van der Flier WM, Fazekas F, van Straaten EC, Pantoni L, Poggesi A, LADIS Study Group. Progression of white matter hyperintensities and incidence of new lacunes over a 3-year period: the Leukoaraiosis and Disability study. Stroke 2008 May;39(5):1414-1420. [doi: 10.1161/STROKEAHA.107.498535] [Medline: 18323505]

6. Hijdra A, Verbeeten B, Verhulst J. Relation of leukoaraiosis to lesion type in stroke patients. Stroke 1990 Jun;21(6):890-894. [doi: 10.1161/01.STR.21.6.890] [Medline: 2349592]
7. Miyao S, Takano A, Teramoto J, Takahashi A. Leukoaraiosis in relation to prognosis for patients with lacunar infarction. Stroke 1992 Oct;23(10):1434-1438. [doi: 10.1161/01] [Medline: 1412580]
8. Fu J, Lu C, Hong Z, Dong Q, Luo Y, Wong K. Extent of white matter lesions is related to acute subcortical infarcts and predicts further stroke risk in patients with first ever ischaemic stroke. J Neurol Neurosurg Psychiatry 2005 Jun;76(6):793-796 [FREE Full text] [doi: 10.1136/jnnp.2003.032771] [Medline: 15897500]
9. Chen Y, Wang A, Tang J, Wei D, Li P, Chen K, et al. Association of white matter integrity and cognitive functions in patients with subcortical silent lacunar infarcts. Stroke 2015 Apr;46(4):1123-1126. [doi: 10.1161/STROKEAHA.115.008998] [Medline: 25737316]
10. Conklin J, Silver F, Mikulis D, Mandell D. Are acute infarcts the cause of leukoaraiosis? Brain mapping for 16 consecutive weeks. Ann Neurol 2014 Dec;76(6):899-904. [doi: 10.1002/ana.24285] [Medline: 25283088]
11. Xu H, Jiang M, Oetjens M, Bowton E, Ramirez A, Jeff J, et al. Facilitating pharmacogenetic studies using electronic health records and natural-language processing: a case study of warfarin. J Am Med Inform Assoc 2011;18(4):387-391 [FREE Full text] [doi: 10.1136/amiajnl-2011-000208] [Medline: 21672908]
12. Grishman R, Huttunen S, Yangarber R. Information extraction for enhanced access to disease outbreak reports. J Biomed Inform 2002 Aug;35(4):236-246 [FREE Full text] [doi: 10.1016/S1532-0464(03)00013-3] [Medline: 12755518]
13. South B, Shen S, Jones M, Garvin J, Samore M, Chapman W, et al. Developing a manually annotated clinical document corpus to identify phenotypic information for inflammatory bowel disease. Summit Transl Bioinform 2009 Mar 1;2009:1-32 [FREE Full text] [doi: 10.1186/1471-2105-10-S9-S12] [Medline: 21347157]
14. Gilbert E, Lowenstein S, Koziol-McLain J, Barta D, Steiner J. Chart reviews in emergency medicine research: where are the methods? Ann Emerg Med 1996 Mar;27(3):305-308. [doi: 10.1016/S0196-0644(96)70264-0] [Medline: 8599488]
15. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Dec;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
16. Murff H, FitzHenry F, Matheny M, Gentry N, Kotter K, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. J Am Med Assoc 2011 Aug 24;306(8):848-855. [doi: 10.1001/jama.2011.1204] [Medline: 21862746]
17. Denny J, Ritchie M, Basford M, Pulley J, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 2010 May 1;26(9):1205-1210 [FREE Full text] [doi: 10.1093/bioinformatics/btq126] [Medline: 20335276]
18. Ferrucci D, Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. Nat Lang Eng 1999;10(3-4):327-348. [doi: 10.1017/S1351324904003523]
19. 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]
20. Church K, Hanks P. Word association norms, mutual information, and lexicography. Comput Linguist 1990;16(1):22-29 [FREE Full text]
21. Kim Y. Association for Computational Linguistics. 2014. Convolutional neural networks for sentence classification URL: https://www.aclweb.org/anthology/D14-1181.pdf [accessed 2019-04-08] [WebCite Cache ID 77TSHIa8O]
22. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation. 2016 Presented at: OSDI'16; November 2–4, 2016; Savannah, GA, USA.
23. Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning. In: Proceedings of the 25th international conference on Machine learning. A unified architecture for natural language processing: Deep neural networks with multitask learning. Proceedings of the 25th international conference on Machine learning; 2008 Presented at: ICML'08; July 5-9, 2008; Helsinki, Finland. [doi: 10.1145/1390156.1390177]
24. Breiman L. Random forests. Mach Learn 2001;45(1):5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
25. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273-297 [FREE Full text] [doi: 10.1023/A:1022627411411]
26. Hosmer JD, Lemeshow S, Sturdivant R. Applied Logistic Regression. Hoboken, New Jersey: John Wiley & Sons; 2013.
27. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform 2018 Nov;87:12-20. [doi: 10.1016/j.jbi.2018.09.008] [Medline: 30217670]
28. Mikolov T, Chen K, Corrado G, Dean J. arXiv. 2013. Efficient estimation of word representations in vector space URL: https://arxiv.org/pdf/1301.3781.pdf [accessed 2019-04-08] [WebCite Cache ID 77TWweBoa]
29. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 2;20(1):37-46. [doi: 10.1177/001316446002000104]
30. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947 Jun;12(2):153-157. [doi: 10.1007/BF02295996] [Medline: 20254758]

31. Dietterich T. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 1998;10(7):1895-1923. [doi: 10.1162/089976698300017197]
32. Sasaki Y. Old Dominion University. 2007. The truth of the F-measure URL: https://www.researchgate.net/publication/268185911_The_truth_of_the_F-measure [accessed 2019-04-08] [WebCite Cache ID 77TXSTW3V]
33. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. arXiv. 2002. SMOTE: synthetic minority over-sampling technique URL: https://arxiv.org/pdf/1106.1813.pdf [accessed 2019-04-08] [WebCite Cache ID 77TXcYm3e]

Abbreviations

CNN: convolutional neural network
CT: computed tomography
EHR: electronic health record
IAA: interannotator agreement
IE: information extraction
MRI: magnetic resonance imaging
NLP: natural language processing
NPV: negative predictive value
PPV: positive predictive value
SBI: silent brain infarction
TMC: Tufts Medical Center
WMD: white matter disease

Edited by G Eysenbach; submitted 05.09.18; peer-reviewed by J Zheng, V Vydiswaran, D Oram; comments to author 07.01.19; revised version received 26.02.19; accepted 30.03.19; published 21.04.19.

Please cite as:
Fu S, Leung LY, Wang Y, Raulli AO, Kallmes DF, Kinsman KA, Nelson KB, Clark MS, Luetmer PH, Kingsbury PR, Kent DM, Liu H
Natural Language Processing for the Identification of Silent Brain Infarcts From Neuroimaging Reports
JMIR Med Inform 2019;7(2):e12109
URL: http://medinform.jmir.org/2019/2/e12109/
doi:10.2196/12109
PMID:31066686

Original Paper

Structure and Content of Drug Monitoring Advices Included in Discharge Letters at Interfaces of Care: Exploratory Analysis Preceding Database Development

Benedict Morath[1,2,3]; Katharina Wien[1,2]; Torsten Hoppe-Tichy[2,3], Dr rer nat; Walter Emil Haefeli[1,2], MD, FBPhS; Hanna Marita Seidling[1,2], PD Dr sc hum

[1]Department of Clinical Pharmacology and Pharmacoepidemiology, Heidelberg University Hospital, Heidelberg University, Heidelberg, Germany
[2]Cooperation Unit Clinical Pharmacy, Heidelberg University Hospital, Heidelberg University, Heidelberg, Germany
[3]Hospital Pharmacy, Heidelberg University Hospital, Heidelberg University, Heidelberg, Germany

Corresponding Author:
Hanna Marita Seidling, PD Dr sc hum
Department of Clinical Pharmacology and Pharmacoepidemiology
Heidelberg University Hospital
Heidelberg University
Im Neuenheimer Feld 410
Heidelberg, 69120
Germany
Phone: 49 62215638736
Fax: 49 6221564642
Email: hanna.seidling@med.uni-heidelberg.de

Abstract

Background: Inadequate drug monitoring of drug therapy after hospital discharge facilitates adverse drug events and preventable hospital readmissions.

Objective: This study aimed to analyze the structure and content of drug monitoring advices of a representative sample of discharge letters as a basis for future electronic information systems.

Methods: On 2 days in November 2016, all discharge letters of 3 departments of a university hospital were extracted from the hospital information system. The frequency, content, and structure of drug monitoring advices in discharge letters were investigated and compared with the theoretical monitoring requirements expressed in the corresponding summaries of product characteristics (SmPC). The quality of the drug monitoring advices in the discharge letters was rated with the domains of an adapted systematic instructions for monitoring (SIM) score.

Results: In total, 154 discharge letters were analyzed containing 1180 brands (240 active pharmaceutical substances), of which 50.42% (595/1180) could theoretically be amended with a monitoring advice according to the SmPC. In reality, 40 discharge letters (26.0%, 40/154) contained a total of 66 monitoring advices for 57 brands (4.83%, 57/1180), comprising 18 different monitoring parameters. Drug monitoring advices only addressed mean 1.9 (SD 0.8) of the 7 domains of the SIM score and frequently did not address reasons for monitoring (86%, 57/66), the timing of monitoring, that is, the start (76%, 50/66), the frequency (94%, 63/66), the stop (95%, 63/66), and how to react (83%, 55/66).

Conclusions: Drug monitoring advices were mostly absent in discharge letters and a gold standard for appropriate drug monitoring advices was lacking. Hence, more effort should be put in the development of tools that facilitate easy presentation of clinically meaningful drug monitoring advices at the point of care.

(JMIR Med Inform 2019;7(2):e10832) doi:10.2196/10832

KEYWORDS

drug monitoring; patient discharge summaries; transition of care

Introduction

Background

Adverse drug events (ADE) frequently occur after the patient transitions across interfaces of care, thus making patients prone to unintended outcomes such as hospital readmissions [1,2]. Indeed, up to 10% of all hospital readmissions occur as a consequence of ADE, and nearly 1 in 4 of these ADE is caused by drugs just started during the index hospitalization [1,3-5]. Since during hospitalization, more than 95% of all prehospital drug therapies are modified, appropriate follow-up monitoring is particularly important [6-9]. Furthermore, in the discharge medication, over half of the drugs are newly prescribed during hospitalization, emphasizing the need for closer monitoring during the initial postdischarge phase [8]. However, after hospital discharge, monitoring of safety (ADE) and efficacy is often lacking, thus causing potentially preventable readmissions [10]. Interestingly, in the ambulatory setting, preventable ADE resulting from inadequate monitoring and leading to hospitalization are more likely associated with commonly prescribed drugs such as drugs with a cardiovascular indication [1,3,11-13]. For instance, about one-third of patients treated with angiotensin-converting-enzyme (ACE) inhibitors do not undergo serum creatinine and potassium controls at least yearly; although, it is well established that monitored patients experience ADE less often [14-17].

Objectives

Hence, it appears useful to include structured and comprehensive drug monitoring advices in discharge letters concerning the safety and efficacy of drug therapy to support general practitioners with drug therapy monitoring and ensure a safe patient transfer across the interfaces of care. Today, the current state of drug monitoring recommendations at interfaces is not well known and except for specific diseases or drugs, a comprehensive and prospectively evaluated gold standard for evidence-based drug monitoring advices is lacking [18,19]. As a first step to develop and subsequently provide suitable drug monitoring advices at interfaces of care, we performed an exploratory analysis of the structure and the patterns of current drug monitoring advices in discharge letters and compared this information with the statutory information provided in the pertinent summary of product characteristics (SmPC).

Methods

Context

We analyzed an exploratory sample of consecutive discharge letters of 3 major departments of a large university hospital to determine the number, structure, and content of the drug monitoring advices that are currently provided in daily practice. Therefore, discharge letters of the divisions of hemato-oncology, gastroenterology, cardiology, endocrinology, general medicine, psychosomatics, visceral surgery, vascular surgery, cardiac surgery, urology, and neurology were included in the analysis. Although a German drug monitoring guideline was published in 2013 by the German College of General Practitioners and Family Physicians, a comprehensive and prospectively evaluated gold standard for evidence-based drug monitoring advices in different settings of care is not established at present in Germany [20]. Therefore, the information in discharge letters was compared with the generic drug monitoring parameters of the SmPC. This study was approved by the responsible Ethics Committee of the Medical Faculty of Heidelberg University (S-402/2016).

Setting and Data Collection of Drug Monitoring Parameters in the Discharge Letter

As a point prevalence analysis, all final discharge letters of the departments of, surgery, internal medicine, and neurology that were issued on November 15 and November 16 of the year 2016 and stored in the hospital information system were screened by 1 author. The departments were chosen to cover a broad spectrum of medications of different specialties and generate a representative overview. All discharge letters containing a discharge medication were selected, printed, and pseudonymized by blacking data of the attending physicians and the patient and attributing a consecutive number code to every letter.

The entire discharge letter was independently read by 2 investigators and screened for drug monitoring advices. The following information was extracted into a predefined Microsoft Excel (Microsoft Corporation) sheet with the following categories: Code of the discharge letter, name of all drugs listed as discharge medication including their strength, dosage, and additional information such as administration advices, as well as potential drug monitoring advices with their content and placement in the letter (eg, directly adjunct to a drug or included in the prose text).

A drug monitoring advice was defined as a statement that was explicitly (eg, “please monitor serum potassium under ramipril therapy”) or by placement (ie, proximity) connected with the recommended drug treatment at discharge. Second, a drug monitoring advice needed to explicitly state tests that should be performed (eg, electrocardiogram) or parameters that should be checked (eg, potassium) either in terms of safety and ADE monitoring (eg, liver function test to detect hepatotoxicity) or in terms of efficacy (eg, target low-density lipoprotein values to identify poor or nonresponders).

We did not differentiate in drug monitoring advices for newly prescribed drugs and those that were already on the patients’ medication list at the time of hospital admission.

Structure and Content of Drug Monitoring Advices in Discharge Letters

To determine the structure and content of current drug monitoring advices, the drug monitoring advices were independently categorized by 2 authors using the domains of an adapted version of the systematic instructions for monitoring (SIM) score [21]. The SIM score contains 7 essential domains of information, which should be addressed in an unequivocal and comprehensive drug monitoring advice: (1) why to monitor, (2) what to monitor, (3) when to start monitoring, (4) how frequently to monitor, (5) what to look for, that is, target values in terms of drug efficacy or specific ADE such as laboratory changes, (6) how to respond to findings, and (7) when to stop drug monitoring. When analyzing the drug monitoring advices,

we specified for every category whether it was included in the drug monitoring advice (=1 point) or not (=0 points).

Extraction of Summary of Product Characteristics Information

The SmPC of all brands reported in the discharge medications were independently screened by 2 authors. When no brand name and only an active pharmaceutical substance was provided in the discharge medication, the SmPC of the brand listed in the hospital formulary was screened because this was the last specific brand the patient received. All eligible text passages concerning drug monitoring of the respective brands were transferred into an excel sheet once a consensus of the 2 reviewers was reached. If consensus was not reached, a third reviewer was involved.

In analogy to the discharge letters, a drug monitoring advice was defined as a parameter that should be measured or a test that should invariably be performed for safety or efficacy reasons at a given time during or after the treatment. Extracted drug monitoring parameters and tests are available in Multimedia Appendix 1.

Analysis and Statistics

Drug monitoring parameters and tests were rated as concordant in discharge letters and the SmpC if (1) the drug monitoring parameter or test was stated explicitly in the SmPC and the discharge letter, for example, “measure potassium” or (2) if the SmPC or the discharge letter recommended drug monitoring parameters or tests that were related to each other. As an example, when the SmPC recommended potassium controls and the drug monitoring advices in the discharge letter recommended controls of electrolytes, these drug monitoring advices were also rated as concordant. The allocation of monitoring parameters was done independently by 2 investigators. The frequency of drug monitoring parameters and tests was determined and averages with SDs were calculated using Microsoft Excel. Cohen kappa was calculated to determine interrater reliability of the SIM score rating.

Results

Characteristics of the Included Discharge Letters and Discharge Medications

On the 2 index days, 158 discharge letters were issued and hence screened for inclusion. Yet, 4 of these discharge letters did not contain any discharge medication and were therefore excluded, leaving 154 discharge letters for analysis. There were 34 discharge letters from the surgery department (22.1%, 34/154), 95 from internal medicine (61.7%, 95/154), and 25 (16.2%, 25/154) from neurology. Overall, the discharge letters contained 1180 brands referring to 240 different active pharmaceutical substances from 51 different 3-digit anatomical therapeutic chemical code (ATC) groups (see Figure 1), resulting in an average of 7.7 (SD 4.3) brands per discharge letter. The most commonly prescribed brands were antithrombotic agents (B01, n=161), drugs for acid-related disorders (A02, n=87), agents acting on the renin-angiotensin system (C09, n=85), beta-blocking agents (C07, n=81), and diuretics (C03, n=80, Figure 1).

Drug Monitoring Advices Provided in the Discharge Letter

Overall, 40 discharge letters (25.9%, 40/154) contained at least 1 drug monitoring advice for, in total, 57 brands (4.83%, 57/1180), and 29 active pharmaceutical substances (details are shown in Multimedia Appendix 2). Phenprocoumon (n=6), tacrolimus (n=5), and levothyroxine (n=5) were the active pharmaceutical substances most frequently accompanied by a drug monitoring advice (see Table 1). Drug monitoring advices most frequently suggested monitoring of renal function (n=9), trough concentrations (n=7), international normalized ratio (n=6), and blood glucose (n=6). Most drug monitoring advices were solely located in the text (n=34), some were included in the discharge medication (n=14), and the advices were rarely found in both text and discharge medication section (n=9; Table 1), leading to a total of 66 drug monitoring advices with a total of 69 suggested drug monitoring parameters and tests (referring to 18 different parameters and tests).

Structure and Content of Drug Monitoring Advices in Discharge Letters

Of the 66 drug monitoring advices, 20 addressed 1 domain, 29 addressed 2 domains, and 15 addressed 3 domains of the SIM score. Only 1 drug monitoring advice addressed 4 domains (what to monitor, when to start monitoring, how frequently to monitor, and when to stop monitoring), and 1 drug monitoring advice (ie, “we ask for regular endocrinological follow-up controls”) was too vague and hence did not meet any of the SIM domains. On average, the drug monitoring advices addressed 1.9 (SD 0.8) domains (see Figure 2).

Nearly all drug monitoring advices (99%, 65/66) contained a definition of the monitoring parameters or tests that should be performed. Around a quarter of the drug monitoring advices specified when drug monitoring should be started (24%, 16/66) and what should be looked for (29%, 19/66, Figure 2). Only few drug monitoring advices gave reasons of drug monitoring, that is, why to monitor (14%, 9/66), or described which actions to take in case of findings, that is, how to respond to deviations (17%, 11/66). Adequate timing of drug monitoring was seldom addressed; almost all drug monitoring advices lacked information on the frequency of monitoring, that is, how frequently to monitor (94%, 62/66) and when monitoring may be stopped (95%, 63/66). Interrater reliability was very good with a Cohen kappa of 0.89.

Figure 1. Most common prescribed drug groups (expressed as 3-digit anatomical therapeutic chemical code class) in 154 consecutive discharge letters of 3 large university departments (internal medicine, neurology, and surgery). ATC: anatomical therapeutic chemical code, RAS: renin-angiotensin system.

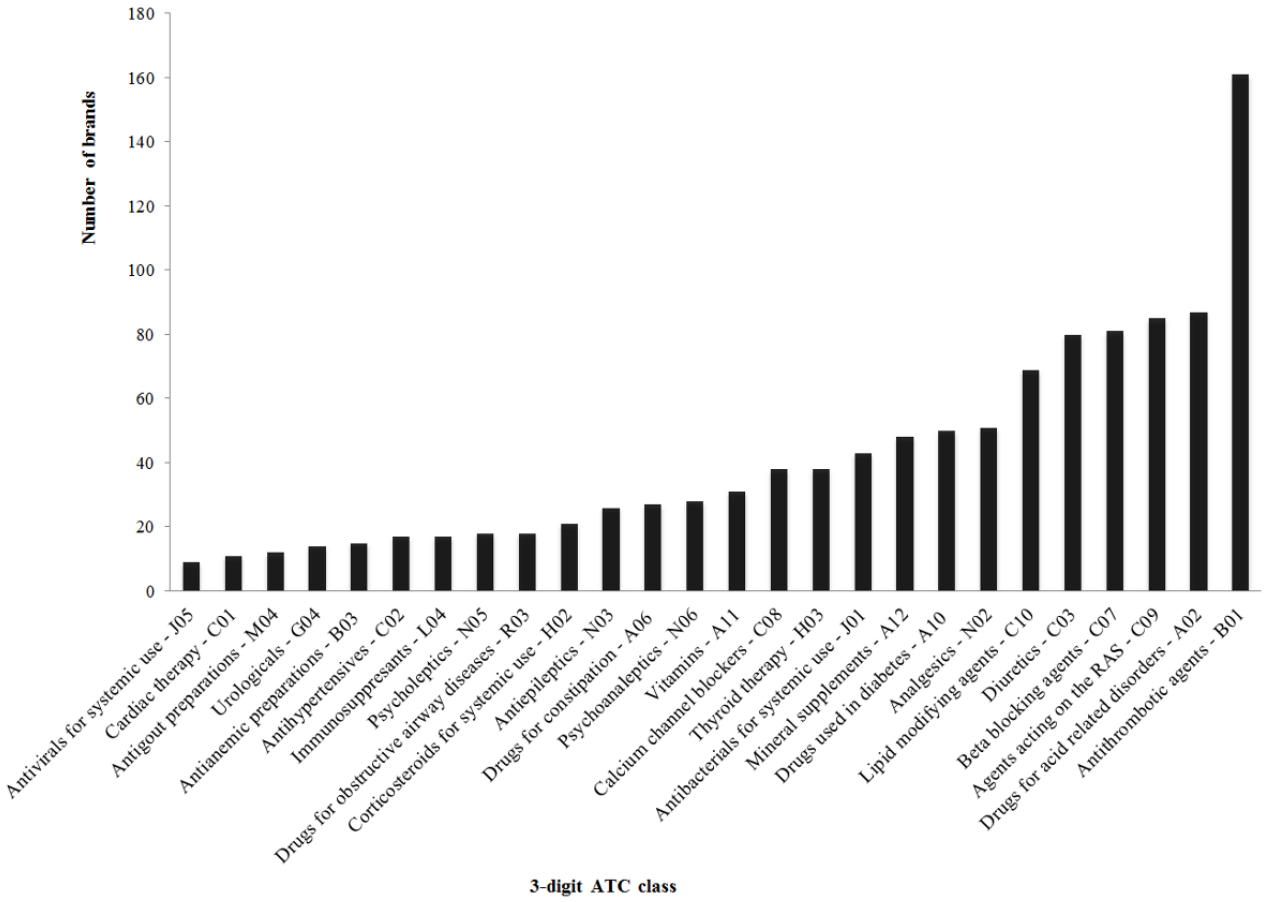


Table 1. Comparison of the drug monitoring parameters reported in the discharge letters with the monitoring recommendations in the corresponding summary of product characteristics.

Active pharmaceutical substance (frequency)[a]	Location in the discharge letter (frequency)[b]	Drug class	Drug monitoring parameter in discharge letter (frequency)[c]	Drug monitoring parameters in the summary of product characteristics
Apixaban (1/8)	TXT[d] (1)	Factor Xa inhibitor	Blood cell count (1)	Bleeding signs
Atorvastatin (3/37)	MED[e] (3), TXT (1)	HMG-CoA[f]-reductase inhibitor	CK[g] (1), LDL[h] (3), liver function (1)	CK, liver function test
Candesartan (1/20)	MED (1), TXT (1)	Angiotensin-II- receptor antagonist	Blood pressure (1), renal function (1)	Only in special patient populations (hypertension and impaired renal function, heart failure)
Carvedilol (1/10)	MED (1), TXT (1)	Nonselective beta blocker	Blood pressure (1), heart rate (1)	Only in special patient populations (heart failure with low blood pressure or ischemic heart disease)
Cefuroxime (1/4)	TXT (1)	Cephalosporin	Inflammatory parameters (1)	No parameters mentioned
Ciclosporin (2/2)	MED (1), TXT (1)	Calcineurin inhibitor	Blood concentration (2)	Serum potassium, serum magnesium, serum lipids, uric acid, renal function, liver function, ciclosporin concentrations, blood pressure, and physical examination
Ciprofloxacin (2/12)	TXT (2)	Fluorquinolone	Inflammatory parameters (2), renal function (1)	No parameters mentioned
Clindamycin (1/1)	TXT (1)	Lincosamide	Inflammatory parameters (1)	Blood cell count, liver function test, and renal function test
Colecalciferol (1/8)	TXT (1)	Vitamin	Serum calcium (1)	Calcium in serum and urine, creatinine
Dabigatran etexilate (1/3)	MED (1)	Thrombin inhibitor	Liver function (1), renal function (1)	Renal function, signs and symptoms of bleeding or anemia
Duloxetine (2/3)	TXT (2)	Selective serotonin and norepinephrine reuptake inhibitor	Serum sodium (2)	Only in special patient populations (old patients, hypertension, or heart disease)
Enoxaparin (2/36)	TXT (2)	Low-molecular-weight heparin	Blood cell count (2)	Platelet count
Eplerenone (1/4)	MED (1)	Mineralocorticoid receptor antagonist	Renal function (1), electrolytes (1)	Serum potassium
Furosemide (2/11)	TXT (2)	Loop diuretics	Renal function (1), electrolytes (1), body weight (1)	Potassium, sodium, calcium, bicarbonate, creatinine, blood urea, uric acid, and blood glucose
Hydrochlorothiazide (1/14)	MED (1), TXT (1)	Thiazide diuretic	Renal function (1), electrolytes (1),	Serum potassium, serum sodium, and serum magnesium
Insulin (6/30)	TXT (6)	Insulin	Blood glucose (6)	Blood glucose
Pancreatic enzyme supplement (2/6)	TXT(2)	Enzymes	Stool consistency (2)	No parameters mentioned
Levetiracetam (1/8)	TXT (1)	Antiepileptic	Renal function (1)	Suicidal ideation
Levothyroxine (5/34)	MED (1), TXT (5)	Thyroid hormone	Thyroid function (5)	No parameters mentioned
Nebivolol (1/7)	TXT (1)	Selective beta blocker	Heart rate (1)	No parameters mentioned
Oxcarbazepine (1/1)	TXT (1)	Antiepileptic	Serum sodium (1)	Serum sodium, suicidal ideation
Phenprocoumon (6/13)	MED (2), TXT (5)	Vitamin K antagonist	INR[i] (6)	Liver function test, INR
Pravastatin (1/16)	MED (1)	HMG-CoA-reductase inhibitor	LDL (1)	Only in special patient population (patients with myopathy, impaired renal function, hypothyroidism, or alcohol abuse)
Ramipril (1/44)	MED (1)	Angiotensin-converting-enzyme inhibitor	Blood pressure (1)	Serum potassium, renal function, and leukocytes

Active pharmaceutical substance (frequency)[a]	Location in the discharge letter (frequency)[b]	Drug class	Drug monitoring parameter in discharge letter (frequency)[c]	Drug monitoring parameters in the summary of product characteristics
Sildenafil (1/2)	MED (1)	Phosphodiesterase type 5 inhibitor	Blood pressure (1), heart rate (1)	No parameters mentioned
Simvastatin (1/9)	MED (1)	HMG-CoA-reductase inhibitor	LDL (1)	CK, liver function test
Spironolactone (2/13)	TXT (2)	Mineralocorticoid receptor antagonist	Renal function (1), electrolytes (1), body weight (1)	Potassium, sodium, calcium, bicarbonate, creatinine, blood urea, uric acid, and acid-base balance
Tacrolimus (5/6)	MED (5), TXT (2)	Calcineurin inhibitor	Blood concentration (5)	Electrolytes, liver function, renal function, fasting blood glucose, hematological parameters, coagulation, plasma proteins, blood concentration, blood pressure, ECG[j], neurologic status, and vision
Torasemide (2/32)	MED (2), TXT (1)	Loop diuretic	Body weight (1), electrolytes (1), renal function (1)	Electrolytes, creatinine, uric acid, blood glucose, lipids, leukocytes, erythrocytes, and platelets

[a]The first number in parenthesis shows the number of active pharmaceutical substances with a drug monitoring advice; the second number in parenthesis indicates the total amount of discharge letters, which had the active pharmaceutical substance included.

[b]The number in parenthesis indicates how often the drug monitoring advice was located in the text or in the discharge medication.

[c]The number in parenthesis indicates how often the drug monitoring parameter was recommended for the corresponding active pharmaceutical substance.

[d]TXT: text.

[e]MED: discharge medication.

[f]HMG-CoA: hydroxymethylglutaryl-coenzyme A.

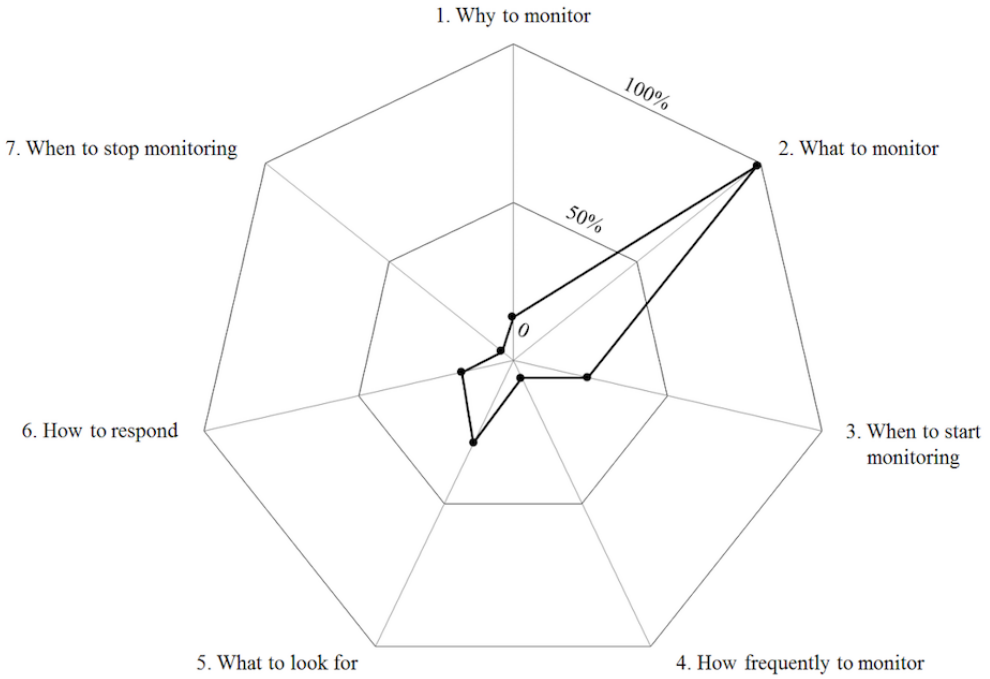
[g]CK: creatine kinase.

[h]LDL: low-density lipoprotein.

[i]INR: international normalized ratio.

[j]ECG: electrocardiogram.

Figure 2. Overview of the frequency of systematic instructions for monitoring score domains used in the drug monitoring advices of the discharge letters.



Comparison With Summary of Product Characteristics Information

For 52 of the 57 brands with an actual drug monitoring advice in the discharge letter, the SmPC also mentioned a drug monitoring advice, but of the 69 drug monitoring parameters and tests mentioned in the discharge letters, only 35 parameters (51%, 35/69) were also listed in the corresponding SmPC. In contrast, 29 of the 71 SmPC parameters (41%, 29/71) were included in the discharge letters, whereas the remaining 42 parameters (59%, 42/71) were not mentioned in the discharge letters at all (Table 1 and Multimedia Appendix 2). However, the SmPC suggested drug monitoring advices for many more drugs. Indeed, for 595 of the 1180 brands (50.42%) included in the discharge letters (referring to 132 out of 240 active pharmaceutical substances; 55.0% (132/240), the SmPC contained suggestions for drug monitoring that could be theoretically applied.

Discussion

Principal Findings

Drug monitoring advices were provided only for about 1 in 20 brands recommended in the discharge medication and most often did not offer all information domains a "best-practice" monitoring advice should contain. Following the SIM score, the advices only infrequently specified what to look for, why one should monitor, and what should be done in case of finding deviations. Regarding the timing of drug monitoring, a quarter of the advices specified a start of drug monitoring, but only 1 in 20 advices stated a frequency or defined an end of drug monitoring.

Evidence Gap Regarding the Need of Drug Monitoring

To finally judge the quality of drug monitoring advices and also to subsequently derive an evidence-based information support tool, information on monitoring advices that have been shown to be clinically relevant is needed. Thereby, clinically relevant information might be particularly determined by the severity and probability of the potential ADE or efficacy loss as well as the chances that the ADE or efficacy loss can be reliably detected and prevented by the monitoring activity. The SmPC rather follows a generic approach and suggest up to 10 times more advices than were currently included in the letters. However, it remains unclear whether these advices are all clinically relevant and need to be followed in all patients [21,22]. Conversely, there are first hints suggesting that even the SmPC lacks relevant advices that are included in clinical guidelines. For instance, the 2016 heart failure guideline of the European Society of Cardiology recommends close monitoring of creatinine, serum potassium, and urea upon ACE inhibitor therapy initiation, which was not similarly mentioned in the German ramipril SmPC [23]. Although clinical guidelines might be expected to be a good source for clinically meaningful drug monitoring advices, this aspect is not a standard request for good guideline development, which mainly focuses on proper guideline development methods [24], and preliminary analyses suggest that drug monitoring advices are included only sporadically and certainly not systematically.

In daily practice, there are only few drugs with precise and unambiguous monitoring recommendations in the SmPC, for example, agranulocytosis monitoring with clozapine [25]. However, in most drugs, the monitoring need is vague and a specification requires clinical context factors such as (1) patient characteristics, (2) stage of therapy (eg, dose titration), and (3) comedication. This is also reflected by the discrepancies between the mentioned drug monitoring parameters and tests in the discharge letters and the SmPC in this study, which can be attributed to the evaluation of clinical context factors. As an example, blood glucose monitoring was recommended for a patient under insulin therapy in a discharge letter. This was consistent with the SmPC recommendations, but it also might appear rather obvious and lead to alert fatigue if integrated routinely in discharge letters. Regarding the clinical context factors, the respective patient had had pancreatectomy and therefore a clear clinical indication for close glucose monitoring in the postoperative phase, justifying the explicit drug monitoring advice.

This study therefore supports the hypothesis that there is an evidence gap in terms of a consistent definition of indications for drug monitoring and populations benefitting of it. Therefore, to close this gap, future research should address changes in ADE incidence over time and evaluate protective and risk factors that might have an impact on the need of drug monitoring. There are first approaches to develop such information tools, such as a recent recommendation providing suggestions for drug monitoring of high-risk medicines in primary care, which were derived from a range of guideline sources and expert opinions [20,26].

Concept and Structure of Comprehensive and Practical Drug Monitoring Advices

If drug monitoring is indicated, the drug monitoring advices should be clearly formulated and support physicians in the development of individual monitoring plans. A comprehensive drug monitoring advice should follow the information clusters suggested in the SIM score [21]. The need of drug monitoring defined by a sole indication of a drug, for example, clozapine therapy, or a combination of clinical context factors define the domains "why to monitor" (SIM score domain 1) and what to monitor (SIM score domain 2). The stage of therapy (eg, drug initiation, maintenance, or tapering) is an important determinant regarding the proper timing of drug monitoring activities and specifies the start (SIM score domain 3), the frequency (SIM score domain 4), and the end of drug monitoring (SIM score domain 7). Timing is a crucial aspect of any monitoring because the risk of ADE varies over time as some drugs have a high risk of ADE early after drug initiation, for example, hyperkalemia with ramipril intake or dosage changes, whereas, other ADE more likely occur after longer time periods, for example, pulmonary toxicity caused by amiodarone [27-30]. The SIM score domain "what to look for" (SIM score domain 5) and "how to respond" (SIM score domain 6) are domains that were rarely addressed and, if addressed, sufficient information was lacking. For instance, the drug monitoring advice "please check liver function" lacks detailed information on what explicitly to look for because drug-induced liver injury occurs in different clinical patterns such as hepatic, cholestatic, or mixed, which

can easily be detected by characteristic laboratory patterns [31,32].

Limitations of the Study Design

This study has several limitations. First, a sample of 154 discharge letters was analyzed, which could be deemed as relatively small. To ensure representability, we included consecutive discharge letters of 2 working days rather than deliberately choosing letters of different patient populations; this approach covered a broad range of different brands (n=1180) and a sizeable number of different 3-digit ATC-codes (n=51). Moreover, the sample size was estimated on the basis of previous studies analyzing the quality and structure of drug monitoring advices in drug labels, which had similar or even lower sample sizes [21,22,33]. Second, direct clinical implications of missing drug monitoring parameters neither were nor could have been assessed in this study, and they neither were in the focus of our study. Furthermore, the clinical implications of infrequent drug monitoring are well known, and there is no obvious reason to omit proper monitoring of pharmacotherapy after discharge from tertiary care [3,34]. Therefore, the study focused on the structure and content of drug monitoring advices at interfaces of care to analyze potential areas for improvement and interventions, targeting the problem of infrequent drug monitoring in patient care. Third, we did not consider the date of onset of a specific drug as this information was scarcely available in the discharge letters. ADE of some active pharmaceutical substances (eg, hyperkalemia with ramipril intake) might occur more likely during dosage titration and monitoring periods could be longer, when long-term maintenance doses are taken uneventfully [27,29]. Consequently, it might be possible that the real monitoring need was overestimated or, on the other hand, that drug monitoring advices were not precise enough. Finally, we solely evaluated the drug monitoring advices of one other data source, that is, the SmPC. Yet, as the legally binding document also in terms of drug therapy monitoring, it could be the first reference consulted by prescribers and information therein should be reliable also in this regard.

Conclusions

Drug monitoring advices were included in discharge letters only for a minority of brands; however, respective SmPC information was broad and unspecific in most parts, suggesting that a future monitoring database should consider not only the drug and its indication but also further patient characteristics, the stage of therapy, and the comedication.

Authors' Contributions

BM planned the study design, extracted the discharge letters, analyzed the data, and wrote the manuscript. KW analyzed the data and wrote the manuscript. THT planned the study design. He wrote parts of the manuscript and critically revised it. WEH planned the study design, analyzed the data, wrote the manuscript, and critically revised it. HS planned the study design, analyzed the data, wrote the manuscript, and critically revised it.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of summary of parameters, tests, and symptoms to monitor mentioned in the summary of product characteristics.

[PDF File (Adobe PDF File), 105KB - medinform_v7i2e10832_app1.pdf]

Multimedia Appendix 2

Overview of the drug monitoring advices recommended in 154 of 158 consecutive discharge letters of a university hospital.

[PDF File (Adobe PDF File), 268KB - medinform_v7i2e10832_app2.pdf]

References

1. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. Adverse drug events occurring following hospital discharge. J Gen Intern Med 2005 Apr;20(4):317-323 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.30390.x] [Medline: 15857487]
2. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med 2003 Feb 4;138(3):161-167. [Medline: 12558354]
3. Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annest JL. National surveillance of emergency department visits for outpatient adverse drug events. J Am Med Assoc 2006 Oct 18;296(15):1858-1866. [doi: 10.1001/jama.296.15.1858] [Medline: 17047216]
4. Dormann H, Neubert A, Criegee-Rieck M, Egger T, Radespiel-Tröger M, Azaz-Livshits T, et al. Readmissions and adverse drug reactions in internal medicine: the economic impact. J Intern Med 2004 Jun;255(6):653-663 [FREE Full text] [doi: 10.1111/j.1365-2796.2004.01326.x] [Medline: 15147529]

5. Davies EC, Green CF, Mottram DR, Rowe PH, Pirmohamed M. Emergency re-admissions to hospital due to adverse drug reactions within 1 year of the index admission. Br J Clin Pharmacol 2010 Nov;70(5):749-755 [FREE Full text] [doi: 10.1111/j.1365-2125.2010.03751.x] [Medline: 21039769]
6. Himmel W, Kochen MM, Sorns U, Hummers-Pradier E. Drug changes at the interface between primary and secondary care. Int J Clin Pharmacol Ther 2004 Feb;42(2):103-109. [Medline: 15180171]
7. Himmel W, Tabache M, Kochen MM. What happens to long-term medication when general practice patients are referred to hospital? Eur J Clin Pharmacol 1996;50(4):253-257. [Medline: 8803514]
8. Grimmsmann T, Schwabe U, Himmel W. The influence of hospitalisation on drug prescription in primary care - a large-scale follow-up study. Eur J Clin Pharmacol 2007 Aug;63(8):783-790. [doi: 10.1007/s00228-007-0325-1] [Medline: 17549465]
9. von Klüchtzner W, Grandt D. Influence of hospitalization on prescribing safety across the continuum of care: an exploratory study. BMC Health Serv Res 2015 May 13;15:197 [FREE Full text] [doi: 10.1186/s12913-015-0844-x] [Medline: 25962594]
10. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. J Am Med Assoc 2007 Feb 28;297(8):831-841. [doi: 10.1001/jama.297.8.831] [Medline: 17327525]
11. Thomsen LA, Winterstein AG, Søndergaard B, Haugbølle LS, Melander A. Systematic review of the incidence and characteristics of preventable adverse drug events in ambulatory care. Ann Pharmacother 2007 Sep;41(9):1411-1426. [doi: 10.1345/aph.1H658] [Medline: 17666582]
12. Kanaan AO, Donovan JL, Duchin NP, Field TS, Tjia J, Cutrona SL, et al. Adverse drug events after hospital discharge in older adults: types, severity, and involvement of Beers Criteria Medications. J Am Geriatr Soc 2013 Nov;61(11):1894-1899 [FREE Full text] [doi: 10.1111/jgs.12504] [Medline: 24116689]
13. Gurwitz JH, Field TS, Harrold LR, Rothschild J, Debellis K, Seger AC, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. J Am Med Assoc 2003 Mar 5;289(9):1107-1116. [Medline: 12622580]
14. Raebel MA, McClure DL, Simon SR, Chan KA, Feldstein A, Andrade SE, et al. Laboratory monitoring of potassium and creatinine in ambulatory patients receiving angiotensin converting enzyme inhibitors and angiotensin receptor blockers. Pharmacoepidemiol Drug Saf 2007 Jan;16(1):55-64. [doi: 10.1002/pds.1217] [Medline: 16470693]
15. Raebel MA, Lyons EE, Andrade SE, Chan KA, Chester EA, Davis RL, et al. Laboratory monitoring of drugs at initiation of therapy in ambulatory care. J Gen Intern Med 2005 Dec;20(12):1120-1126 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0257.x] [Medline: 16423101]
16. Tjia J, Fischer SH, Raebel MA, Peterson D, Zhao Y, Gagne SJ, et al. Baseline and follow-up laboratory monitoring of cardiovascular medications. Ann Pharmacother 2011 Sep;45(9):1077-1084. [doi: 10.1345/aph.1Q158] [Medline: 21852593]
17. Raebel MA, Ross C, Xu S, Roblin DW, Cheetham C, Blanchette CM, et al. Diabetes and drug-associated hyperkalemia: effect of potassium monitoring. J Gen Intern Med 2010 Apr;25(4):326-333 [FREE Full text] [doi: 10.1007/s11606-009-1228-x] [Medline: 20087674]
18. Hiemke C, Bergemann N, Clement HW, Conca A, Deckert J, Domschke K, et al. Consensus guidelines for therapeutic drug monitoring in neuropsychopharmacology: update 2017. Pharmacopsychiatry 2018 Jan;51(1-02):9-62 [FREE Full text] [doi: 10.1055/s-0043-116492] [Medline: 28910830]
19. Feuerstein JD, Nguyen GC, Kupfer SS, Falck-Ytter Y, Singh S, American Gastroenterological Association Institute Clinical Guidelines Committee. American gastroenterological association institute guideline on therapeutic drug monitoring in inflammatory bowel disease. Gastroenterology 2017 Dec;153(3):827-834. [doi: 10.1053/j.gastro.2017.07.032] [Medline: 28780013]
20. Mainz A, Kötter T, Brockmann S, Egidi G. German Society for General Medicine and Family Medicine. Drug monitoring URL: http://www.degam.de/files/Inhalte/Leitlinien-Inhalte/Dokumente/DEGAM-S1-Handlungsempfehlung/053-037%20Medikamentenmonitoring/S1-HE_Medikamentenmonitoring_Langfassung_201406.pdf [accessed 2018-04-09] [WebCite Cache ID 6yXo0PTX4]
21. Ferner RE, Coleman J, Pirmohamed M, Constable SA, Rouse A. The quality of information on monitoring for haematological adverse drug reactions. Br J Clin Pharmacol 2005 Oct;60(4):448-451 [FREE Full text] [doi: 10.1111/j.1365-2125.2005.02440.x] [Medline: 16187979]
22. Geerts AF, De Koning FH, Van Solinge WW, De Smet PA, Egberts TC. Instructions on laboratory monitoring in 200 drug labels. Clin Chem Lab Med 2012 Feb 2;50(8):1351-1358. [doi: 10.1515/cclm-2011-0753] [Medline: 22868799]
23. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, ESC Scientific Document Group. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016 Dec 14;37(27):2129-2200. [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]
24. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P, Board of Trustees of the Guidelines International Network. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med 2012 Apr 3;156(7):525-531. [doi: 10.7326/0003-4819-156-7-201204030-00009] [Medline: 22473437]
25. Food and Drug Administration. Clozaril - Summary of product characteristics URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2010/019758s062lbl.pdf [accessed 2019-01-17] [WebCite Cache ID 75UsbGcvy]

26. Erskine D. Specialist Pharmacy Service. 2017 Oct. Suggestions for therapeutic drug monitoring in adults in primary care URL: https://www.sps.nhs.uk/wp-content/uploads/2017/12/Drug-monitoring-final-December-2017a.pdf [accessed 2018-04-05] [WebCite Cache ID 6ySBO6VEp]
27. Park IW, Sheen SS, Yoon D, Lee SH, Shin GT, Kim H, et al. Onset time of hyperkalaemia after angiotensin receptor blocker initiation: when should we start serum potassium monitoring? J Clin Pharm Ther 2014 Feb;39(1):61-68. [doi: 10.1111/jcpt.12109] [Medline: 24262001]
28. Epstein AE, Olshansky B, Naccarelli GV, Kennedy Jr JI, Murphy EJ, Goldschlager N. Practical management guide for clinicians who treat patients with amiodarone. Am J Med 2016 May;129(5):468-475. [doi: 10.1016/j.amjmed.2015.08.039] [Medline: 26497904]
29. Barber J, McKeever TM, McDowell SE, Clayton JA, Ferner RE, Gordon RD, et al. A systematic review and meta-analysis of thiazide-induced hyponatraemia: time to reconsider electrolyte monitoring regimens after thiazide initiation? Br J Clin Pharmacol 2015 Apr;79(4):566-577 [FREE Full text] [doi: 10.1111/bcp.12499] [Medline: 25139696]
30. Connolly S, Cairns J, Gent M, Roberts R, Yusuf S. Effect of prophylactic amiodarone on mortality after acute myocardial infarction and in congestive heart failure: meta-analysis of individual data from 6500 patients in randomised trials. Amiodarone Trials Meta-Analysis Investigators. Lancet 1997 Nov 15;350(9089):1417-1424. [Medline: 9371164]
31. Leise MD, Poterucha JJ, Talwalkar JA. Drug-induced liver injury. Mayo Clin Proc 2014 Jan;89(1):95-106. [doi: 10.1016/j.mayocp.2013.09.016] [Medline: 24388027]
32. Fontana RJ, Seeff LB, Andrade RJ, Björnsson E, Day CP, Serrano J, et al. Standardization of nomenclature and causality assessment in drug-induced liver injury: summary of a clinical research workshop. Hepatology 2010 Aug;52(2):730-742 [FREE Full text] [doi: 10.1002/hep.23696] [Medline: 20564754]
33. Nederlof M, Stoker LJ, Egberts TC, Heerdink ER. Instructions for clinical and biomarker monitoring in the summary of product characteristics (SmPC) for psychotropic drugs: overview and applicability in clinical practice. J Psychopharmacol 2015 Dec;29(12):1248-1254. [doi: 10.1177/0269881115609016] [Medline: 26464457]
34. Auerbach AD, Kripalani S, Vasilevskis EE, Sehgal N, Lindenauer PK, Metlay JP, et al. Preventability and causes of readmissions in a national cohort of general medicine patients. JAMA Intern Med 2016 Apr;176(4):484-493. [doi: 10.1001/jamainternmed.2015.7863] [Medline: 26954564]

Abbreviations

ACE: angiotensin-converting-enzyme
ADE: adverse drug event
ATC: anatomical therapeutic chemical code
CK: creatine kinase
ECG: electrocardiogram
HMG-CoA: hydroxymethylglutaryl-coenzyme A
INR: international normalized ratio
LDL: low-density lipoprotein
MED: discharge medication
SIM: systematic instructions for monitoring
SmPC: summary of product characteristics
TXT: text

Edited by G Eysenbach; submitted 20.04.18; peer-reviewed by Y Li; comments to author 09.10.18; revised version received 06.11.18; accepted 10.12.18; published 08.04.19.

Please cite as:
Morath B, Wien K, Hoppe-Tichy T, Haefeli WE, Seidling HM
Structure and Content of Drug Monitoring Advices Included in Discharge Letters at Interfaces of Care: Exploratory Analysis Preceding Database Development
JMIR Med Inform 2019;7(2):e10832
URL: https://medinform.jmir.org/2019/2/e10832/
doi:10.2196/10832
PMID:30958278

Original Paper

A Computer Application to Predict Adverse Events in the Short-Term Evolution of Patients With Exacerbation of Chronic Obstructive Pulmonary Disease

Inmaculada Arostegui[1,2,3], MSc, PhD; María José Legarreta[1,4], MSc; Irantzu Barrio[1,3], MSc, PhD; Cristobal Esteban[3,5], MD, PhD; Susana Garcia-Gutierrez[3,4], MD, PhD; Urko Aguirre[3,4], MSc; José María Quintana[3,4], MD, PhD; IRYSS-COPD Group[3]

[1]Departamento de Matemática Aplicada y Estadística e Investigación Operativa, The University of the Basque Country UPV/EHU, Leioa, Spain

[2]Research Institute, Basque Center for Applied Mathematics, Bilbao, Spain

[3]Red de Investigación en Servicios Sanitarios en Enfermedades Crónicas, Galdakao, Spain

[4]Unidad de Epidemiología Clínica, Hospital Galdakao, Galdakao, Spain

[5]Servicio de Neumología, Hospital Galdakao, Galdakao, Spain

Corresponding Author:
Inmaculada Arostegui, MSc, PhD
Departamento de Matemática Aplicada y Estadística e Investigación Operativa
The University of the Basque Country UPV/EHU
Barrio Sarriena S/N
Leioa, 48940
Spain
Phone: 34 946015346
Fax: 34 946013500
Email: inmaculada.arostegui@ehu.eus

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is a common chronic disease. Exacerbations of COPD (eCOPD) contribute to the worsening of the disease and the patient's evolution. There are some clinical prediction rules that may help to stratify patients with eCOPD by their risk of poor evolution or adverse events. The translation of these clinical prediction rules into computer applications would allow their implementation in clinical practice.

Objective: The goal of this study was to create a computer application to predict various outcomes related to adverse events of short-term evolution in eCOPD patients attending an emergency department (ED) based on valid and reliable clinical prediction rules.

Methods: A computer application, Prediction of Evolution of patients with eCOPD (PrEveCOPD), was created to predict 2 outcomes related to adverse events: (1) mortality during hospital admission or within a week after an ED visit and (2) admission to an intensive care unit (ICU) or an intermediate respiratory care unit (IRCU) during the eCOPD episode. The algorithms included in the computer tool were based on clinical prediction rules previously developed and validated within the Investigación en Resultados y Servicios de Salud COPD study. The app was developed for Windows and Android systems, using Visual Studio 2008 and Eclipse, respectively.

Results: The PrEveCOPD computer application implements the prediction models previously developed and validated for 2 relevant adverse events in the short-term evolution of patients with eCOPD. The application runs under Windows and Android systems and it can be used locally or remotely as a Web application. Full description of the clinical prediction rules as well as the original references is included on the screen. Input of the predictive variables is controlled for out-of-range and missing values. Language can be switched between English and Spanish. The application is available for downloading and installing on a computer, as a mobile app, or to be used remotely via internet.

Conclusions: The PrEveCOPD app shows how clinical prediction rules can be summarized into simple and easy to use tools, which allow for the estimation of the risk of short-term mortality and ICU or IRCU admission for patients with eCOPD. The app can be used on any computer device, including mobile phones or tablets, and it can guide the clinicians to a valid stratification of patients attending the ED with eCOPD.

Trial Registration: ClinicalTrials.gov NCT00102401; https://clinicaltrials.gov/ct2/show/results/NCT02434536 (Archived by WebCite at http://www.webcitation.org/76iwTxYuA)

International Registered Report Identifier (IRRID): RR2-10.1186/1472-6963-11-322

(JMIR Med Inform 2019;7(2):e10773) doi:10.2196/10773

KEYWORDS

COPD; disease exacerbation; mortality; intensive care; clinical prediction rule; mobile app

Introduction

Chronic obstructive pulmonary disease (COPD) is one of the most common chronic diseases, and its prevalence is expected to increase over the next few decades [1]. COPD is a leading cause of death in developed countries, and patients with COPD generally suffer a substantial deterioration in their quality of life [2]. COPD is a complex and heterogeneous condition with different clinical manifestations and variable disease activity. There is a continuing interest in using clinical and pulmonary function variables and other disease indicators that may help predict outcomes [3].

The exacerbation of COPD (eCOPD) is defined as an event in the natural course of a patient's COPD characterized by a change in baseline dyspnea, cough, or sputum, that is beyond normal day-to-day variations and that may have warranted a change in medication or treatment [4]. Exacerbations are common among patients with COPD [5]. These sudden worsenings of COPD contribute to disease progression, reduce quality of life, increase the risk of death, and account for substantial use of health care resources [2,6,7]. Currently, emergency department (ED) physicians must rely largely on their experience and the patient's personal criteria to gauge how an eCOPD will evolve. Clinical prediction rules that could help predict eCOPD evolution would allow ED physicians to make better-informed decisions about treatment [8].

Prediction models are gaining importance as a support for decision-making processes. Decisions such as the most appropriate treatment for a disease; whether or not a given patient should be discharged; or the development of effective, acceptable, and cost-efficient prevention strategies are based on the individual patient's risk of suffering some undesirable event. Clinical prediction models provide estimates for an individual's risk of an adverse event over a specific period on the basis of a combination of a number of patient characteristics, which we call variables. Often, clinical prediction models are extended to include clinical prediction rules, risk scores, or prognostic models. The literature includes well-known prediction models, which have been developed to predict the development of a disease, death, or poor evolution caused by a current disease, including eCOPD. More precisely, the Investigación en Resultados y Servicios de Salud COPD (IRYSS-COPD) Appropriateness Study group has developed clinical prediction rules for short-term outcomes in eCOPD patients attending an ED. These outcomes include (1) mortality during hospital admission or within a week after the ED visit [9] and (2) admission to an intensive care unit (ICU) or an intermediate respiratory care unit (IRCU) during the eCOPD episode [10].

Nowadays, clinicians and patients are both actively involved in deciding therapeutic interventions or choosing medical treatments in a shared decision-making process. It is well known that the estimation of an individual's risks of various adverse events by means of prediction models may provide the necessary input for shared decision-making [11]. Therefore, the application of clinical prediction rules in daily clinical practice is one more step in this process. The translation of clinical prediction rules into easy-to-use computer tools would allow the use of these models in clinical practice. The goal of this work was to create a computer application to predict various outcomes related to adverse events of short-term evolution in eCOPD patients attending an ED based on valid and reliable clinical prediction rules. We present the Prediction of Evolution of patients with eCOPD (PrEveCOPD) tool for prediction of 2 outcomes: (1) mortality during hospital admission or within a week after the ED visit and (2) admission to an ICU or IRCU during the eCOPD episode. The algorithms included in the computer tool are based on the clinical prediction rules previously published by Quintana et al [9,10] for the IRYSS-COPD study.

The rest of the paper is organized as follows. The Methods section presents a brief description of the IRYSS-COPD study, including the development of the predictive models and clinical rules, and provides the methodology used to create the PrEveCOPD computer tool for different environments. The Results section describes the PrEveCOPD tool and shows how it runs with individual cases. Finally, the paper closes with a discussion in which the novelty and usefulness of the application, some limitations, and future work are reviewed and conclusions are drawn.

Methods

The Investigación en Resultados y Servicios de Salud-Chronic Obstructive Pulmonary Disease Study: Description and Outcome Prediction Rules

A detailed description of the IRYSS-COPD study has been reported in depth in the study protocol [12]. In brief, this prospective cohort study included subjects with an eCOPD attending the ED of 16 hospitals in Spain between June 2008 and September 2010. The study was approved by the institutional review boards of the participating hospitals, in accordance with all applicable regulations. All patients were informed of the goals of the study and invited to voluntarily participate in it; confidentiality was guaranteed. All who agreed to participate provided written consent.

Data from several time points were collected in the study. However, for the purpose of this study, we concentrated on variables collected at 2 time points. First, data were collected

when the decision was made to hospitalize the patient or discharge him or her home. If the patient was hospitalized, then additional data were collected in the medical ward up to 1 week. Otherwise, if the patient was discharged, he or she was contacted by phone, and similar information was recorded up to 1 week after the index ED visit. The selected predictive variables were previously described [12]. The selected 2 outcome variables were also previously described when the predictive models were developed [9,10]. However, because of the importance of the 2 outcomes for the purpose of this study, we present a brief definition of them. The 2 outcome variables were as follows:

- Death, if it occurred during the hospital admission or within 7 days of the index ED visit among patients discharged to home.
- ICU or IRCU admission: The patient needs an ICU admission or invasive mechanical ventilation (IMV) or suffers a cardiac arrest; or the patient needs a noninvasive mechanical ventilation (NIMV) for 2 or more days, when mechanical ventilation was not used at home before admission or needs an admission to an IRCU for 2 or more days. A minimum of 2 days was chosen to include only those patients needing more intensive and prolonged therapeutic interventions.

This description is restricted to the variables finally considered for the development of the 2 prediction rules. Table 1 shows the distribution of the selected predictive variables by outcome.

The 2 clinical prediction rules were developed following similar methodological approaches. Detailed description is provided elsewhere [9,10], although a brief summary is given below.

Univariate logistic regression analysis was initially performed, and variables with statistically significant results at $P<.20$ were posteriorly entered into a multiple logistic regression model. Internal validation of the variable selection process and modeling was performed until the final predictive model was reached. A score was developed by assigning a weight to each variable or category in the final multiple logistic regression model, as suggested in the literature [13]. Finally, the score was categorized into a manageable number of risk classes based mainly on the estimated risk of event for each outcome.

Discrimination of the score and the risk categories was assessed by the area under the receiver operating characteristic curve. All the modeling, scoring, and categorization processes were validated by split-sample validation (50% development and 50% validation). Figure 1 shows the whole process of score development and categorization and the resulting risk categories for the 2 outcomes, death and ICU or IRCU admission. The Cochran-Armitage trending statistic was performed to assess whether classification provided by the score could differentiate low-risk patients from high-risk patients in a fashion of graded response based on the level of risk present.

Results of the developed risk categories and association with the 2 outcomes are shown in Table 2. Note that because of missing values in predictor or response variables, the total number of subjects for which the 2 risk scores were estimated differed. Detailed information regarding missing values can be obtained in the original papers where these scores were developed.

The Computer Application: Prediction of Evolution of Patients With Exacerbation of Chronic Obstructive Pulmonary Disease

The PrEveCOPD computer application has been implemented to be installed both in Windows and Android systems and can also be used on the Web without installing any application.

The application for Windows and Web platforms has been developed using Microsoft Visual Studio 2008 [14], and a tool called Eclipse [15] was used to develop the instrument to be run on an Android system.

For the Windows application, we used C# programming language [16] to develop the application and then install and run locally on the user's computer. The Web application was also created in C# and implemented on a computer workstation so that users could access it remotely. The application is available for downloading and installing on the computer or to be used remotely as a Web application [17]. Therefore, anyone with an internet connection and browser could access the website and run the application. The application operates exactly in the same way when the access is local and remote. The performance of the Windows application has been checked under Windows 7, in a 32-bits personal computer. The most common browsers (Internet Explorer, Firefox, Chrome, and Safari), with updated plugins installed for Java version 8 or posterior, have been tested for the Web application. For the Android app, we used the Java programming language to develop the app in the Eclipse development environment and install it and run it on any device with an Android operating system. The Android app is available on Google Play under the medicine category, with the name PrEveCOPD. The performance of the app has been tested under Android version 7.0 or posterior.

The minimum equipment requirements that we recommend to run the application are Windows 7 with 32 bits and 4 GB of RAM memory for local access under Windows, Java version 8, and one of the following browsers: Internet Explorer 11, Firefox 59, or Chrome 69 for remote access or Android Nougat 7.0 release for the Android app.

Figures 2 and 3 show a screenshot of the Android app running on a mobile phone (Figure 2) and the tool under Windows (Figure 3). The computer application has been developed in English and Spanish. For electronic devices running under Android, the language is automatically detected depending upon the default settings, with English being the default option for any language other than Spanish. For a computer running under Windows, the application has an option to switch between the 2 languages, with Spanish being the default option.

Table 1. Distribution of the predictive variables by outcome. The 2 outcomes are mortality during hospital admission or within a week after the emergency department visit and admission to an intensive care unit or intermediate respiratory care unit during the exacerbation of chronic obstructive pulmonary disease episode (N=2487).

Predictive variable	Sample, n (%)	Mortality		Admission to intensive care unit or intermediate respiratory care unit	
		n (%)	*P* value[a]	n (%)	*P* value[a]
Total		59 (2.37)	—[b]	258 (10.37)	—
Age (years)					
<75	1271 (51.11)	18 (1.42)	<.001	166 (13.06)	<.001
75-85	1050 (42.22)	28 (2.67)	<.001	80 (7.62)	<.001
>85	165 (6.63)	13 (7.88)	<.001	12 (7.27)	<.001
Previous long-term home oxygen therapy or noninvasive mechanical ventilation					
Yes	841(33.82)	43 (5.11)	<.001	190 (22.59)	<.001
No	1646 (66.18)	16 (0.97)	<.001	68 (4.13)	<.001
Altered consciousness					
Yes	70 (2.81)	9 (12.86)	<.001	36 (51.43)	<.001
No	2415 (97.10)	49 (2.03)	<.001	220 (9.11)	<.001
Use of inspiratory accessory muscle					
Yes	535 (21.51)	34 (6.36)	<.001	111 (20.75)	<.001
No	1952 (78.49)	25 (1.28)	<.001	147 (7.53)	<.001
Dyspnea (Medical Research Council)					
Missing	250 (10.05)	18 (7.20)	<.001	21 (8.40)	<.001
Grade 1	188 (7.56)	0 (0)	<.001	10 (5.32)	<.001
Grade 2	600 (24.13)	1 (0.17)	<.001	44 (7.33)	<.001
Grade 3	501 (20.14)	7 (1.40)	<.001	45 (8.98)	<.001
Grade 4	672 (27.02)	10 (1.49)	<.001	81(12.05)	<.001
Grade 5	276 (11.10)	23 (8.33)	<.001	57 (20.65)	<.001
pH					
≥7.35	1991 (86.75)	40 (2.01)	.02	121 (6.08)	<.001
7.26-7.35	250 (10.89)	11 (4.40)	.02	87 (34.80)	<.001
<7.26	54 (2.35)	3 (5.56)	.02	38 (70.37)	<.001
Pressure of carbon dioxide (P_{CO2})					
≤45	1232 (57.20)	16 (1.30)	<.001	32 (2.60)	<.001
45-55	484 (22.47)	14 (2.89)	<.001	47 (9.71)	<.001
55-65	241 (11.19)	10 (4.15)	<.001	58 (24.07)	<.001
>65	197 (9.15)	13 (6.60)	<.001	107 (54.31)	<.001

[a]Chi-square test for homogeneity.

[b]Not applicable.

The main screen incorporates a help button, where the specific definition of all the predictive variables is detailed exactly the same as in the manuscripts where prediction rules were developed [9,10]. The computer tool also incorporates a predefined range of acceptable values for each variable to control for typing mistakes or out-of-range values. An error message prevents invalid values to be introduced, with strict instructions about the accepted range of values. The application accepts a missing value in any of the predictive variables, leading in that case to a lower bound for the corresponding score.

A button with information for users about the legal responsibility derived from the use of the application is also incorporated in the Android platform, and this information is displayed on the main screen in the Windows and Web platforms.

Figure 1. Summary of the process for the 2 outcomes (death and intensive care unit or intermediate respiratory care unit admission): score development and stratification into risk categories. ED: emergency department; ICU: intensive care unit; IRCU: intermediate respiratory care unit; LTHOT: long-term home oxygen therapy; MRC: Medical Research Council; NIMV: noninvasive mechanical ventilation; P_{CO2}: pressure of carbon dioxide.

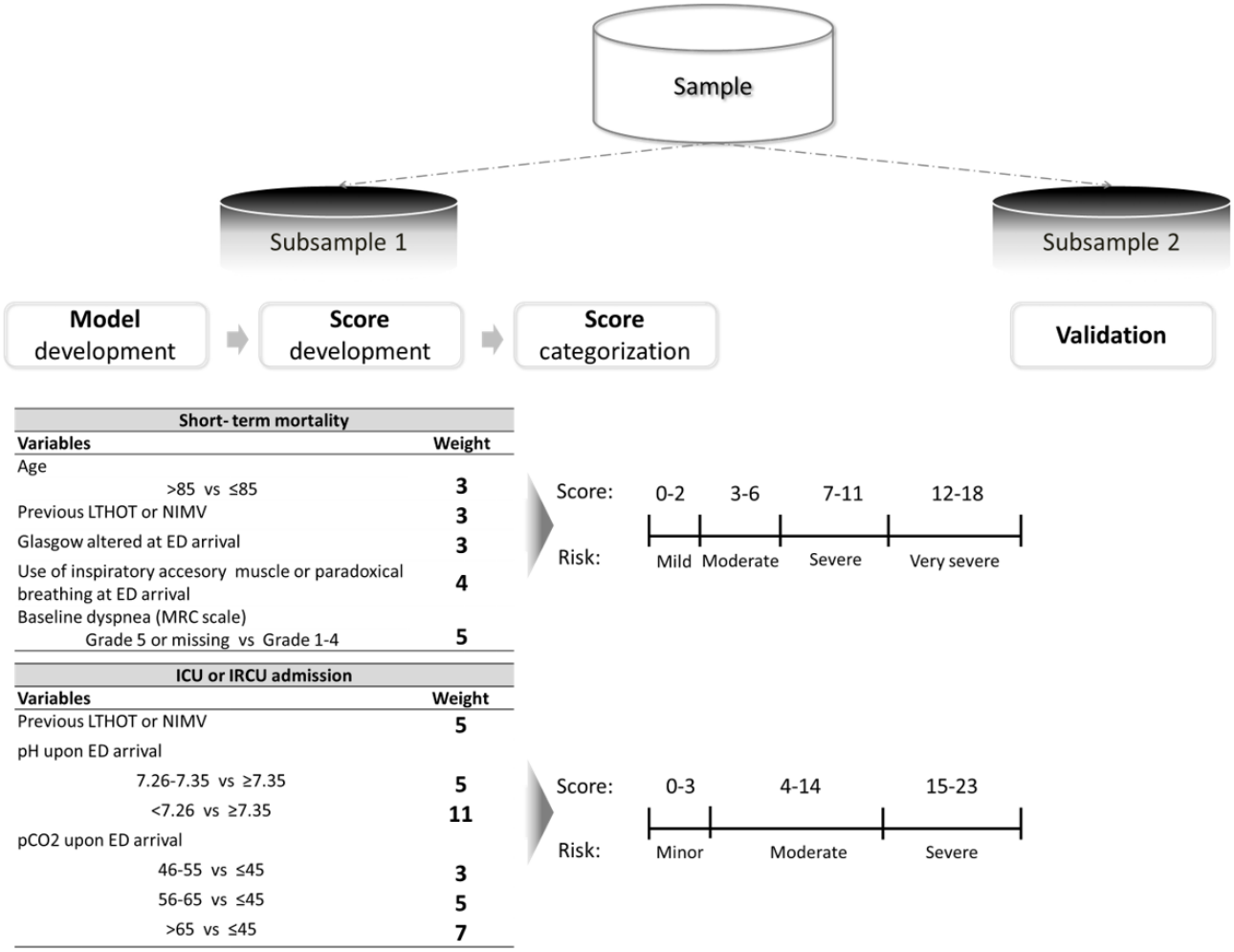


Table 2. Distribution of the developed risk categories for each of the outcomes.

Outcome	Yes, n (%)	No, n (%)	*P* value[a]
Short-term mortality risk			
Mild (n=1081)	3 (0.28)	1078 (99.72)	<.001
Moderate (n=865)	11 (1.27)	854 (98.73)	<.001
Severe (n=441)	20 (4.54)	421 (95.46)	<.001
Very severe (n=97)	24 (24.74)	73 (75.26)	<.001
Intensive care unit or intermediate respiratory care unit admission risk			
Minor (n=1203)	12 (1.00)	1191 (99.00)	<.001
Moderate (n=803)	144 (17.93)	659 (82.07)	<.001
Severe (n=148)	88 (59.46)	60 (40.54)	<.001

[a]Cochran-Armitage trend-test.

Figure 2. Screenshot of the application running under the Android platform. Data for an imaginary subject with complete information displayed as an example. ED: emergency department; ICU: intensive care unit; IRCU: intermediate respiratory care unit; LTHOT: long-term home oxygen therapy; MRC: Medical Research Council; NIMV: noninvasive mechanical ventilation; P_{CO2}: pressure of carbon dioxide.

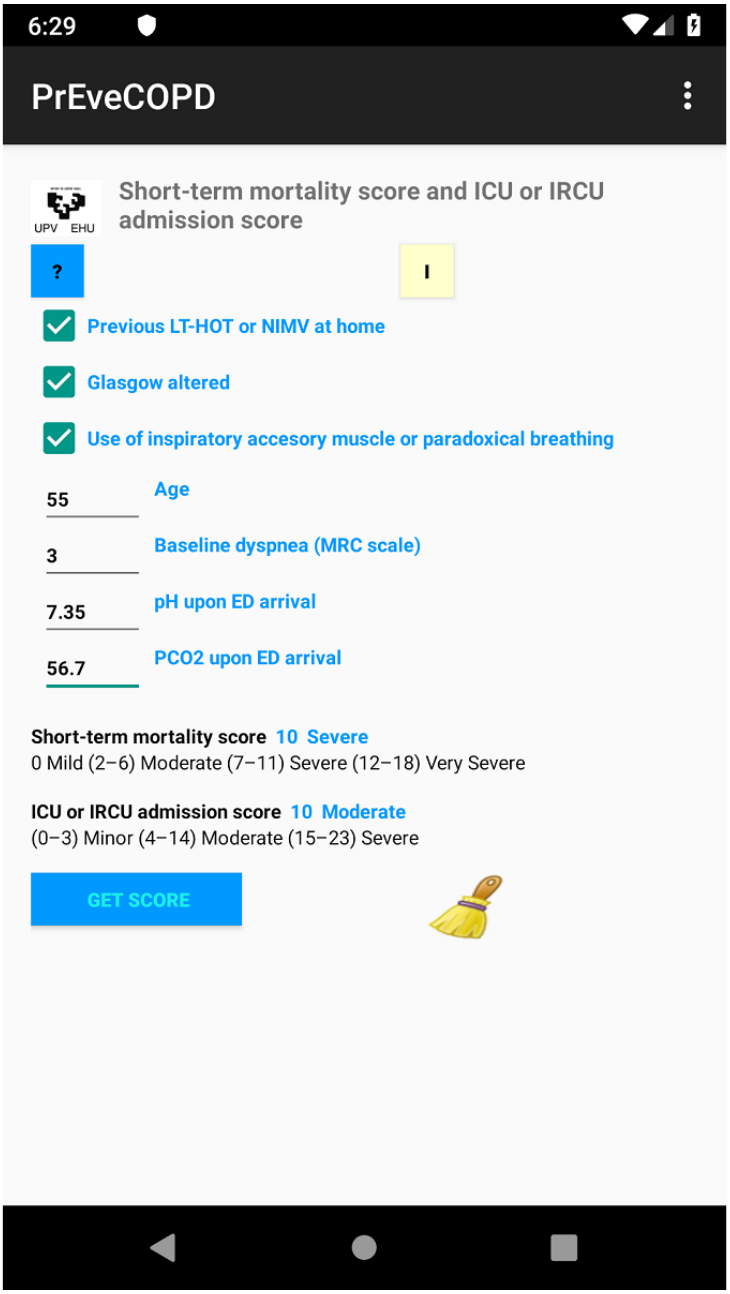
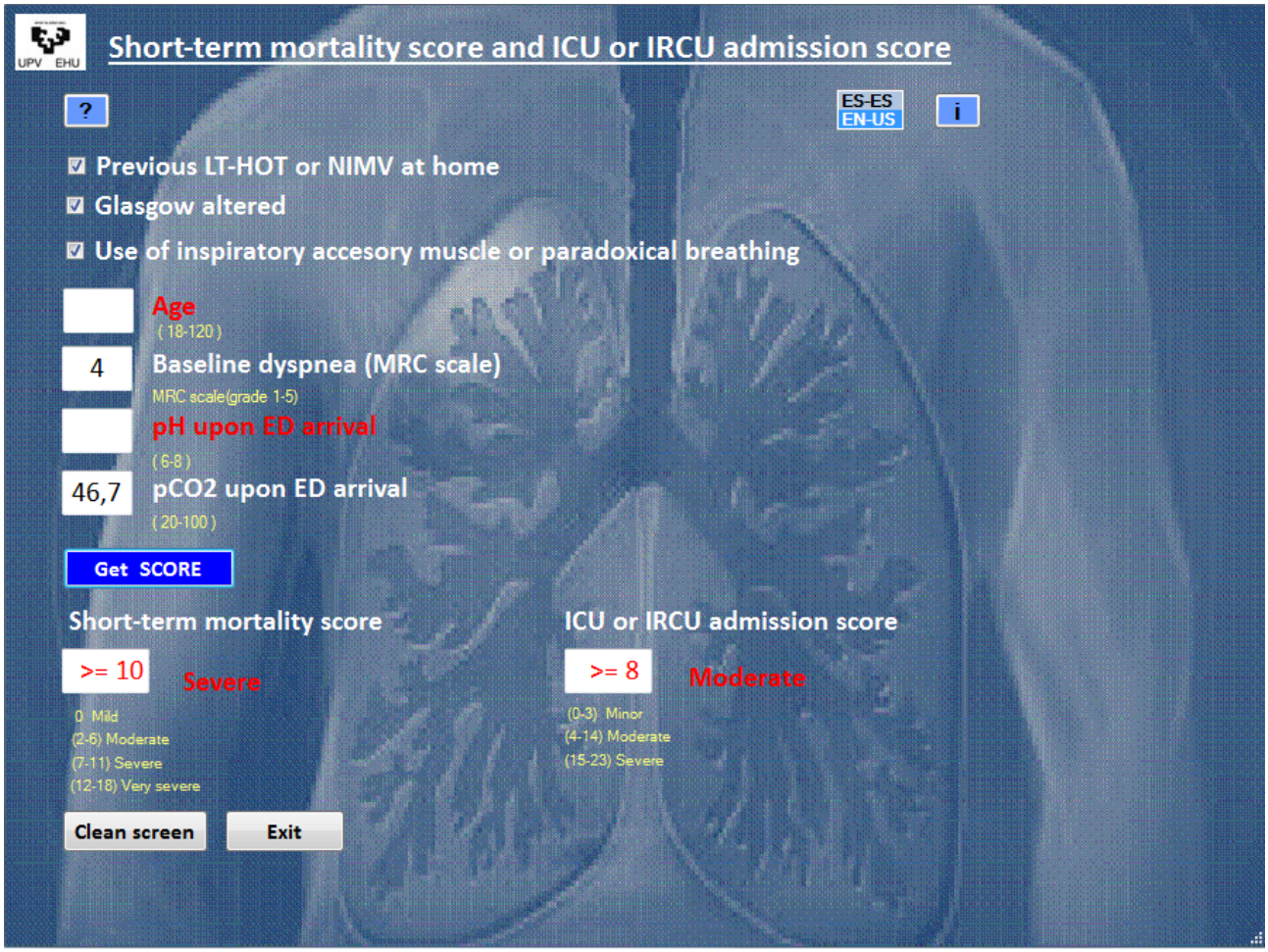


Figure 3. Screenshot of the application running under Windows and Web platforms. Data for an imaginary subject with incomplete information displayed as an example. ED: emergency department; ICU: intensive care unit; IRCU: intermediate respiratory care unit; LTHOT: long-term home oxygen therapy; MRC: Medical Research Council; NIMV: noninvasive mechanical ventilation; P_{CO2}: pressure of carbon dioxide.



Results

The final product is an application with a user-friendly interface that comprises a screen where the values of the specific predictive variables are introduced. Then, by pushing the *Get SCORE* button, the estimated score for the 2 outcomes, death and ICU or IRCU admission, are automatically shown. The screen also shows the stratification of risk into categories for both scores.

Furthermore, 5 parameters defined the final model for predicting death during hospital admission or within 1 week of discharge from the ED to home: age, previous history of long-term home oxygen therapy (LTHOT) or need for NIMV, altered consciousness measured by Glasgow coma scale (GCS), use of accessory inspiratory muscles or paradoxical breathing upon ED arrival, and baseline dyspnea measured by the Medical Research Council (MRC) scale. The final predictive model for ICU or IRCU admission was defined by 3 variables. One of them was the same as in the previous model for death, namely, previous history of LTHOT or need for NIMV. The other 2 were elevated P_{CO2} and decreased pH upon ED arrival. Previous history of LTHOT or need for NIMV, altered consciousness measured by GCS, and the use of accessory inspiratory muscles or paradoxical breathing upon ED arrival are tick variables. It means that by default they were stated as *No*, whereas selecting them with a tick changes their state to *Yes*. Age and baseline dyspnea (Grade 1-5) must be introduced in the integer format. P_{CO2} and pH are numerical values formatted with 1 and 2 decimal digits, respectively. The application does not allow data outside the established range or erroneous data entry, as stated on the help screen. If values for any of the variables included in the application are missing, the names of these variables as well as the estimated score and the risk category will appear in red. Moreover, it is indicated that the real value will be greater than or equal to the value on screen.

For instance, Figure 2 shows how data on a 55-year-old patient who arrives at ED with eCOPD, pH=7.35, P_{CO2}=56.7, level of dyspnea-MRC=3, previous history of LTHOT or need for NIMV, use of accessory inspiratory muscles, and altered consciousness measured by GCS were introduced in the app running under Android. For a patient with these specific characteristics, the application estimates a value of 10 for the score that measures the risk of death during the first 7 days, which means a severe risk of death. The same patient, or another one with these characteristics, has an estimated value of 10 for the score that measures the risk of admission to ICU or IRCU, which is translated to a moderate risk of admission to ICU or IRCU.

Figure 3 shows how data on a patient arriving at ED with eCOPD who has a previous history of LTHOT or need for NIMV, use of accessory inspiratory muscles, altered consciousness measured by GCS, P_{CO2}=46.7, level of dyspnea-MRC=4, and missing values for age and pH were introduced in the application running under Windows. For a patient with these specific characteristics, the application estimates a value greater than or equal to 10 for the score that measures the risk of death during the first 7 days, which means a severe or very severe risk of death. The same patient, or another one with these characteristics, has an estimated value of 8 or higher for the score that measures the risk of admission to ICU or IRCU, which is translated to a moderate or severe risk of admission to ICU or IRCU.

Discussion

Principal Findings

We have developed a computer application that implements the prediction models previously developed for 2 relevant adverse events in the short-term evolution of patients with eCOPD. The 2 adverse events selected as outcomes were mortality during hospital admission or within a week after the ED visit and admission to an ICU or IRCU during the eCOPD episode. The main strength of the app is that it is based on clinical predictive rules derived from models previously developed and validated for both outcomes.

The short-term evolution of patients with eCOPD is a critical issue regarding the health care provided at the EDs. Decision on medication, treatment, or hospitalization could be extremely benefited by any reliable information of the estimated risk of adverse evolution. Previous studies showed that relevant events in terms of bad evolution during the initial days would be death, ICU admission, need for IMV, cardiac arrest, need for NIMV if mechanical ventilation was not used at home before, or admission to an IRCU for some days [18,19]. Although some of the adverse events are obviously more severe than others, there is no continuum on the severity of all of them. Therefore, measuring the risk of any such events at the same time and through the same instrument could have a potential benefit over individual tools or crude predictive models.

As stated in the literature, the development and validation of prediction models require strict methodological norms [11]. When prediction models are developed, it may be necessary to make several assumptions regarding the structure of the data or the relation between covariates. If the aim is to apply the prediction model in practice, it is important to show that it is valuable when applied to new data, which is called validation. Internal validation evaluates the validity of the model when it is applied to data derived from the same sample in which they have been developed. Conversely, external validation examines the generalizability of the model to other samples. Usually, there are no data or funding available to do external validation. Hence, when a prediction model is developed, a good internal validation should be ensured at the least. The 2 logistic models we have selected to develop the app have been developed following proper procedures for derivation and validation, and they provide very good predictive validity. In addition, both models were derived from a large multicenter prospective cohort, and they use clinical data generally available in the ED and also at the primary care level.

Nowadays, the transference from clinical research to clinical practice is a relevant issue. The development of a clinical prediction rule goes one step further than predictive modeling. The development of a model does not mean that results predicted by the model would be used in daily clinical practice. Moreover, the success of a well-validated prediction model in practice will depend on 2 factors: its transfer to a reliable clinical rule and its availability in an easy-to-use tool. Implementation of a validated model into a user-friendly tool is a key step in developing risk models, which can increase the uptake of the model [20]. Thecalculator.co provides all kinds of free Web tools such as calculators, where one of the areas of interest is devoted to health [21]. Specifically for COPD, the website offers calculators for the well-known BODE Index (based on the body-mass index (B), the degree of airflow obstruction (O) and dyspnea (D), and exercise capacity (E), measured by the six-minute–walk test) [22] and for COPD stages classification by the Global Initiative for Chronic Obstructive Lung Disease GOLD guidelines [3]. Nevertheless, these tools are not all based on prediction models or clinical prediction rules.

Other studies have developed prediction models of evolution for patients with eCOPD [18,19] or have validated existing prediction models for other respiratory diseases [23,24]. Some of them have been translated into clinical prediction rules or scores for predicting short-term outcomes or stratifying patients based on their probability of adverse evolution [18,24,25]. However, as far as we know, none of them have been incorporated on an available and easy-to-use computer application that only needs to be downloaded to a computer device, such as a tablet or mobile phone, to be used. The implementation of a theoretical model into an easy-to-use application would allow its rapid and easy incorporation to the clinical management of eCOPD patients at the ED to guide their treatment. Nowadays, information systems are created differently across regions and countries. For the moment, we have stored our tool on a server so that it can be used in any health system in the world. As technology advances in each health system, our instrument could serve as the basis to automatically include information relative to the individual patient at the bed-side where decisions should be made. We are aware that until these tools are able to use information from electronic health record directly, emergency physicians will have to duplicate introduction of data, and this fact is a limitation for the generalization of the use of prediction models in clinical practice. We recommend lead efforts in this direction. Strictly, the use of these models in practice will allow us to properly validate them and, if necessary, update them.

Regarding other clinical fields, we have found some prediction rules that have been integrated into computer applications [26-28]. For instance, in the context of the Framingham Heart Study, several risk prediction models have been developed [29,30]. These risk scores are available either as an interactive calculator or a spreadsheet [26]. Another example is showed by Moreno-Cid et al, who performed a systematic review of the clinical prediction rules for the risk of Down syndrome based

on ultrasound findings in pregnancy [31]. These authors showed that only 3 of the rules were validated (2 internally and 1 externally) and 4 of them were incorporated into a software application [32-35]. Moreover, a recent systematic review evaluated Web-based cardiovascular disease risk calculators in terms of clinical validity, understandability, and actionability [36]. The authors concluded that although the number of available Web-based tools is high, developers need to address actionability as well as clinical validity and understandability to improve usefulness. We believe that with regard to the prediction of evolution in the context of eCOPD, our software application verifies the 3 conditions highlighted by the authors, namely, validity, understandability, and actionability.

Limitations and Future Work

This study inherits the limitations derived from the development of the 2 clinical prediction rules that have been translated into the application. These limitations were missing data for some key variables and the absence of biomarkers. These limitations were already previously cited and discussed in the original papers [9,10]. However, we would like to incorporate some discussion related to a third limitation, which was the lack of external validation of the developed predictive models. Authors of the clinical prediction rules for adverse events in the short-term evolution of patients with eCOPD asseverate that proper validation in future studies should further demonstrate their value in clinical practice. The use of the computer application that we present could easily allow for the storing of new data on patients attending to an ED with eCOPD, which could be posteriorly used to externally validate the original models and prediction rules in different populations. This easy-to-get bank of data would also allow for the description of types and profiles of patients attending an ED with an eCOPD. We should mention a new limitation, restricted to the app and not to the prediction rules, which is the fact that the selected outcomes were predefined. The application in its actual form does not allow for prediction of other outcomes apart from the ones included in the original prediction rules and clearly stated before. The prediction of any different outcome would require a previous development and validation of a new prediction rule and posterior incorporation into the app. Finally, we have reported the characteristics of the computer, the operating system, and the software versions under which the app has been developed and tested. We are not able to guarantee the correct performance of the application under different conditions.

Conclusions

The proposed computer application shows how clinical prediction rules derived from multiple logistic regression models can be summarized into simple and easy-to-use tools that allow the estimation of the risk of short-term mortality and ICU or IRCU admission for patients with eCOPD. The app can be used in any computer device, including mobile phone or tablets, and it can guide the clinicians to a valid stratification of patients attending the ED with eCOPD.

Acknowledgments

This work was financially supported in part by grants from the Fondo de Investigación Sanitaria (PI 06/1010, PI06/1017, PI06/714, PI06/0326, PI06/0664); Departamento de Salud del Gobierno Vasco (2012111008); Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT620-13); Ministerio de Economía y Competitividad del Gobierno Español and FEDER (BCAM Severo Ochoa excellence accreditation SEV-2017-0718 and MTM2016-74931-P); Basque Government through the BERC 2018-2021; the Research Committee of the Hospital Galdakao; the thematic networks—REDISSEC (Red de Investigación en Servicios de Salud en Enfermedades Crónicas)—of the Instituto de Salud Carlos III; and the University of the Basque Country (PES18/26). There was no other source of financial support or sponsor. The authors also wish to thank the anonymous referees for providing comments that led to substantial improvement of the paper.

The IRYSS-COPD group included the following co-investigators: Dr Jesús Martínez-Tapias (Hospital Virgen de las Nieves, Granada); Alba Ruiz (Hospital de Motril, Granada); Dr Eduardo Briones (Unidad de Epidemiología, Distrito Sanitario Sevilla); Dr Silvia Vidal (Unidad de Investigación, Hospital Costa del Sol, Marbella); Dr Emilio Perea-Milla and Francisco Rivas (Servicio de Epidemiología, Hospital Costa del Sol, Málaga-REDISSEC); Dr Maximino Redondo (Servicio de Laboratorio, Hospital Costa del Sol, Málaga-REDISSEC); Javier Rodríguez Ruiz (Responsable de Enfermería del Área de Urgencias, Hospital Costa del Sol, Málaga); Dr Marisa Baré (Epidemiología y Evaluación, Corporació Sanitaria Parc Taulí-CSPT, Sabadell REDISSEC); Dr Manel Lujan and Dr Concepción Montón (Servicio de Neumología, CSPT/REDISSEC); Dr Amalia Moreno, Dr Josune Ormaza, and Dr Javier Pomares (Servicio de Neumología, CSPT); Dr Juli Font (Medicina, Servicio de Urgencias; CSPT); Dr Cristina Estirado and Dr Joaquín Gea (Servicio de Neumología, Hospital del Mar/CIBERES, Barcelona); Dr Elena Andradas (subdirectora de Promoción de la Salud y Epidemiología del Ministerio de Sanidad, Servicios Sociales e Igualdad); Dr Juan Antonio Blasco (Unidad de Evaluación de Tecnologías Sanitarias, Agencia Laín Entralgo, Madrid); Dr Nerea Fernández de Larrea (Subdirección General de Tecnología e Innovación Sanitarias, Consejería de Sanidad de la Comunidad de Madrid/REDISSEC); Rosa Girón (Hospital de La Princesa, Madrid); María del Puerto Cano Aguirre (Hospital de Torrejón, Madrid); Dr Jose Luis Lobo (Servicio de Neumología, Hospital Txagorritxu, Araba); Dr Esther Pulido (Servicio de Urgencias, Hospital Galdakao-Usansolo, Bizkaia); Dr Mikel Sánchez (Servicio de Urgencias, Hospital Galdakao-Usansolo); Dr Luis Alberto Ruiz (Servicio de Respiratorio, Hospital de Cruces, Bizkaia); Dr Ane Miren Gastaminza (Hospital San Eloy, Bizkaia); Dr Eva Tabernero (Servicio de Neumología, Hospital de Santa Marina); Carmen Haro (Servicio de Urgencias, Hospital de Santa Marina, Bizkaia); Dr Ramon Agüero (Servicio de Neumología, Hospital Marqués de Valdecilla, Santander); Dr Gabriel Gutiérrez (Servicio de Urgencias, Hospital Cruces, Bizkaia); Dr Belén Elizalde (Dirección Territorial de Gipuzkoa); Dr Felipe Aizpuru (Unidad de Investigación, Hospital Txagorritxu,

Araba/REDISSEC); Dr Inmaculada Arostegui and Dr Irantzu Barrio (Departamento de Matemática Aplicada, Estadística e Investigación Operativa, UPV/EHU-REDISSEC); Amaia Bilbao (Hospital Universitario Basurto/REDISSEC); Dr Cristóbal Esteban (Servicio de Neumología, Hospital Galdakao-Usansolo, Bizkaia/REDISSEC); Dr Nerea González, Susana Garcia, Iratxe Lafuente, and Urko Aguirre; and Miren Orive, Ane Anton, and Dr Jose M Quintana (Unidad de Investigación, Hospital Galdakao-Usansolo, Bizkaia/REDISSEC).

Conflicts of Interest

None declared.

References

1. Buist AS, Vollmer WM, McBurnie MA. Worldwide burden of COPD in high- and low-income countries. Part I. The burden of obstructive lung disease (BOLD) initiative. Int J Tuberc Lung Dis 2008 Jul;12(7):703-708. [Medline: 18544191]
2. Esteban C, Quintana JM, Moraza J, Aburto M, Egurrola M, España PP, et al. Impact of hospitalisations for exacerbations of COPD on health-related quality of life. Respir Med 2009 Aug;103(8):1201-1208 [FREE Full text] [doi: 10.1016/j.rmed.2009.02.002] [Medline: 19272762]
3. Global Initiative for Chronic Obstructive Lung Disease (GOLD). URL: https://goldcopd.org/ [accessed 2019-03-08] [WebCite Cache ID 76ipSZbYh]
4. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2007 Sep 15;176(6):532-555. [doi: 10.1164/rccm.200703-456SO] [Medline: 17507545]
5. Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med 2010 Sep 16;363(12):1128-1138. [doi: 10.1056/NEJMoa0909883] [Medline: 20843247]
6. Anzueto A, Leimer I, Kesten S. Impact of frequency of COPD exacerbations on pulmonary function, health status and clinical outcomes. Int J Chron Obstruct Pulmon Dis 2009;4:245-251 [FREE Full text] [Medline: 19657398]
7. Perera PN, Armstrong EP, Sherrill DL, Skrepnek GH. Acute exacerbations of COPD in the United States: inpatient burden and predictors of costs and mortality. COPD 2012 Apr;9(2):131-141. [doi: 10.3109/15412555.2011.650239] [Medline: 22409371]
8. Roche N, Rabbat A, Zureik M, Huchon G. Chronic obstructive pulmonary disease exacerbations in emergency departments: predictors of outcome. Curr Opin Pulm Med 2010 Mar;16(2):112-117. [doi: 10.1097/MCP.0b013e328335f039] [Medline: 20019614]
9. Quintana JM, Esteban C, Unzurrunzaga A, Garcia-Gutierrez S, Gonzalez N, Barrio I, IRYSS-COPD group. Predictive score for mortality in patients with COPD exacerbations attending hospital emergency departments. BMC Med 2014 Apr 23;12:66 [FREE Full text] [doi: 10.1186/1741-7015-12-66] [Medline: 24758312]
10. Quintana JM, Esteban C, Unzurrunzaga A, Garcia-Gutierrez S, Gonzalez N, Lafuente I, IRYSS-COPD Group. Prognostic severity scores for patients with COPD exacerbations attending emergency departments. Int J Tuberc Lung Dis 2014 Dec;18(12):1415-1420. [doi: 10.5588/ijtld.14.0312] [Medline: 25517805]
11. Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.
12. Quintana JM, Esteban C, Barrio I, Garcia-Gutierrez S, Gonzalez N, Arostegui I, IRYSS-COPD Group. The IRYSS-COPD appropriateness study: objectives, methodology, and description of the prospective cohort. BMC Health Serv Res 2011 Nov 24;11:322 [FREE Full text] [doi: 10.1186/1472-6963-11-322] [Medline: 22115318]
13. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. Stat Med 2004 May 30;23(10):1631-1660. [doi: 10.1002/sim.1742] [Medline: 15122742]
14. Randolph N, Gardner D. Professional Visual Studio 2008. Indianapolis: John Wiley & Sons, Inc; 2008.
15. Daum B. Professional Eclipse 3 For Java Developers. Indianapolis: John Wiley & Sons, Inc; 2004.
16. Lee W. C# 2008 Programmer's Reference. Indianapolis: John Wiley & Sons, Inc; 2008.
17. Legarreta M, Barrio I, Arostegui I, Quintana J, García-Gutierrez S, Esteban C. University of Pais Vasco. PrEveCOPD URL: https://www.ehu.eus/en/web/biostit/prevecopd [accessed 2018-04-13] [WebCite Cache ID 6yeNJu7J7]
18. Roche N, Zureik M, Soussan D, Neukirch F, Perrotin D, Urgence BPCO (COPD Emergency) Scientific Committee. Predictors of outcomes in COPD exacerbation cases presenting to the emergency department. Eur Respir J 2008 Oct;32(4):953-961 [FREE Full text] [doi: 10.1183/09031936.00129507] [Medline: 18508819]
19. Matkovic Z, Huerta A, Soler N, Domingo R, Gabarrús A, Torres A, et al. Predictors of adverse outcome in patients hospitalised for exacerbation of chronic obstructive pulmonary disease. Respiration 2012;84(1):17-26 [FREE Full text] [doi: 10.1159/000335467] [Medline: 22327370]
20. Wynants L, Collins GS, Van Valster B. Key steps and common pitfalls in developing and validating risk models. BJOG 2017 Feb;124(3):423-432 [FREE Full text] [doi: 10.1111/1471-0528.14170] [Medline: 27362778]
21. The Calculator. 2014. URL: https://www.thecalculator.co/ [accessed 2018-04-13] [WebCite Cache ID 6yeNsd0BL]

22. Celli BR, Cote CG, Marin JM, Casanova C, Montes DO, Mendez RA, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med 2004 Mar 04;350(10):1005-1012. [doi: 10.1056/NEJMoa021322] [Medline: 14999112]
23. Shorr AF, Sun X, Johannes RS, Yaitanes A, Tabak YP. Validation of a novel risk score for severity of illness in acute exacerbations of COPD. Chest 2011 Nov;140(5):1177-1183. [doi: 10.1378/chest.10-3035] [Medline: 21527510]
24. Chang CL, Sullivan GD, Karalus NC, Mills GD, McLachlan JD, Hancox RJ. Predicting early mortality in acute exacerbation of chronic obstructive pulmonary disease using the CURB65 score. Respirology 2011 Jan;16(1):146-151. [doi: 10.1111/j.1440-1843.2010.01866.x] [Medline: 20920140]
25. Steer J, Gibson J, Bourke SC. The DECAF Score: predicting hospital mortality in exacerbations of chronic obstructive pulmonary disease. Thorax 2012 Nov;67(11):970-976. [doi: 10.1136/thoraxjnl-2012-202103] [Medline: 22895999]
26. Framingham Heart Study. URL: https://www.framinghamheartstudy.org/ [accessed 2018-04-13] [WebCite Cache ID 6yeOdm8rQ]
27. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. Circulation 2006 Mar 21;113(11):1424-1433. [doi: 10.1161/CIRCULATIONAHA.105.584102] [Medline: 16534009]
28. Badawi O, Breslow MJ. Readmissions and death after ICU discharge: development and validation of two predictive models. PLoS One 2012 Nov;7(11):e48758 [FREE Full text] [doi: 10.1371/journal.pone.0048758] [Medline: 23144958]
29. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-753 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.699579] [Medline: 18212285]
30. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. Lancet 2009 Feb 28;373(9665):739-745 [FREE Full text] [doi: 10.1016/S0140-6736(09)60443-8] [Medline: 19249635]
31. Moreno-Cid M, Tenías Burillo JM, Rubio-Lorente A, Rodríguez MJ, Bueno-Pacheco G, Román-Ortiz C, et al. Systematic review of the clinical prediction rules for the calculation of the risk of Down syndrome based on ultrasound findings in the second trimester of pregnancy. Prenat Diagn 2014 Mar;34(3):265-272. [doi: 10.1002/pd.4304] [Medline: 24420977]
32. Nicolaides KH. Screening for chromosomal defects. Ultrasound Obstet Gynecol 2003 Apr;21(4):313-321 [FREE Full text] [doi: 10.1002/uog.128] [Medline: 12704736]
33. Schluter PJ, Pritchard G. Mid trimester sonographic findings for the prediction of Down syndrome in a sonographically screened population. Am J Obstet Gynecol 2005 Jan;192(1):10-16. [doi: 10.1016/j.ajog.2004.08.036] [Medline: 15671995]
34. Vergani P, Ghidini A, Weiner S, Locatelli A, Pozzi E, Biffi A. Risk assessment for Down syndrome with genetic sonogram in women at risk. Prenat Diagn 2008 Dec;28(12):1144-1148. [doi: 10.1002/pd.2138] [Medline: 19006200]
35. Zhong Y, Longman R, Bradshaw R, Odibo AO. The genetic sonogram: comparing the use of likelihood ratios versus logistic regression coefficients for Down syndrome screening. J Ultrasound Med 2011 Apr;30(4):463-469. [Medline: 21460145]
36. Bonner C, Fajardo MA, Hui S, Stubbs R, Trevena L. Clinical validity, understandability, and actionability of online cardiovascular disease risk calculators: systematic review. J Med Internet Res 2018 Feb 01;20(2):e29 [FREE Full text] [doi: 10.2196/jmir.8538] [Medline: 29391344]

Abbreviations

COPD: chronic obstructive pulmonary disease
eCOPD: exacerbation of chronic obstructive pulmonary disease
ED: emergency department
GCS: Glasgow coma scale
ICU: intensive care unit
IMV: invasive mechanical ventilation
IRCU: intermediate respiratory care unit
IRYSS-COPD: Investigación en Resultados y Servicios de Salud–COPD
LTHOT: long-term home oxygen therapy
MRC: Medical Research Council
NIMV: noninvasive mechanical ventilation
P_{CO2}: pressure of carbon dioxide
PrEveCOPD: prediction of evolution of patients with eCOPD

Edited by G Eysenbach; submitted 13.04.18; peer-reviewed by K Langohr, J op den Buijs; comments to author 02.09.18; revised version received 21.01.19; accepted 17.02.19; published 17.04.19.

Please cite as:
Arostegui I, Legarreta MJ, Barrio I, Esteban C, Garcia-Gutierrez S, Aguirre U, Quintana JM, IRYSS-COPD Group
A Computer Application to Predict Adverse Events in the Short-Term Evolution of Patients With Exacerbation of Chronic Obstructive Pulmonary Disease
JMIR Med Inform 2019;7(2):e10773
URL: http://medinform.jmir.org/2019/2/e10773/
doi:10.2196/10773
PMID:30994471

Original Paper

Exploring the Impact of the Prescription Automatic Screening System in Health Care Services: Quasi-Experiment

Yan Li[1*], MSc; Xitong Guo[1*], PhD; Carol Hsu[2*], PhD; Xiaoxiao Liu[1*], MSc; Doug Vogel[1*], PhD

[1]eHealth Research Institute, School of Management, Harbin Institute of Technology, Harbin, China

[2]Management Science and Engineering, Tongji University, Shanghai, China

[*]all authors contributed equally

Corresponding Author:
Xitong Guo, PhD
eHealth Research Institute
School of Management
Harbin Institute of Technology
92 West Dazhi Street
Nangang District
Harbin, 15000
China
Phone: 86 45186414022
Fax: 86 45186414022
Email: xitongguo@hit.edu.cn

Abstract

Background: Hospitals have deployed various types of technologies to alleviate the problem of high medical costs. The cost of pharmaceuticals is one of the main drivers of medical costs. The Prescription Automatic Screening System (PASS) aims to monitor physicians' prescribing behavior, which has the potential to decrease prescription errors and medical treatment costs. However, a substantial number of cases with unsatisfactory results related to the effects of PASS have been noted.

Objective: The objectives of this study were to systematically explore the imperative role of PASS on hospitals' prescription errors and medical treatment costs and examine its contingency factors to clarify the various factors associated with the effective use of PASS.

Methods: To systematically examine the various effects of PASS, we adopted a quasi-experiment methodology by using a 2-year observation dataset from 2 hospitals in China. We then analyzed the data related to physicians' prescriptions both before and after the deployment of PASS and eliminated influences from a variety of perplexing factors by utilizing a control hospital that did not use a PASS system. In total, 754 physicians were included in this experiment comprising 11,054 patients: 400 physicians in the treatment group and 354 physicians in the control group. This study was also preceded by a series of interviews, which were employed to identify moderators. Thereafter, we adopted propensity score matching integrated with difference-in-differences to isolate the effects of PASS.

Results: The effects of PASS on prescription errors and medical treatment costs were all significant (error: 95% CI –0.40 to –0.11, P=.001; costs: 95% CI –0.75 to –0.12, P=.007). Pressure from organizational rules and workload decreased the effect of PASS on prescription errors (95% CI 0.18-0.39; P<.001) and medical treatment costs (95% CI 0.07-0.55; P=.01), respectively. We also suspected that other pressures (eg, clinical title and risk categories of illness) also impaired physicians' attention to alerts from PASS. However, the effects of PASS did not change among physicians with a higher clinical title or when treating diseases demonstrating high risk. This may be attributed to the fact that these physicians will focus more on their patients in these situations, regardless of having access to an intelligent system.

Conclusions: Although implementation of PASS decreases prescription errors and medical treatment costs, workload and organizational rules remain problematic, as they tend to impair the positive effects of auxiliary diagnosis systems on performance. This again highlights the importance of considering both technical and organizational issues to obtain the highest level of effectiveness when deploying information technology in hospitals.

(JMIR Med Inform 2019;7(2):e11663) doi:10.2196/11663

KEYWORDS

prescription drug monitoring programs; hospital information system; quality of health care; medical errors

Introduction

Background

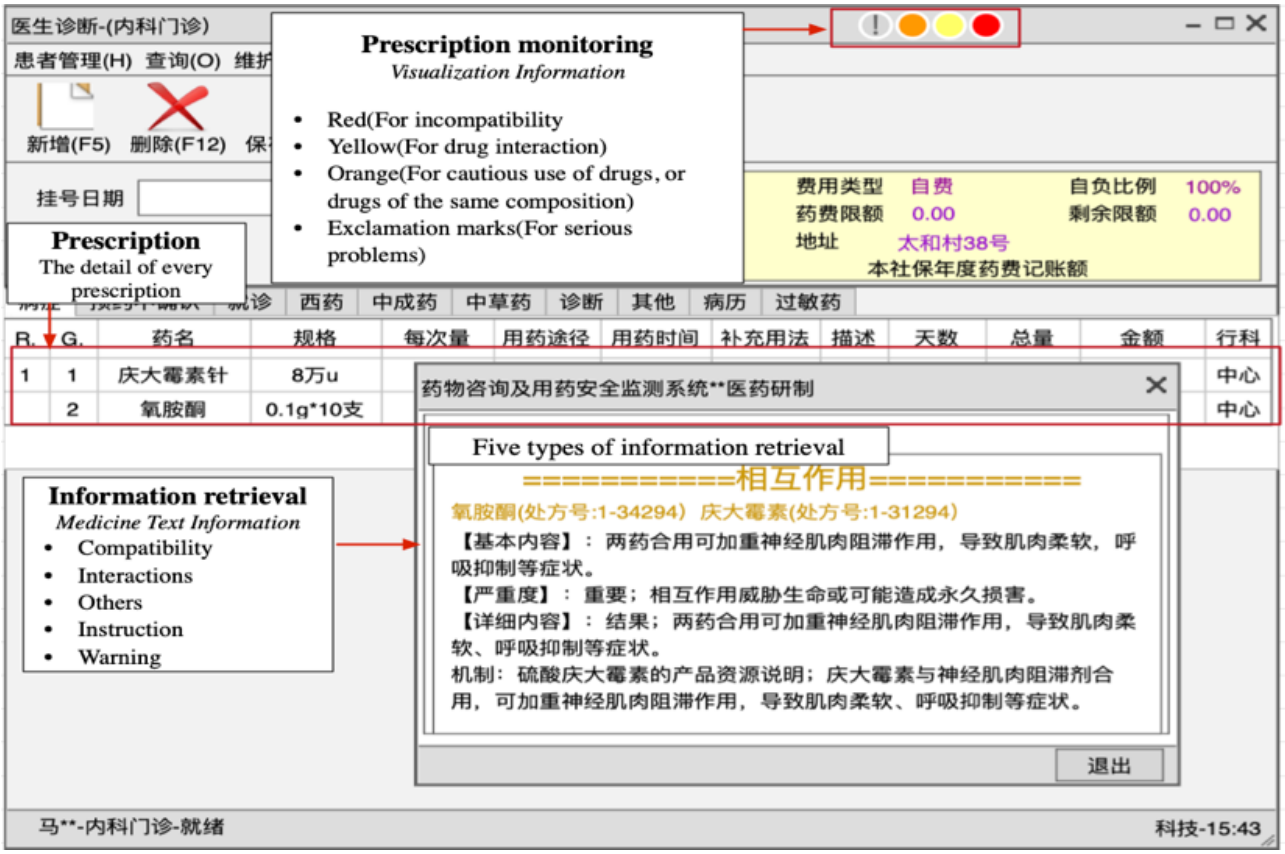
Hospital information systems (HISs) have been adopted as effective tools to mitigate medical treatment costs and improve quality of care [1]. According to the report from Healthcare Information and Management Systems Society analytics, the rates of adoption of hospital intelligence solutions for the US health care market in 2017 increased up to 62% [2] and those for Chinese health care market during 2017-2018 increased up to 55% [3]. However, from the HIS users' perspective, 36.36% of users show their dissatisfaction toward the effects of hospital information technology (HIT). This indicates that despite extensive investments on HISs, the effects of related systems remain somewhat questionable and controversial, in particular, under various environments and for various individuals [4].

Most HISs in China provide informative guidance to the decision makers, which further aim to support their judgments and adoption of the systems [5]. As a significant part of the HIS, the Prescription Automatic Screening System (PASS) focuses on rational drug usage (eg, drug interactions, pharmacologic antagonism, and chemical incompatibility) and searches for available drug information. Past literature posited that PASS could provide high-quality alerts for users [6] and is an effective approach to improve prescribing behavior [7]. Compared with other HISs (eg, electronic medical record [EMR] and HIS), PASS could directly help in the avoidance of adverse drug reactions (ADRs) and adverse drug events (ADEs) and in the reduction of high medical treatment costs caused by inappropriate drug usage [1,8]. Considering the benefits of PASS, many hospitals have already adopted this system. However, the effects of PASS and ways to improve the effectiveness of using PASS are still unclear [9]. Organizations generally develop related rules to enhance the positive effects of PASS, such as a weekly report of dosage; how these rules influence the effects of PASS on medical performance is less explored. Studies have also presented an insignificant correlation between the system in hospitals and prescription behavior [4,9]. Hence, exploring the factors that may influence the effectiveness of PASS implementation, such as individual characteristics and environmental factors, would contribute to further understanding the mechanism of effects of PASS.

Research Context

PASS is an aided diagnosis system that aims to monitor physicians' prescription behavior via reminders to avoid the potential risks and uncertainties inherent in preparing the prescriptions. The primary affordances include *supporting the information retrieval* and *monitoring a prescription. Information retrieval* provides a related knowledge database for physicians to search for information when they are unsure about a situation. PASS provides the detailed usage information of medicines including incompatibility, interactions, drug instructions, warnings, and others. *Prescription monitoring* provides support for physicians to avoid repeated diagnoses, drug interactions, and medication contraindications and to ensure dosage control through the reminders. During the decision process, a knowledge-centered design system promotes more direct interactions between physicians and the system using differently colored alerting lights. These lights represent different risks and facilitate increased communication between physicians and patients, thus helping acquire more patient information to avoid ADRs/ADEs. Specifically, the level of risks existing in prescriptions is presented by these alerting lights, including exclamation marks (for serious problems), red (for incompatibility), yellow (for drug interactions), and orange (for cautious use of drugs, or drugs of the same composition). Such reminders are a form of *in-process* control and serve as the foundation for our subsequent statistical analysis. Figure 1 presents all the above-mentioned information in detail.

Figure 1. Example of the Prescription Automatic Screening System interface.



Literature Review in Hospital Information Systems

Previous literature on HISs has 2 streams, with one focusing on the exploring factors related to HIS adoption [10] and the other focusing on the effects of implementing an HIS in medical institutions. Increasing number of empirical studies have explored the effects of HISs, with studies suggesting that HISs have inconsistent effects on medical performance, such as the medial cost and prescription errors [11,12]. To unfold the black box of the impact of HIS investment, studies further explored various types of HISs to investigate how HISs could improve the medical performance [12-14]. Studies also investigated the mediators between the HIS and performance to explore the underlying mechanism of HIS effects [12]. However, the HIS (eg, EMR) mainly improves the efficiency of physicians' work, which then influences physicians' performance. Limited studies explore the effects of systems such as PASS on physicians' performance by using actual behavior data. Such systems (eg, PASS) have specific functions, such as monitoring and controlling, which differ those of the EMR. Besides, previous studies mainly adopted a descriptive analysis or regression model to explore the relationships in the literature on HIS effects; limited studies deployed a quasi-experiment to explore the effects of an HIS, which is a well-designed methodology to explore the causal relationships [12-14]. Hence, considering these research gaps, this study further explored the contingency factors related to PASS effects by a quasi-experiment in hospitals, which could be generalized to the literature of HIS effects.

Theoretical Foundation

This study was theoretically based on the knowledge-based view [15] and pressure perspective [16]. As tacit technical knowledge of physicians is quite difficult to transfer to another individual [17], coordination among individuals who have expertise in various domains of knowledge becomes a key factor to enhance the quality of performance of hospitals [18]. To improve the coordination efficiency among different groups of individuals within the hospital setting, information technology (IT) provides a technical framework to facilitate the integration of knowledge among various individuals [19]. In this way, IT could not only promote effective knowledge exchange [20] but also provide a pathway for the mapping of knowledge to fulfill the knowledge gaps of a particular individual or group of individuals. Hence, benefiting from the knowledge integration, PASS could provide more information for physicians' decision making, which could improve medical performance.

However, the effectiveness of the effects of knowledge-based view on decision making may vary when individuals face both internal and external pressures. Hence, to fulfill the research gaps of exploring contingency factors between PASS and medical performance, this study built a theoretical model from the pressure perspective and adopted the measurements based on theoretical literature and qualitative results from interviews with caregivers. Regarding the definition of pressure, pressure refers to "those organizational events which cause the individual anxiety, restlessness and irritation" [21]. Hence, in line with this theoretical logic, individual pressure refers to the psychological emotion coming from individual factors (eg, clinical title and perceived illness risks), whereas organizational

pressure comes from organizational factors (eg, medical policy and workload). Given the psychological aspects, physicians generally make decisions based on the balance and appropriateness between evidence (eg, patients' real-time conditions) and their cognition (eg, a domain of knowledge and previous experience) [22]. Furthermore, anchoring bias is quite common in the decision-making process, most particularly in situations with a high level of uncertainty regarding individual heterogeneity in both patients (eg, health evidence) and physicians (eg, cognition). Hence, under pressure from an individual and organization, the effects of PASS vary with various types of pressure by having different reactions to the implementation of PASS.

In the context of this study, PASS provides not only the basic functions for integrating knowledge from diverse sources (eg, information retrieval) but also supports the decision through the integration of knowledge presented through advanced technology (eg, alerting light). Despite the similarities of intelligent decision-making support between PASS and smart diagnoses, the visualization information derived from PASS could provide more efficiency in updating the knowledge of physicians to enhance their decision-making processes. For example, physicians will know what types of new drugs are inappropriate for specific cases by merely observing the alerting lights. Hence, based on the limited research exploring related topics, this study will investigate the correlation between the adoption of PASS and medical performance from the knowledge-based view.

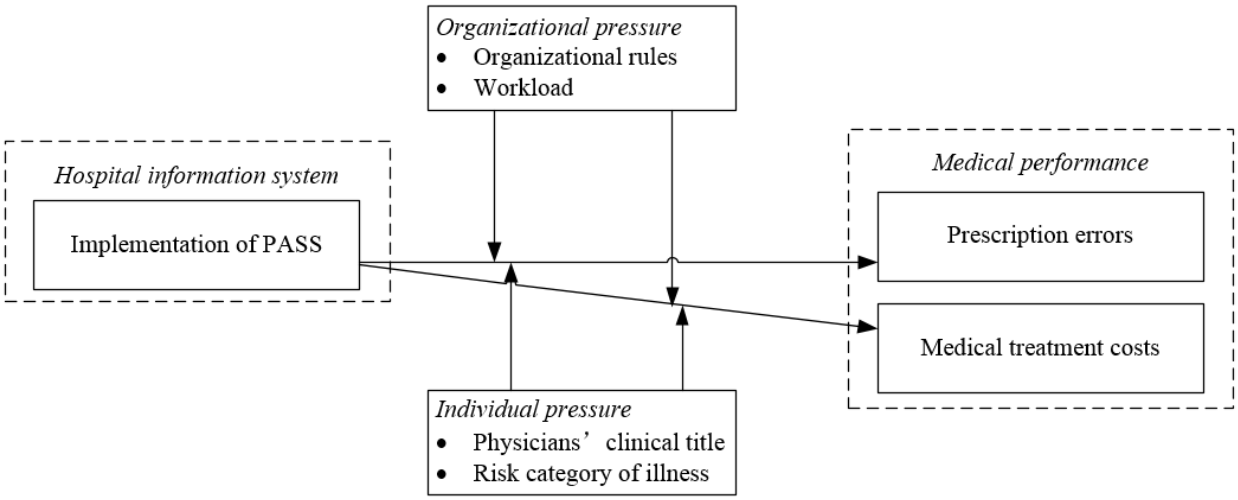
To further explore the moderators in this context, we conducted several interviews with caregivers working in hospitals (ie, 3 physicians, 2 nurses, 4 IT employees, and 3 administrators), so as to further explore the actual usage behavior of PASS (details of interview results are provided in Multimedia Appendix 1). The findings of these interviews are presented as follows:

- Physicians may ignore the alerting information when working under high workload.
- The weekly report of medicine usage may influence physicians' performance. In some hospitals, departments will have a statistic report of medicine usage for each physician every week. This report will calculate the specific dose of each medicine per physician, such as antibiotics.
- Physicians who have an elevated position and level of responsibility may have different opinions that may contradict the alerting information.
- Physicians may spend more time to make a decision when confronting cases with a high risk (details shown in Multimedia Appendix 1).

On the basis of the above-mentioned findings, we identify 4 moderators: workload, organizational rules, clinical title, and risk of diseases. In line with the knowledge-based view and pressure theory, workload and organizational rules are organizational events that will create pressure on physicians; physicians' clinical title and the risk of diseases are individual issues. High pressure may distract physicians' attention, which, to some extent, will influence the effects of PASS on medical performance. For example, with high pressure of workload, physicians may accelerate the decision process and follow their anchoring cognitions, which may impair the effects of PASS. The research framework is shown in Figure 2.

The objectives of this study were to (1) explore the effects of PASS (an auxiliary system of the HIS) on prescription errors and medical treatment costs and (2) examine how the effects of PASS vary when physicians are under pressure from the organization (eg, a high workload and organizational rules from administrators) and individual (eg, clinical title and risk of diseases).

Figure 2. Framework of the Prescription Automatic Screening System (PASS) working principle.



Methods

Experiment Design

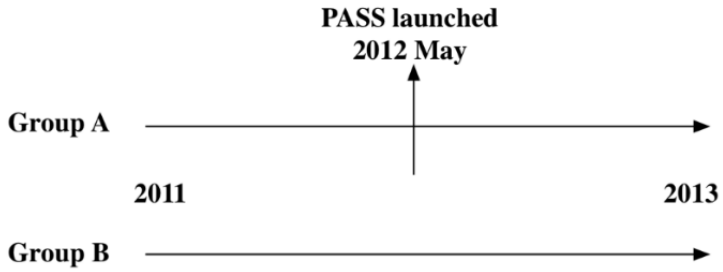
To test the proposed relationships, we used the quasi-experiment methodology in this study. We used 2 groups to compare the outcomes, that is, the control group and the treatment group. To be specific, in our context, we chose a hospital that had already deployed a PASS into practice as the treatment group (group A) and a hospital that had not deployed a PASS as the control group (group B) as presented in Figure 3.

There may exist endogeneity problems because of the effects of physician-specific unobserved factors on prescription errors and costs. To be specific, physicians' individual preferences toward the system may simultaneously influence their using behavior of the system. Thus, to solve the endogeneity problems arising from the self-selection, this study adopted the propensity score matching (PSM) integrated with the difference-in-differences (DID) analysis as recommended by the current literature [23,24].

Furthermore, the methodology of DID estimates the difference in pre- and postbehavior or outcome differences between the 2 groups of physicians, the treatment group (ie, physicians who use PASS) and the control group (ie, physicians who do not use PASS). Because a comparison of the differences between pre- and postbehavior could not eliminate the extraneous factors, the DID provides a method to adopt the benchmark physicians who do not use the PASS to control the influence of extraneous factors. Thereafter, to eliminate the differences between groups, the physicians in these 2 groups should have similar individual features, such as similar clinical title and gender. Hence, this study used the PSM to match the similar physicians in 2 groups, which could impair the temporally invariant estimation bias and also simulate a randomized experimental setup [25].

Finally, the endogeneity problem may arise from the actual use behavior of physicians. First, because PASS is an assistant tool in HIS, caregivers will use it from the day the system is launched, as the implementation of this system is mandatory for hospitals. Furthermore, in the evidence of the interviews with physicians, IT employees, and administrators, their answers support that physicians adopt PASS while they are making prescriptions. Hence, the physicians in the treatment group actually use PASS.

Figure 3. Research design. PASS: Prescription Automatic Screening System.



Data Collection

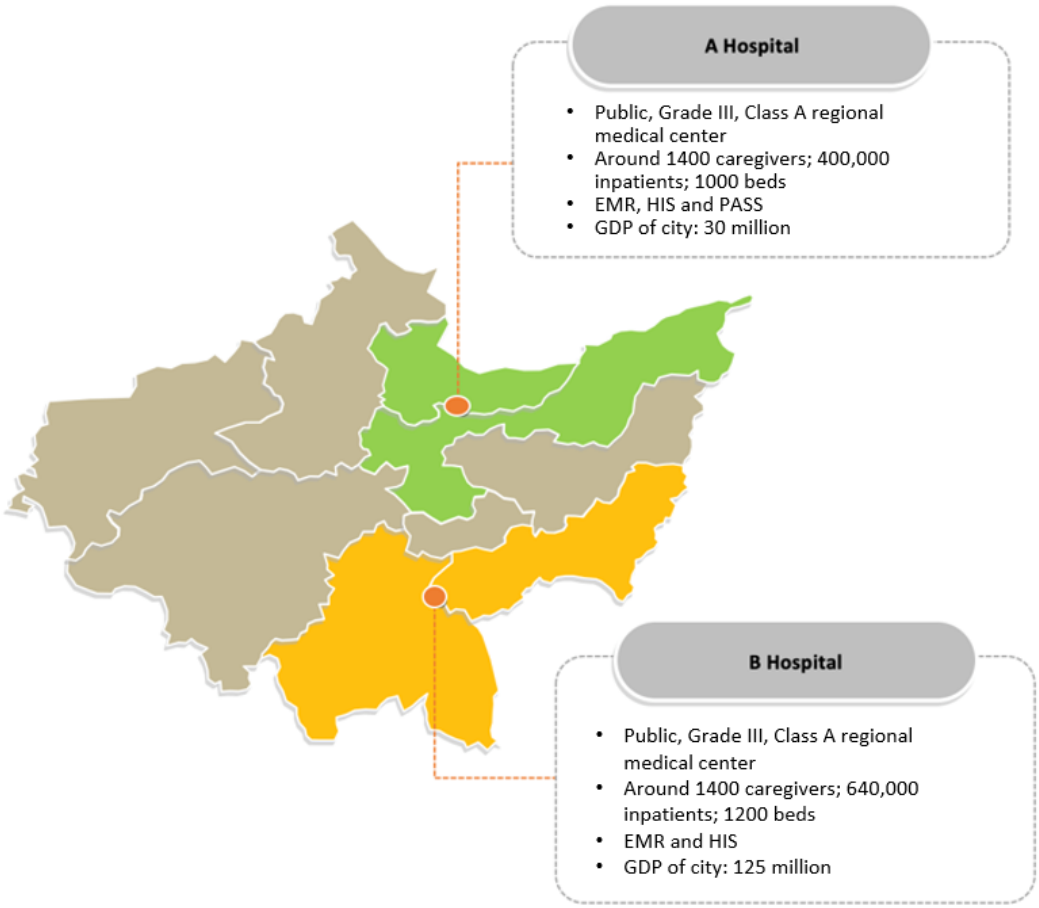
To test the model, we collected data from 2 hospitals in the same province in China. They are all public grade III hospitals, which are also the best hospitals in each city. They have similar features in basic medical conditions, technology facilities, and organizational environment. Specifically, both hospitals, have approximately 1000 medical caregivers and 0.5 million inpatients, and they make use of advanced HITs (ie, EMR and HIS). Moreover, these 2 hospitals are located in the same province having a similar natural environment, such as the cold weather, which causes respiratory illnesses in residents. The control hospital has a higher gross domestic product (GDP) than the treatment hospital, but this may have little impact on the experiment. This is because both hospitals invested similar advanced information systems; physicians and patients share similar economic status, which explains that GDP has little impact on physicians' decisions and medical costs. Visualization of basic information has been shown in Figure 4. Robustness check also proves that the confounding effect is not from the differences between the 2 hospitals.

PASS deployed in hospital A is developed by 1 of the 2 biggest software firms. PASS provides information and decision support to physicians and pharmacists. To ensure the information quality, a database of PASS collects information (eg, medicine information, medical policy, and rules) from an authoritative medical dataset and updates every 2 years. Hence, PASS could provide reliable information for physicians to assist their decisions.

We chose the general inpatients as the sample frame of this study. This is because inpatients would be taking an array of medicines; as such, prescribing another medication could easily lead to incompatibility and interactions among them. Next, we used medicine usage information and medical information based on our selected hospitals in China during the period between 2011 and 2013. In group A, the hospital using a PASS, this system was introduced in May 2012, which enabled examination before and after system implementation of PASS.

According to descriptive statistics, the overall data included 754 physicians (400 in group A and 354 in group B), and 11,054 patients (5199 in group A and 5855 in group B). Group A included 55.5% (222/400) of male physicians and 44.5% (178/400) of female physicians, whereas group B included 58.2% (206/354) of male physicians and 41.8% (148/354) of female physicians. Moreover, our data indicated 26.7% (201/754) of high clinical title physicians (ie, chief physicians), 28.9% (218/754) of physicians of medium clinical title (ie, attending physicians), and 44.4% (335/754) of low clinical title physicians (ie, physicians).

Figure 4. Hospital descriptions. EMR: electronic medical record; HIS: hospital information systems; PASS: Prescription Automatic Screening System; GDP: gross domestic product.



Variables Description

To measure medical performance, we adopted medical treatment costs and prescription errors as the dependent variables of this study. We calculated the medical treatment costs by $\sum(a_{ij}/n_i)$, where a_{ij} denotes the fee of each medical category *j* for patient *i*. We then calculated the number of prescriptions withdrawn within 10 min, and we excluded the data of prescriptions that were withdrawn because of patients' reasons. We assumed that withdrawing prescriptions within 10 min was unnecessary, which could be well avoided during the process of prescription making.

Hospitals usually implement relevant policy to ensure the effectiveness of information systems. Hence, we used a dummy variable to measure the organizational rules, that is, if the department implements rules related to the use of PASS, then the dummy variable is 1, otherwise the dummy variable is 0. We calculated the total number of patients per physician to measure the workload. Finally, based on a reference to the existing classification about the severity of illness, this study identified 4 categories (ie, 1-4) of illness risk by manual labeling [26]. Because the decisions of physicians depend partially on the individual's characteristics, we used gender as the control variable as presented in Table 1.

Table 1. Variable definitions.

Variables	Symbols	Measurements
Dependent variable		
Prescription errors	Error	The average number of prescriptions withdrawn within 10 min per physician in a month
Medical treatment costs	Cost	The average medical costs per physician, $\sum(a_{ij}/n_i)$
Independent variable: implementation of PASS[a]		
Time of PASS	InSys	The time of the system launched, 0=No, 1=Yes
Treatment	Treatment	Whether hospitals implement the PASS, 0=No, 1=Yes
Moderator: organizational pressure		
Workload	WorkLoad	The daily number of patients seen by a physician per month
Organizational rules	Ins_pres	Whether the physicians stay in the department having a statistical report related to usage of PASS every month, 0=No, 1=Yes
Moderator: individual pressure		
Risk category of illness	Risk	The risk category based on the case information, for example, readmission times, age, inpatient health condition, and ICD-10[b]
Physicians' clinical title	Title	Dummy variables of physicians' clinical title, such as chief physician and attending physician
Control variables		
Gender	Gender	Gender of physicians, that is, male or female

[a]PASS: Prescription Automatic Screening System.

[b]ICD-10: International Classification of Diseases, Tenth Revision.

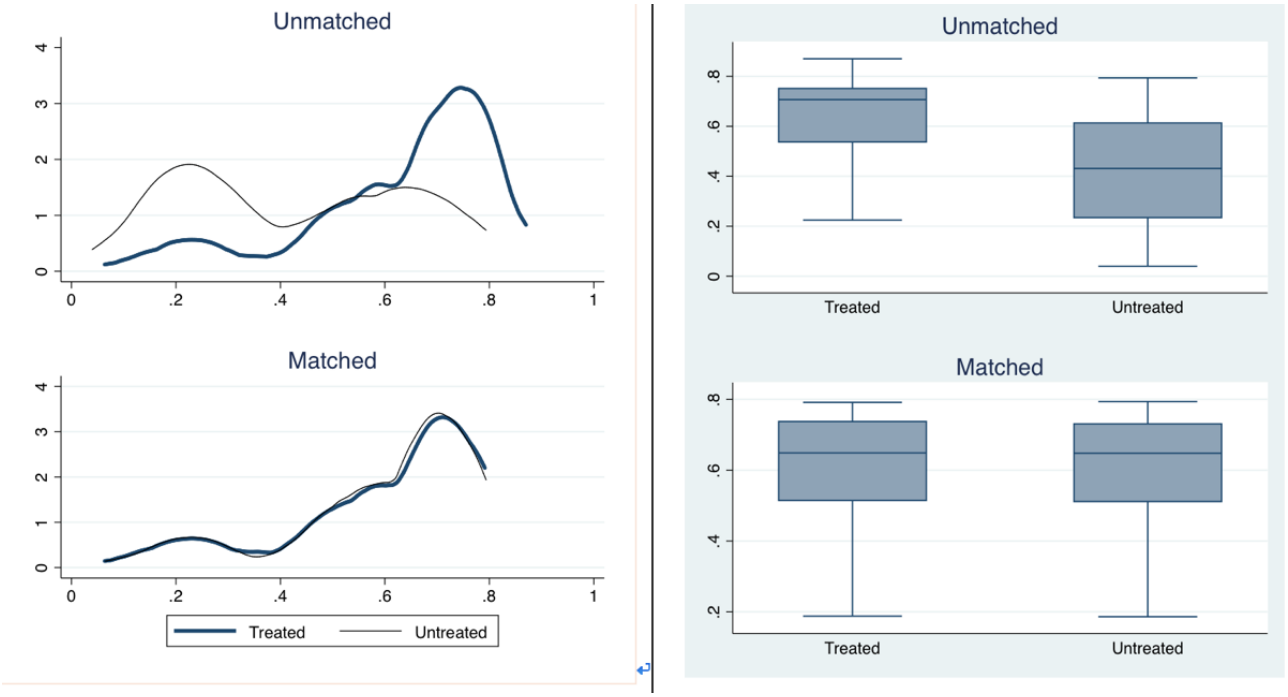
Model Specifications

Propensity Score Matching

To measure the causality between the system and user performance and to eliminate the sample differences between the 2 hospitals, we formed group A (the treatment group) and group B (the control group) using the PSM method to compare the effects. We created a statistical equivalence to balance all relevant characteristics that existed before the system launch [27]. Data before the intervention were available in both groups, and we used 12 months of data pertaining to patient characteristics for both groups before the launch of the PASS.

We used the kernel-based method in PSM. As physicians' gender partly influences prescriptions, this reflects differences in attitude toward technology adoption [28]. In addition, the physicians' clinical title and the number of patients per physician also play a role in physicians' use intentions, as an overworked physician will possess lower work efficacy and face greater pressure. On the basis of the preceding section, we considered these variables as covariant variables (see Multimedia Appendix 1). After matching the 2 groups, we had 695 physicians. To check the substantial overlap in the characteristics of the physicians who adopted PASS and those who did not (ie, common support conditions), we conducted a visual analysis of the propensity score distributions through box plots and histograms (see Figure 5) and found evidence for the existence of common support.

Figure 5. A visual analysis of propensity score distributions through box plots and histograms.



Effects of Prescription Automatic Screening System on Prescription Errors and Medical Treatment Costs

This study combines the PSM and DID methods to verify the before and after effects of (1) prescription errors and (2) medical treatment costs. For treatment and control groups, the logarithm of the error of the prescription is modeled as follows:

$$\mathrm{Ln(Erro_{it})} = \delta_{oj} + \delta_1 \mathrm{Treat_i} + \delta_2 \mathrm{InSys_{it}} + \delta_3 \mathrm{Treat_i} \times \mathrm{InSys_{it}} + \Theta X_i + \xi_{it} \quad (1)$$

The independent variables in the equation of the other 2 dependent variables are the same. In equation 1, *i* denotes a treatment group or a control group physician and *t* denotes the time period. $Treat_i$ is the treatment dummy variable (1 denotes that the physician is in the treatment group and 0 denotes that the physician is in the control group), whereas $InSys_{it}$ is a dummy variable denoting the launch of PASS, taking the values 0 and 1 for periods before and after the system launch, respectively. For physicians belonging to the matched pair *i*, X_i represents a vector of control variables, with Θ being their corresponding estimated coefficients.

δ*oj* refers to the physician-specific fixed effects that capture the differences in baseline relationship intensity, which enable the controlling of unobserved heterogeneity among physicians. It is to be noted that in the above formula consisting of matched treatments and control group physicians, monthly data that span both pre- and postlaunch time periods of PASS of this study are used (December 2011-December 2012), resulting in time-series data that are then stacked for estimation. The main parameter is δ_3, which captures the changes in the average length of stay for treatment and physicians post adoption compared with physicians of the control group who did not adopt the system.

Moderating Effects of Individual and Organizational Pressure

Next, we described an alternative version of the model to the one presented previously (in equation 1), which enabled us to test the hypotheses posited earlier. Therefore, to investigate the impact of moderating variables, we used the following formula:

$$\mathrm{Ln(Error_{it})} = \gamma_{oj} + \gamma_1 \mathrm{TreatD_i} + \gamma_2 \mathrm{InSys_{it}} + \gamma_3 \mathrm{TreatD_i} \times \mathrm{InSys_{it}} + \gamma_4\ \mathrm{TreatD_i} \times \mathrm{Moderator_i} + \gamma_5 \mathrm{InSys_{it}} \times \mathrm{Moderator_i} + \gamma_6 \mathrm{TreatD_i} \times \mathrm{InSys_{it}} \times \mathrm{Moderator_i} + \Omega X_i + \varepsilon\_it \quad (2)$$

In equation 2, the variables have the same meaning as in equation 1, and $Moderator_i$ refers to the moderators in this study including organizational rules and workload per physicians, the risk category of illness, and the clinical title of physicians.

Results

The Impacts of Prescription Automatic Screening System on Errors of Prescription and Medical Treatment Costs

This paper estimated a series of alternative models to measure the results and statistical fit of our DID model. We used a basic DID model without any control variables in model 1 while inserting control variables and physicians' information into model 2. Next, we included moderators in model 3. According to the results presented in Multimedia Appendix 2, the result of fit statistic (R^2) was seen to increase from model 1 to model 3 toward every variable, which supports the validity of the results.

With reference to the error of prescription, the results in Multimedia Appendix 2 indicate that the parameter of interactions between system onset and time lapse was continuously significant and negative from model 1 to model

2 (beta=–.246 *P*=.001; beta=–.257 *P*<.001). This indicates that physicians withdraw fewer times of prescriptions in hospital after system use. With the control variables to measure validity, the interaction variable parameter was still significantly negative in model 2 (95% CI –0.40 to –0.11; *P*<.001), indicating a negative impact of the system on the error of prescription. With reference to medical treatment costs, the results in Multimedia Appendix 2 show that the parameters of interaction between the system and cost are continuously significant and negative from model 1 to model 2 (beta=–.371 *P*=.007; beta=–.389; *P*=.007). With the control variables to measure the validity, the interaction variable parameter was still significantly negative in model 2 (95% CI –0.75 to –0.12; *P*=.007). This illustrates that PASS used in group A helped in lowering the costs for patients' hospital stay, thus reflecting a lowered medical burden. However, according to R^2, the model in this study could explain more about physicians' medical performance.

The Impact of Individual and Organizational Pressure

The effects of PASS implementation for the different moderating variables and the results of the difference-in-difference-in-difference (DDD) model are presented in model 3 in Multimedia Appendix 2. Moderating results relatively impact the moderating variable on dependent variables based on the 2-way interactions of the treatment effects in the DDD model. Moreover, the effects of system implementation on the error of prescriptions will differ depending on whether physicians belong to the department having PASS-related organizational rules. The parameters of interactions between the implementation of organizational rules and treatment effect (Treatment×InSys×Ins_pres) in model 3 (Multimedia Appendix 2) showed that PASS has fewer effects on prescription errors when physicians perceive high pressure from organizations (95% CI 0.07-0.55; *P*=.01). With respect to the medical treatment costs, the parameter of the interaction between workload and treatment effect (Treatment×InSys×Workload) was positively significant (95% CI 0.18-0.39; *P*<.001).

Additional Analysis

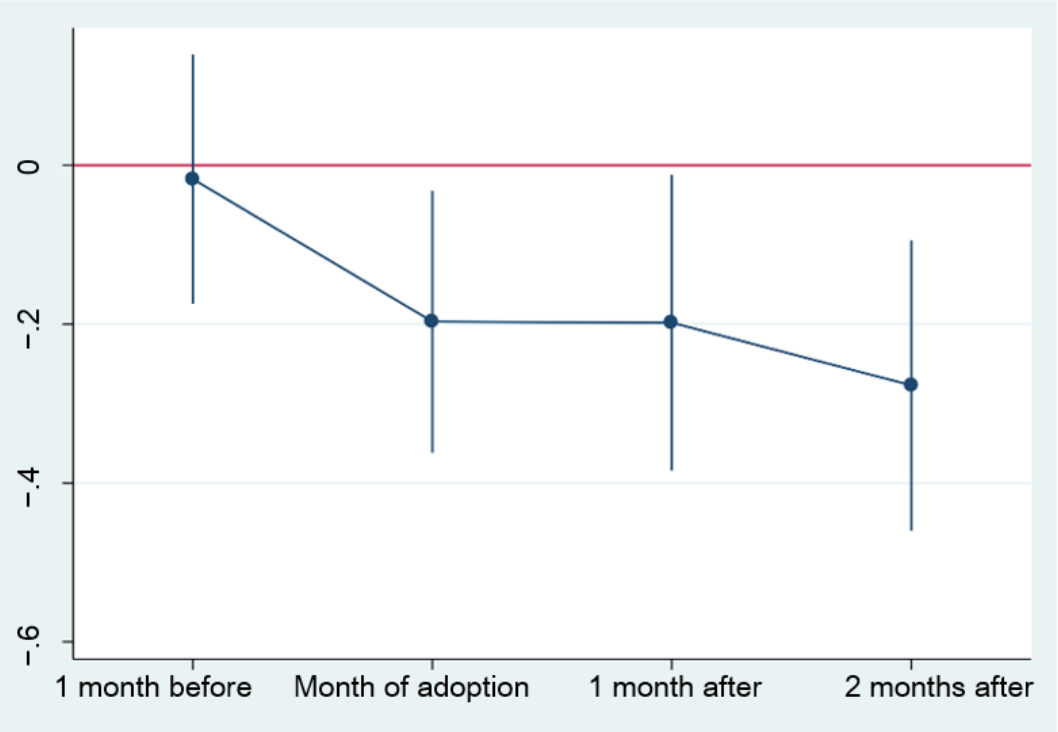
To examine the differences in hospitals before the implementation of PASS, we constructed the parallel trend test. Specifically, we added 2 indicator variables for each month before the system change, 3 indicators for each month after the system change, and the interaction terms of indicators and treatment. We chose 2 months before PASS as the baseline; the final results are presented in Table 2. According to the results in Table 2, there are no significant differences between the 2 hospitals before PASS implementation. Significant decrease in cost was observed in the month of adoption. The results also showed that the overall costs continued to decrease after the system change. We visualized this pattern in Figure 6.

Table 2. Results of parallel trend test.

Variables	Ln (cost)
Treatment	–0.286[a]
1 month before	–0.016
Month of adoption	–0.196[a]
1 month after	–0.197[b]
2 months after	–0.276[a]
Title_dummy1	0.129
Title_dummy2	0.067
Title_dummy3	0.191[b]
Ln(Workload)	–0.071
Gender	0.257[a]
_cons	9.244[c]
R^2	0.097

[a]*P*<.05.
[b]*P*<.01.
[c]*P*<.001.

Figure 6. Time trends relative to month of adoption of the Prescription Automatic Screening System (PASS).



Discussion

Prescription Automatic Screening System Reduces Prescription Errors and Medical Treatment Costs

Because the interaction of *Treatment* and *InSys* is negatively significant, applying PASS will reduce the prescription errors (beta=–.246, *P*=.001; beta=–.257, *P*<.001) and medical treatment costs (beta=–.371, *P*=.007; beta=–.389, *P*=.007). The results are consistent with the previous study that HISs decrease the prescription errors [14] and medical treatment costs [29]. As the primary role of IT, as applied to hospitals, the PASS system integrates knowledge derived from diverse individual specialists to support prescription-making decisions. This shortens the time in the treatment process and improves the effectiveness of using IT to support the diagnoses. Moreover, these findings highlight the various roles of IT to promote more appropriate coordination among individuals within the hospital setting, which could be used to improve the quality of health care. However, IT does exert a variety of effects when applied to different environmental considerations.

Individual Pressure Presents No Impact

Contrary to the hypothesis, the moderators, risk categories (beta=.057, *P*=.33) and clinical title (beta=.272, *P*=.29; beta=.190, *P*=.47; beta=.200, *P*=.47) exerted insignificant effects, which is inconsistent with previous studies [30]. This may be due to the features of the medical domain; physicians will focus high attention on patients' health conditions regardless of whether patients stay in high- or low-risk conditions. Further referring to the moderating effect of physicians' title, the effects of the system on mitigating the anchoring bias are quite low because this factor will also influence the physicians' behavior for physicians with both high and low clinical statuses. Hence, in the context of health care, personal characteristics will not limit the effects of the intelligent diagnosis system, which manifests a significant potential if hospitals could deploy more IT into their protocols to provide supplemental support for physicians' decisions.

Workload and Implementation of Organizational Rules Related to Usage of Prescription Automatic Screening System Decrease the Impact of Prescription Automatic Screening System

The parameters of interactions between organizational rules and the main model show that PASS has fewer effects on prescription errors when physicians perceive high pressure from organizations (95% CI 0.18-0.39; *P*<.001). With respect to the medical treatment costs, the parameter of the interaction between workload and main model is positively significant (95% CI 0.07-0.55; *P*=.01). The findings of the moderating effect indicate that environmental factors affect how IT alters users' performance. Previous literature depicted an insignificant impact concerning related IT systems on physicians' performance [24]. However, the findings of this study expanded on the previous model and proved that organizational rules might help clarify that the impact of IT will decrease when physicians perceive high pressure from organizations [31]. These results highlight the critical roles of management within organizations when they adopt IT in the workplace. The findings also emphasize that further exploration is needed to determine why pressure tends to eliminate the impact of IT from a psychological perspective.

Concerning the medical treatment costs, the results of the moderating effects related to the workload indicate that when physicians have more patients awaiting treatment, the effects of IT will decrease. This may be due to the fact that a high level of work pressure motivates physicians to depend more on their knowledge and experience, which will then lead them to ignore or disregard important alerting information, which is consistent

with the previous study [32]. On the basis of this premise, the performance will not lead to significant differences even after the hospital deploys the PASS system.

In general, when hospitals adopted the IT system to enhance medical performance, they also implemented a corresponding policy designed to increase the effectiveness of IT usage. However, based on our findings, when the policy places too much pressure on the physicians, it will have a paradoxical result. Hence, our findings provide some significant new insights for policy implementation in hospitals, such as how to appropriately balance the policy between IT and organizational management protocols and how to effectively enact the evaluation criteria with regard to physicians' performance.

Strengths of This Study

In this study, we examined the impact of the PASS system by conducting a quasi-experiment, which could help eliminate the effects from various confounding factors and further highlight the causality between PASS and medical performance. According to the previous study, the impact of PASS may be quite varied based on the level of pressure exerted by the organizational environment. Through several cooperative interviews with physicians and administrators in hospitals, we obtained detailed information relating to the physicians' attitudes and actual usage of PASS as further guidance in the exploration of environmental impact. Thus, we were able to evaluate how PASS plays its role in hospitals. We firmly believe these findings will provide practical suggestions for hospitals and their administration to garner a higher level of performance from their workforce after deploying the related systems.

Limitations and Directions for Future Research

This study has 2 limitations. First, because some of the moderating effects in this study are insignificant, further exploration of these factors (ie, title and risk) is necessary, particularly, when emerging technologies (eg, artificial intelligence) are considered for use in hospitals. The problems to mitigate the new uncertainty, which derives from the new adoption behavior, are critical to improving the effectiveness of implementing HIS. To ascertain extended performance from PASS, attention must be paid to more categories of different hospitals with different characteristics. Second, based on the preliminary investigations, we need to proceed to provide a more specific classification of alerts, that is, physicians' decision stages. This process requires further discussions with physicians, and the results will provide support for hypotheses on physicians' performance and decision stages. We are aware that much more cooperation and data are required. In addition, although this paper examines the effects of the adoption of the PASS, the effective use of such a system will attract greater attention because of a higher quality of the treatment process, which we will investigate in our future research.

Conclusions

This study found that PASS, a potential tool to integrate knowledge from various expertise, has positive effects on medical performance; however, organizational pressure raises a concern on the effectiveness of PASS. Specifically, we found that compared with individual pressure (eg, clinical title and disease risk), it is the pressure from the organization (eg, organizational rules and workload) that reduces the effectiveness of PASS. Hence, the strategies adopted by hospitals, which are used to improve the effectiveness of HIS implementation, may not work. The findings indicate that management in hospitals needs to balance the relationship between HIS implementation and policy making to augment the positive effects of HIS.

Acknowledgments

The authors are grateful to the editors for their guidance and to the anonymous reviewers for their constructive suggestions. This study was partially funded by the National Natural Science of China (71622002, 71531007, 71471048, 71871074, and 71871073).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of interview results from 2 hospitals.

[DOCX File, 28KB - medinform_v7i2e11663_app1.docx]

Multimedia Appendix 2

Results of regression analyses.

[DOCX File, 17KB - medinform_v7i2e11663_app2.docx]

References

1. Bennani K, Bukassa G, Khattabi A, Akrim M, Maaroufi A. Evaluation of drug-resistant tuberculosis surveillance system before and after implementing Genexpert in Morocco, 2007-2016. IProc 2018 Mar 29;4(1):e10535. [doi: 10.2196/10535]

2. Healthcare Information and Management Systems Society. 2017. Essentials BriefClinical & Business Intelligence Study URL: https://www.himssanalytics.org/sites/himssanalytics/files/2017_Essentials%20Brief_Mobile_SNAPSHOT%20REPORT.pdf [accessed 2019-05-30] [WebCite Cache ID 78l4CmuOI]
3. CHIMA. 2018. 2017-2018 Hospital Informationization Report in China URL: http://www.chima.org.cn/index.php?m=content&c=index&a=show&catid=10&id=170 [accessed 2019-02-03] [WebCite Cache ID 75ukIhxi7]
4. Wipfli R, Ehrler F, Bediang G, Bétrancourt M, Lovis C. How regrouping alerts in computerized physician order entry layout influences physicians' prescription behavior: results of a crossover randomized trial. JMIR Hum Factors 2016 Jun 2;3(1):e15 [FREE Full text] [doi: 10.2196/humanfactors.5320] [Medline: 27255612]
5. Nanji KC, Seger DL, Slight SP, Amato MG, Beeler PE, Her QL, et al. Medication-related clinical decision support alert overrides in inpatients. J Am Med Inform Assoc 2018 May 1;25(5):476-481. [doi: 10.1093/jamia/ocx115] [Medline: 29092059]
6. Schiff GD, Volk LA, Volodarskaya M, Williams DH, Walsh L, Myers SG, et al. Screening for medication errors using an outlier detection system. J Am Med Inform Assoc 2017 Mar 1;24(2):281-287. [doi: 10.1093/jamia/ocw171] [Medline: 28104826]
7. Baypinar F, Kingma H, van der Hoeven RT, Becker M. Physicians' compliance with a clinical decision support system alerting during the prescribing process. J Med Syst 2017 Jun;41(6):96 [FREE Full text] [doi: 10.1007/s10916-017-0717-4] [Medline: 28480481]
8. Kane-Gill SL, Achanta A, Kellum JA, Handler SM. Clinical decision support for drug related events: moving towards better prevention. World J Crit Care Med 2016 Nov;5(4):204-211. [doi: 10.5492/wjccm.v5.i4.204]
9. Desmedt S, Spinewine A, Jadoul M, Henrard S, Wouters D, Dalleur O. Impact of a clinical decision support system for drug dosage in patients with renal failure. Int J Clin Pharm 2018 Oct;40(5):1225-1233. [doi: 10.1007/s11096-018-0612-1] [Medline: 29785684]
10. Gagnon M, Simonyan D, Ghandour E, Godin G, Labrecque M, Ouimet M, et al. Factors influencing electronic health record adoption by physicians: a multilevel analysis. Int J Inf Manage 2016 Jun;36(3):258-270. [doi: 10.1016/j.ijinfomgt.2015.12.002]
11. 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]
12. Angst C, Devaraj S, D'Arcy J. Dual role of IT-assisted communication in patient care: a validated structure-process-outcome framework. J Manage Inf Syst 2014 Dec 8;29(2):257-292. [doi: 10.2753/MIS0742-1222290209]
13. Sharma L, Chandrasekaran A, Boyer K, McDermott C. The impact of Health Information Technology bundles on Hospital performance: an econometric study. J Oper Manage 2016 Nov;41(1):25-41. [doi: 10.1016/j.jom.2015.10.001]
14. Aron R, Dutta S, Janakiraman R, Pathak PA. The impact of automation of systems on medical errors: evidence from field research. Inf Syst Res 2011 Sep;22(3):429-446. [doi: 10.1287/isre.1110.0350]
15. Grant R. Toward a knowledge-based theory of the firm. Strategic Manage J 1996;17(S2):109-122. [doi: 10.1002/smj.4250171110]
16. Premkumar G, Ramamurthy K. The role of interorganizational and organizational factors on the decision mode for adoption of interorganizational systems. Decis Sci 1995 May;26(3):303-336. [doi: 10.1111/j.1540-5915.1995.tb01431.x]
17. Alavi M, Leidner DE. Review: knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Q 2001 Mar 21;25(1):107-7 concl. [doi: 10.2307/3250961]
18. Giordanengo A, Øzturk P, Hansen AH, Årsand E, Grøttland A, Hartvigsen G. Design and development of a context-aware knowledge-based module for identifying relevant information and information gaps in patients with type 1 diabetes self-collected health data. JMIR Diabetes 2018 Jul 11;3(3):e10431. [doi: 10.2196/10431]
19. Hsiao C, King J, Hing E, Simon AE. The role of health information technology in care coordination in the United States. Med Care 2015 Feb;53(2):184-190. [doi: 10.1097/MLR.0000000000000276] [Medline: 25464164]
20. Zolhavarieh S, Parry D, Bai Q. Issues associated with the use of semantic web technology in knowledge acquisition for clinical decision support systems: systematic review of the literature. JMIR Med Inform 2017 Jul 5;5(3):e18 [FREE Full text] [doi: 10.2196/medinform.6169] [Medline: 28679487]
21. Yang D, Yang L. Organizational pressure dimensions measuring of “post-80s” employees and its implications for talent strategy. Open J Soc Sci 2014;02(12):13-18. [doi: 10.4236/jss.2014.212003]
22. Furnham A, Boo HC. A literature review of the anchoring effect. J Socioecon 2011 Feb;40(1):35-42. [doi: 10.1016/j.socec.2010.10.008]
23. Rishika R, Kumar A, Janakiraman R, Bezawada R. The effect of customers' social media participation on customer visit frequency and profitability: an empirical investigation. Inform Syst Res 2013 Mar;24(1):108-127. [doi: 10.1287/isre.1120.0460]
24. Huang Q, Nijs VR, Hansen K, Anderson ET. Wal-Mart's impact on supplier profits. J Market Res 2012 Apr;49(2):131-143. [doi: 10.1509/jmr.10.0256]
25. Rubin D. Matched Sampling For Causal Effects. United States of America: Cambridge University Press; 2006.
26. Gonnella J. Healthcare Cost and Utilization Project. 2001. Disease Staging Clinical And Coded Criteria URL: https://www.hcup-us.ahrq.gov/db/nation/nis/DiseaseStagingV5_26ClinicalandCodedCriteria.pdf.Accessed: 2019-05-30 [accessed 2019-05-30] [WebCite Cache ID 78l5BqGrw]

27. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55. [doi: 10.1093/biomet/70.1.41]
28. Nov O, Schecter W. Dispositional resistance to change and hospital physicians' use of electronic medical records: a multidimensional perspective. J Am Soc Inf Sci 2012 Feb 1;63(4):648-656. [doi: 10.1002/asi.22602]
29. Atasoy H, Chen P, Ganju K. The spillover effects of health IT investments on regional healthcare costs. Manage Sci 2018 Jun;64(6):2515-2534. [doi: 10.1287/mnsc.2017.2750]
30. Paul J, MacDonald L, Hariharan G. Modeling risk factors and disease conditions to study associated lifetime medical costs. Serv Sci 2014 Mar 10;6(1):47-62. [doi: 10.1287/serv.2014.0063]
31. Barrett M, Oborn E, Orlikowski W. Creating value in online communities: the sociomaterial configuring of strategy, platform, and stakeholder engagement. Inf Syst Res 2016 Dec;27(4):704-723. [doi: 10.1287/isre.2016.0648]
32. Mazur L, Mosaly PR, Moore C, Comitz E, Yu F, Falchook AD, et al. Toward a better understanding of task demands, workload, and performance during physician-computer interactions. J Am Med Inform Assoc 2016 Dec;23(6):1113-1120. [doi: 10.1093/jamia/ocw016] [Medline: 27026617]

Abbreviations

ADE: adverse drug events
ADR: adverse drug reactions
DDD: difference-in-difference-in-difference
DID: difference-in-differences
EMR: electronic medical record
GDP: gross domestic product
HIS: hospital information system
HIT: hospital information technology
IT: information technology
PASS: Prescription Automatic Screening System
PSM: propensity score matching

Edited by G Eysenbach; submitted 24.07.18; peer-reviewed by Y Tong, B Kim, A Sharafoddini, X Zhang; comments to author 08.10.18; revised version received 10.12.18; accepted 17.02.19; published 14.06.19.

Please cite as:
Li Y, Guo X, Hsu C, Liu X, Vogel D
Exploring the Impact of the Prescription Automatic Screening System in Health Care Services: Quasi-Experiment
JMIR Med Inform 2019;7(2):e11663
URL: http://medinform.jmir.org/2019/2/e11663/
doi:10.2196/11663
PMID:31199314

Original Paper

Understanding of and Barriers to Electronic Health Record Patient Portal Access in a Culturally Diverse Pediatric Population

Daniel J Miklin[1], MD; Sameera S Vangara[1], MD; Alan M Delamater[2], PhD, ABPP; Kenneth W Goodman[3], PhD, FACMI

[1]University of Miami Miller School of Medicine, Miami, FL, United States

[2]University of Miami Miller School of Medicine, Mailman Center for Child Development, Miami, FL, United States

[3]University of Miami Miller School of Medicine, Institute for Bioethics and Health Policy, Miami, FL, United States

Corresponding Author:
Daniel J Miklin, MD
University of Miami Miller School of Medicine
1600 NW 10th Ave
Miami, FL, 33136
United States
Phone: 1 3035188936
Email: dmiklin@med.miami.edu

Abstract

Background: Electronic health records (EHRs) have become a standard in the health care setting. In an effort to improve health literacy, foster doctor-patient communication, and ease the transition from adolescent to adult care, our institution created a policy that allows patients aged between 13 and 17 years to have EHR portal access. A literature review revealed predictable differences in portal registration among different ethnicities and socioeconomic statuses. Consequently, a cross-sectional survey was developed to investigate barriers to EHR portal access in a sample of culturally diverse adolescents.

Objective: The aim of this study was to assess for barriers to EHR portal access in a culturally diverse adolescent population.

Methods: A 42-item anonymous survey was completed by 97 adolescents aged between 13 and 18 years, attending general pediatrics clinics. The results were analyzed using descriptive statistics and *t* tests.

Results: The average participant age was 15.5 (SD 1.5) years with 60% (58/97) male and 40% (39/97) female. Participants were 44% (43/97) black, 41% (40/97) Hispanic, 9% (9/97) Caucasian, 3% (3/97) Asian, and 2% (2/97) others. There were statistically significant differences in perceptions of confidentiality in age (13 to 15 years vs 16 to 18 years; P=.001) and insurance status (government vs private; P=.012) but not in gender, ethnicity, or parental education level. Younger adolescents with governmental insurance were more confident in the level of confidentiality with their physician. A total of 94% of participants had heard of the term *EHR*, but only 55% were familiar with its function. Furthermore, 77% of patients primarily accessed the internet through phones, and 50% of participants knew that patients aged under 18 years could obtain care for mental health, substance abuse, sexual health, and pregnancy.

Conclusions: This research has identified gaps in EHR technology with regard to the pediatric patient population. The results of our survey show that adolescents may have misconceptions regarding the doctor-patient relationship, their ability to obtain care, and the modalities present in an EHR. As technology progresses, it is essential to have a deeper understanding of adolescents' perceptions of confidentiality, technology, and available resources to design an EHR system that encourages patient education and communication while limiting barriers to care.

(JMIR Med Inform 2019;7(2):e11570) doi:10.2196/11570

KEYWORDS

electronic health records; physician-patient relations; adolescent health; patient-accessible electronic health records; patient portals

Introduction

Background

Since the passing of the Health Information Technology for Economic and Clinical Health Act within the American Recovery and Reinvestment Act of 2009, electronic health record (EHR) systems have become a standard of care within the health care field [1]. Providers are not only encouraged to use these systems but are also reimbursed for their *meaningful use*. These systems have been shown to provide benefits for patients and clinicians in terms of improved access to information, reduction of errors, faster test results, and improved outcomes [2-4]. However, EHR systems may lack some of the resources needed to address the complexities that exist in adolescent care settings with regard to confidentiality, autonomy, and access to care [3,5].

Adolescent patients, although formally considered part of the realm of pediatrics, provide an interesting challenge to health care providers because of their physical, social, and emotional changes. These patients are distinctive owing to the rapid development of autonomy, self-awareness, and responsibility. The Society for Adolescent Health and Medicine has affirmed that confidentiality is essential for the development of maturity, autonomy, and the willingness of adolescent patients to disclose sensitive information and seek care [6-8]. Special considerations are therefore needed when caring for this population, particularly when dealing with sensitive information such as sexual activity, pregnancy, substance use, and mental health. All or most of this *information* is available through patient EHR portals, although the portals lack functionality to manage confidentiality in an adequate manner. What information is present, how it is accessible, and what resources exist for learning and communicating with the physician can play a vital role in the care of an adolescent [8,9].

Objectives

The University of Miami and Jackson Memorial Health Systems are 2 major academic medical centers in an urban setting that are frequently presented with a number of medically and ethically challenging pediatric cases. Miami-Dade County is unique in its ethnically and socially diverse community and has the highest rate of new HIV infections in the country, with up to 34% of new cases occurring between ages 13 and 29 years [10]. In Florida, like many other states, pediatric patients are legally able to consent for care that involves sexually transmitted infections (STIs), drug and alcohol abuse, pregnancy, and mental health [11]. This presents a challenge toward confidentiality when trying to document and bill for care in any of these areas. Although the patient can legally consent for these services, it is often their parent/guardian who pays for them. Furthermore, because of their limited interaction with the nonclinical aspects of health care, adolescents often lack knowledge of the resources and privileges available to them [6,12,13]. Therefore, in an effort to improve health literacy, foster communication between patients and providers, and ease the transition from adolescent to adult care, members of the Ethics Committee at Holtz Children's Hospital sought to create a new policy for allowing pediatric patients to have proxy access to their EHR through a Web-based portal. A literature review and direct contact with 8 major academic and community hospitals around the country suggested that very few institutions have formal policies regarding pediatric patient access to their EHR portals [14-16]. On the basis of the information collected, the decision was made to create a policy allowing adolescent patients aged between 13 and 17 years to have proxy access to their EHR portal.

Studies suggest that there are predictable differences in portal registration patterns among different ethnicities and socioeconomic statuses [17-20]. These studies demonstrated a decrease in portal enrollment and activation by black and Hispanic adolescents, particularly those with governmental insurance [17]. Miami-Dade County has a very ethnically diverse population, with 18% black and 69% Hispanic residents and 20% of residents under the age of 18 years [21]. To our knowledge, no studies exist regarding pediatric patient perception and preferences for an EHR portal. Therefore, as a follow-up to our institution's newly approved policy for proxy EHR portal access, a 42-item anonymous survey was created to investigate barriers to accessing EHR portals in this culturally distinct population.

Methods

Survey Creation

A 42-item survey was created to assess for barriers to and understanding of EHR portals for participants aged between 13 and 18 years. A literature review of existing surveys regarding EHR use demonstrated that there were very few published questions that would be applicable to a pediatric patient population. Survey items were therefore created de novo, including questions concerning the perception of confidentiality in the doctor-patient relationship, understanding of EHR systems, and knowledge of access to care as a pediatric patient. The completed survey was sent for review for clinical relevance by practicing pediatricians from Jackson Memorial Hospital and the University of Miami Health Systems. The University of Miami Biostatistics Collaboration and Consulting Core assessed the survey for question design and organization. Once the final survey was completed, questions were grouped together based on the content and separated into 6 sections: (1) what is a medical record and how does it work, (2) abilities when using a computer, (3) statement questions, (4) assessing knowledge of confidentiality, (5) obtaining care, and (6) understanding cultural barriers (Multimedia Appendix 1). These groupings were created to assist with the survey ease of use and eventual data analysis. Survey packets containing an informational flyer, consent forms, demographic section, and the 42-item survey were assembled and individually labeled for distribution.

The study was approved by the University of Miami Institutional Review Board (IRB; study number 20160624).

Survey Distribution

Participants were identified by the age criteria before their arrival and were recruited in the waiting rooms at 3 different IRB-approved general pediatric clinics at both institutions. Inclusion criteria included subjects aged between 13 and 18 years, proficient in written and spoken English, and who were

comfortable filling out the survey without parental assistance. Although the intended proxy portal would not be applicable to those aged 18 years and above, it was believed that the feedback from those soon to be transitioning into adult care was vital for the study. Upon arrival to the clinic, participants and their guardians were greeted and informed of the nature of the study. If both the adolescents and their parents or guardians expressed interest, written consent was obtained by the guardians or adolescents aged 18 years and written assent was obtained by all adolescents aged between 13 and 17 years. For accuracy purposes, the demographic portion of the survey was completed separately by the parent or guardian. Surveys were completed in the clinic waiting rooms. To increase validity, guardians were encouraged to allow participants to complete surveys by being apart from them, without assistance. Participants who completed the survey were given a US $5 gift card. The survey distribution was performed by IRB-approved study coordinators.

Analysis

The survey responses were coded into IBM's SPSS statistical software and questions were assessed for response frequency, mean, mode, and SD. To compare responses between groups of participants, 2 different subgroups were created using the mean score of select questions addressing similar themes. The first subsection, *Sum of Perceptions Regarding Confidentiality in the Doctor-Patient Relationship*, comprised the sum and mean of survey questions 21, 22, and 23. The second subsection, *Cultural Barriers*, comprised the sum and mean of questions 20, 31, 32, and 33 (Multimedia Appendix 1). These questions were chosen for analysis owing to previous literature reporting differences in communication, technology, and services rendered among different ethnicities and age groups [9,12,13,22-24]. The mean scores for these subsections were then compared using demographic parameters including ethnicity, highest level of parental education, gender, and age. The scores were compared using an independent sample *t* test.

The objectives of this study included understanding (1) perceptions of adolescents with regard to patient-physician relationship, (2) adolescents' knowledge of EHR systems, and (3) practical concerns of designing an EHR system that best protects patient confidentiality while helping to facilitate patient-physician communication to improve health outcomes and patient adherence.

Results

Survey Demographics

There were 97 completed surveys (96% (97/101) of those approached) with an average participant age of 15.5 (SD 1.5) years with 60% (58/97) male and 40% (39/97) female participants. The majority of survey participants were black (44%, 43/97) and Hispanic (41%, 40/97), with substantially fewer Caucasian, Asian, and other participants. Insurance status was mainly governmental or public (48%, 46/97) or private (27%, 26/97), followed by hospital-based (5%, 5/97) or unknown (20%, 18/97). The level of parental education was stratified into 5 categories but grouped together into high school or below (46%, 45/97) or college and above (54%, 52/97). In total, 4 individuals declined to participate in the survey with an average age of 16.00 (SD 2.00) years. These data points included 1 male and 3 females, 1 private insurance and 3 governmental or public insurance, and 2 with parental education of high school or below and 2 with parental education of college or above.

Statistical Analyses

Both descriptive statistics and *t* tests were employed in this study. For descriptive statistics, clinically relevant questions from the 42-item survey were classified under 5 categories: (1) perceptions regarding confidentiality in the doctor-patient relationship, (2) understanding of EHR systems, (3) practical concerns for the design of an EHR system, (4) cultural barriers affecting care, and (5) knowledge of access to care (Tables 1-6). Participants rated each statement on the survey on the Likert scale with the value *1* classified as *least true* and the value *10* classified as *most true*. The mean scores, SDs, and mode values for the descriptive statistics were interpreted on the Likert scale (Tables 7 and 8). Independent *t* tests were performed for the subsections (1) perceptions regarding confidentiality in the doctor-patient relationship and (2) cultural barriers (Tables 9 and 10). The *t* tests examined the patient ethnicity (black vs Hispanic), parents' highest level of education (high school vs college and above), patient gender (male vs female), and age (13 to 15 years vs 16 to 18 years).

Table 1. Questions regarding confidentiality in the doctor-patient relationship.

Survey question	Mean (SD)	Mode	Frequency true (%)	Frequency false (%)
21. My doctor tells my parents about all our conversations, even those we have in private	5.794 (3.649)	10	—[a]	—
22. My doctor will tell my parents if I am smoking marijuana or drinking alcohol	4.505 (3.773)	1	—	—
23. My doctor will not tell my parents if I am having sex	5.041 (3.6)	1	—	—
26. Parents must be in the room when kids under 18 years see the doctor	—	—	34	66

[a]Denotes inapplicable data points.

Table 2. Questions regarding understanding of electronic medical record systems.

Survey question	Frequency true (%)	Frequency false (%)
1. I have heard the term medical record before	94	6
2. When my parents log in to the hospital website, they can read about what my doctor and I talked about during the visit	55	45
5. Patients can go online to see information about their visit to the doctor	70	30

Table 3. Questions regarding practical concerns for the design of an electronic medical record system.

Survey question	Mean (SD)	Mode	Frequency true (%)	Frequency false (%)
3. I would like to log in to an electronic medical record so I can see my health information	__[a]	—	84	16
16. I want my record to show what medicines I am taking	8.629 (2.205)	10	—	—
17. I want to be able to see my test results online	8.464 (2.471)	10	—	—
18. I would be comfortable sending my doctor messages online	7.155 (2.848)	10	—	—

[a]Denotes inapplicable data points.

Table 4. Preferences toward internet access and use.

Survey question and modality	Percent (%)
11. I access internet mainly through (select all that apply)	
Phone	77
Home computer	26
Public library	4
School	5
Others	4
12. I use the internet mainly for (select all that apply)	
School work	63
Social media	50
Email	28
Games	44
Others	8

Table 5. Questions regarding cultural barriers affecting care.

Survey question	Mean (SD)	Mode	Frequency true (%)	Frequency false (%)
20. I trust my doctor	9.33 (1.539)	10	__[a]	—
31. I would be willing to talk to my doctor about my boyfriend or girlfriend	6.67 (3.201)	10	—	—
32. I would be willing to talk to my doctor about sex	7.258 (2.91)	10	—	—
33. I would be comfortable talking to the doctor alone	8.082 (2.656)	10	—	—
25. Sometimes it's hard for me to tell the doctor everything as I'm afraid he or she might judge me	—	—	34	66

[a]Denotes inapplicable data points.

Table 6. Questions regarding knowledge of access to care.

Survey question	Mean (SD)	Mode	Frequency true (%)	Frequency false (%)
24. Kids under 18 years can get screened for sexually transmitted diseases without their parents' knowledge or consent	6.299 (3.395)	10	__[a]	—
28. Patients under 18 years can get a pregnancy test at the doctor's office without parents' permission	—	—	41	59
29. Patients under 18 years can get treatment for certain conditions such as depression, drug addiction, and sexually transmitted diseases without their parents' knowledge or consent	—	—	50	50

[a]Denotes inapplicable data points.

Table 7. Mean response to questions regarding the perception of confidentiality in the doctor-patient relationship stratified by ethnicity.

Ethnicity	Mean value of response			Average of 3 questions
	Question 21[a]	Question 22[b]	Question 23[c]	
Caucasian	7.556	7.333	5.667	6.852
Hispanic	5.900	4.975	5.125	5.333
Black	5.349	3.535	4.721	4.535
Asian	5.667	3.00	6.333	5.000

[a]My doctor tells my parents about all our conversations, even those we have in private.

[b]My doctor will tell my parents if I am smoking marijuana or drinking alcohol.

[c]My doctor will not tell my parents if I am having sex.

Table 8. Mean response to questions regarding cultural barriers stratified by ethnicity.

Ethnicity	Mean value of response				Average of 4 questions
	Question 20[a]	Question 31[b]	Question 32[c]	Question 33[d]	
Caucasian	9.444	7.556	7.667	8.778	8.361
Hispanic	9.325	6.775	7.100	8.650	7.963
Black	9.372	6.302	7.209	7.465	7.587
Asian	9.333	7.000	7.000	9.000	8.083

[a]I trust my doctor.

[b]I would be willing to talk to my doctor about my boyfriend or girlfriend.

[c]I would be willing to talk to my doctor about sex.

[d]I would be comfortable talking to the doctor alone.

Table 9. Independent *t* test: questions regarding perception of confidentiality in the doctor-patient relationship. Statistically significant data points have been italicized.

Variable	*t* (*df*)	*P* value	Mean difference (SE)	95% CI
Ethnicity	1.515 (81)	.134	0.798 (0.527)	−0.250 to 1.847
Level of parental education	0.606 (94)	.546	0.303 (0.500)	−0.689 to 1.295
Gender	1.851 (95)	.067	0.919 (0.497)	−0.067 to 1.904
Age	−3.781 (95)	*<.001*	−1.747 (0.462)	−2.664 to −0.830
Insurance	2.578 (70)	*.012*	1.477 (0.573)	0.334 to 2.621

Table 10. Independent *t* test: questions regarding cultural barriers.

Variable	*t* (*df*)	*P* value	Mean difference (SE)	95% CI
Ethnicity	0.873 (81)	.386	0.375 (0.430)	−0.481 to 1.231
Level of parental education	0.634 (94)	.527	0.232 (0.366)	−0.495 to 0.959
Gender	−0.533 (95)	.595	−0.212 (0.397)	−1.000 to 0.577
Age	−1.802 (95)	.075	−0.691 (0.384)	−1.453 to 0.070
Insurance	1.591 (70)	.116	0.747 (0.469)	−0.189 to 1.683

In section 1 of the survey, 94% of participants acknowledged they had heard the term *electronic health record* before but only 55% answered positively when asked if they knew how EHRs function (Multimedia Appendix 1). In total, 84% of participants expressed interest in viewing their records online and reported a mean score of 8.46 out of 10 when asked if they would like to see the test results online. The majority of participants (77%) used their cell phones as their primary internet access. Only 50% of participants knew that patients aged under 18 years could get treatment for sexual health, mental health, and drug/alcohol abuse without parental consent.

Independent *t* tests performed on *subsection 1: questions regarding perception of confidentiality in the doctor-patient relationship* revealed significant differences for patient age and insurance status but not for ethnicity, level of parental education, or gender. The independent *t* tests for *subsection 2: questions regarding cultural barriers* did not result in significant differences for any of the categories.

Discussion

The objectives of this study included understanding the (1) perceptions of adolescents with regard to patient-physician relationship, (2) adolescents' knowledge of EHR systems, and (3) practical concerns of designing an EHR portal that best protects patient confidentiality while helping to facilitate patient-physician communication to improve health outcomes and patient adherence.

Confidentiality and Cultural Barriers

For the subsection, *perceptions regarding confidentiality in the doctor-patient relationship*, there were statistically significant differences in age (13 to 15 years vs 16 to 18 years) and insurance status (governmental vs private) but not gender, ethnicity, or highest parental education level. These results indicate that the younger age group (aged 13 to 15 years) and those with governmental insurance have more confidence in the confidentiality of their discussions with their physician. These results are concordant with the study by Carlisle et al, which reported older teenagers to be more concerned about confidentiality than younger teenagers [24]. However, if one is to presume that governmental insurance status is correlative to lower socioeconomic status, our results would be contradictory to Ford et al, who reported that adolescents with higher socioeconomic status were more willing to disclose confidential information [6]. Interestingly, there were no statistically significant differences found between gender, ethnicity, or parental education with regard to confidentiality in the physician-patient relationship. These results are discordant with previous studies which demonstrated significant differences between gender and ethnicity [12,24]. The lack of gender results may be due to our limited power owing to the decreased sample size, as well as a male predominance (58 males vs 39 females). With regard to ethnicities, we only compared black with Hispanic patients because of a marginal sample size of other ethnicities; thus, a direct comparison with previous studies is unfortunately inapplicable.

The subsection *questions regarding cultural barriers* had no significant differences for any of the categories including age, ethnicity, gender, or highest level of education. These results are promising in that there were no identifiable cultural barriers to patient's trust of physicians, which should facilitate improved communication and care given upon implementation of the new patient portal.

For the questions regarding perceptions of confidentiality in the physician-patient relationship, there did not appear to be a strong positive or negative response, with mean results between 4.51 and 5.79 out of 10, indicating that the survey participants may be unsure of what topics are considered to be confidential between them and their physician (Table 1). However, when the results were stratified by race, there seemed to be a marked variation in response to question 22: *My doctor will tell my parents if I am smoking marijuana or drinking alcohol*, ranging from 3.00 and 3.53 for Asian and black participants compared with 4.98 and 7.33 for Hispanic and white participants, respectively. This variation may be due to the differences in how often these individuals see their doctor and the level of communication occurring during those visits, a phenomenon that has been previously documented in the literature [13]. This markedly neutral response to questions regarding confidentiality may also explain some of the results about the knowledge of access to care. Only 50% of survey participants positively indicated that patients aged under 18 years could be treated for depression, drug addiction, and sexually transmitted diseases without their parent's consent. Furthermore, even fewer (41%) participants were aware that a pregnancy test could be ordered without parental consent (Table 6). These findings demonstrate a possible disconnect in communication and education between providers and the adolescent population. This issue presents as a significant concern as there has been evidence that adolescents who have concerns about confidentiality may forego seeking necessary care [12,23]. If these individuals do forego care, it may lead to significant morbidity, and possibly even mortality, as these sensitive conditions (pregnancy/STIs/substance use/mental health) can have significant short- and long-term side effects.

The findings regarding cultural barriers to care showed that the adolescent population places a great amount of trust in their doctors, averaging 9.33 out of 10 with a minimal score variation between ethnicities. This was offset however by a slightly lower score for questions regarding talking about sex or one's significant other, with average responses of 6.67 and 7.26, respectively (Tables 5 and 8). These results, although still positive, may be the result of the decreased perception of doctor-patient confidentiality seen in our results, combined with the average age of our study population of 15.5 years. Previous studies have shown that although adolescents do encourage confidentiality, it often comes toward the end of their teenage years [23].

Understanding and Use of Electronic Health Record Portals

Although 94% of participants have heard of the term *electronic health records*, only 55% knew of their function (Table 2). These findings were supplemented with a noteworthy 84% of patients expressing interest in being able to log in to their EHRs and see their information. These data indicate that the adolescent population is not naïve to the concept of EHRs and are very interested in their content. They, as minors, may simply not have the exposure and formal education about accessing records that an adult with a full EHR access would have. As an essential component of the administration of care, we recommend that physicians educate all developmentally appropriate patients on the content, function, and limitations of their EHR portal. This not only prophylactically addresses potential issues regarding the confidentiality of EHR content but may also further the health literacy and autonomy of adolescent patients [6].

Design of an Electronic Health Record Portal

The final objective of this study was the practical concerns for the design of an EHR portal. Our findings demonstrate that adolescent patients are very interested in an EHR portal that contains the ability to view test results, current medications, and send messages to their provider (Table 3). These aspects are commonly offered in most adult EHR portals but have been a controversial topic in the pediatric population because of the risk of sensitive content appearing in the portal [16]. Although a proxy access EHR portal would not necessarily prevent these issues, it is expected that this concept would be discussed before parental consent for access, therefore limiting potential future issues. It is also important to note the methods in which EHR portals can be accessed. Previous literature describes limited adoption and access of EHR portals in ethnically diverse populations with low incomes and limited internet access [22]. A total of 77% of survey participants reported that they used their cell phone as their primary tool for accessing the internet. Owing to the relatively low cost and high prevalence of internet-capable mobile phones present in low income populations, the previously described barriers may not be as applicable. Therefore, it is highly recommended that when designing an EHR portal, the system should be easily accessible from a mobile device.

Limitations

This study has several limitations. The first is that because of the limited sample size, statistical comparisons were only able to be performed between black and Hispanic patients. This somewhat hinders the generalizability of our dataset and limits the identification of ethnically based barriers to EHR portal access. Second, our survey as the first adolescent-directed EHR survey to our knowledge required creation of a de novo and unvalidated question set. Finally, our data collection was primarily performed in a large, urban, academic center, which may not accurately reflect the demographic of patients in the suburban or rural community. In addition, because of the limitations with access to translator services, the survey was only able to be offered in English. This limited the number of patients available to take the survey and may have skewed the dataset away from the large immigrant population present in South Florida.

Conclusions

The objectives of this study included understanding the (1) perceptions of adolescents with regard to patient-physician relationship, (2) adolescents' knowledge of EHR systems, and (3) practical concerns of designing an EHR portal that best protects patient confidentiality while helping to facilitate patient-physician communication to improve health outcomes and patient adherence. Our findings demonstrated that although cultural barriers to EHR portal access may be limited, there is a noteworthy deficit in the adolescent population's understanding of confidentiality and access to care. The results, although preliminary, lend support to the previous studies assessing differences in the understanding of and access to EHR portals in different ethnicities and socioeconomic groups [17-19]. Physicians treating adolescent patients must take an active effort to educate their patients on what topics remain confidential and the availability of resources present in caring for sensitive issues such as substance use, STIs, pregnancy, and mental health. We also recommend that when designing an EHR portal for adolescents, it should be accessible by a mobile device and that it should contain test results, current medications, and the ability to securely message providers. This will allow for easier patient access to their EHR portal, as well as an increase in autonomy, health literacy, and more communication for those who participate. We plan to use the results of this study as the basis for future investigations on EHR portal access in which we may obtain a larger sample size and control group.

Acknowledgments

The development of this project was undertaken by DM, SV, AD, and KG in conjunction with the members of the Pediatric Bioethics Committee at Holtz Children's Hospital at the Jackson Health System and the University of Miami Institute for Bioethics and Health Policy. Survey development and design was completed in conjunction with the University of Miami Miller School of Medicine Pediatric Faculty, as well as the University of Miami Biostatistics Collaboration and Consulting Core.

Conflicts of Interest

None declared.

Multimedia Appendix 1

42-Item survey distributed to participants.

[PDF File (Adobe PDF File), 108KB - medinform_v7i2e11570_app1.pdf]

References

1. CCH Tax Law Editors. American Recovery and Reinvestment Act of 2009: Law, Explanation and Analysis. Chicago: CCH, Inc; 2009.
2. Khalifa M. Perceived benefits of implementing and using hospital information systems and electronic medical records. Stud Health Technol Inform 2017;238:165-168. [doi: 10.3233/978-1-61499-781-8-165] [Medline: 28679914]
3. King J, Patel V, Jamoom EW, Furukawa MF. Clinical benefits of electronic health record use: national findings. Health Serv Res 2014 Feb;49(1 Pt 2):392-404 [FREE Full text] [doi: 10.1111/1475-6773.12135] [Medline: 24359580]
4. Adler-Milstein J, Everson J, Lee SD. EHR adoption and hospital performance: time‐related effects. Health Serv Res 2015 Dec;50(6):1751-1771 [FREE Full text] [doi: 10.1111/1475-6773.12406] [Medline: 26473506]
5. Lehmann CU, Council on Clinical Information Technology. Pediatric aspects of inpatient health information technology systems. Pediatrics 2015 Mar;135(3):e756-e768 [FREE Full text] [doi: 10.1542/peds.2014-4148] [Medline: 25713282]
6. Ford C, English A, Sigman G. Confidential health care for adolescents: position paper for the society for adolescent medicine. J Adolesc Health 2004 Aug;35(2):160-167. [doi: 10.1016/S1054-139X(97)00171-7] [Medline: 15298005]
7. Bayer R, Santelli J, Klitzman R. New challenges for electronic health records: confidentiality and access to sensitive health information about parents and adolescents. J Am Med Assoc 2015 Jan 6;313(1):29-30. [doi: 10.1001/jama.2014.15391] [Medline: 25562260]
8. 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]
9. McKee MD, Rubin SE, Campos G, O'Sullivan LF. Challenges of providing confidential care to adolescents in urban primary care: clinician perspectives. Ann Fam Med 2011;9(1):37-43 [FREE Full text] [doi: 10.1370/afm.1186] [Medline: 21242559]
10. Centers for Disease Control and Prevention. 2017 Nov. 2016 HIV Surveillance Report URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2019-04-10] [WebCite Cache ID 77XLwpb2i]
11. Florida Legislature. Online Sunshine. The 2018 Florida Statutes URL: http://www.leg.state.fl.us/statutes/ [accessed 2019-04-10] [WebCite Cache ID 77XMJpVuF]
12. Cheng TS, Savageau JA, Sattler AL, DeWitt TG. Confidentiality in health care. A survey of knowledge, perceptions, and attitudes among high school students. J Am Med Assoc 1993 Mar 17;269(11):1404-1407. [Medline: 8441216]
13. Irwin C, Adams SH, Park MJ, Newacheck PW. Preventive care for adolescents: few get visits and fewer get services. Pediatrics 2009 Apr;123(4):e565-e572. [doi: 10.1542/peds.2008-2601] [Medline: 19336348]
14. Anoshiravani A, Gaskin G, Kopetsky E, Sandborg C, Longhurst CA. Implementing an interoperable personal health record in pediatrics: lessons learned at an academic children's hospital. J Particip Med 2011 Jul 10;3 [FREE Full text] [Medline: 21853160]
15. Nakamura M, Ferris TG, DesRoches CM, Jha AK. Electronic health record adoption by children's hospitals in the United States. Arch Pediatr Adolesc Med 2010 Dec;164(12):1145-1151. [doi: 10.1001/archpediatrics.2010.234] [Medline: 21135344]
16. Sittig D, Singh H. Legal, ethical, and financial dilemmas in electronic health record adoption and use. Pediatrics 2011 Apr;127(4):e1042-e1047 [FREE Full text] [doi: 10.1542/peds.2010-2184] [Medline: 21422090]
17. Ketterer T, West DW, Sanders VP, Hossain J, Kondo MC, Sharif I. Correlates of patient portal enrollment and activation in primary care pediatrics. Acad Pediatr 2013;13(3):264-271. [doi: 10.1016/j.acap.2013.02.002] [Medline: 23680344]
18. Gracy D, Weisman J, Grant R, Pruitt J, Brito A. Content barriers to pediatric uptake of electronic health records. Adv Pediatr 2012;59(1):159-181. [doi: 10.1016/j.yapd.2012.04.004] [Medline: 22789578]
19. Byczkowski T, Munafo JK, Britto MT. Variation in use of Internet-based patient portals by parents of children with chronic disease. Arch Pediatr Adolesc Med 2011 May;165(5):405-411. [doi: 10.1001/archpediatrics.2011.55] [Medline: 21536954]
20. Britto M, 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):280-1.e1. [doi: 10.1016/j.jpeds.2013.02.041] [Medline: 23541773]
21. United States Census Bureau. QuickFacts: Miami-Dade County Florida URL: https://www.census.gov/quickfacts/fact/table/miamidadecountyflorida/POP060210 [accessed 2019-04-10] [WebCite Cache ID 77XNIhc1i]
22. Gibbons MC. Use of health information technology among racial and ethnic underserved communities. Perspect Health Inf Manag 2011 Jan 1;8:1f [FREE Full text] [Medline: 21307989]
23. Song X, Klein JD, Yan H, Catallozzi M, Wang X, Heitel J, et al. Parent and adolescent attitudes towards preventive care and confidentiality. J Adolesc Health 2019 Feb;64(2):235-241. [doi: 10.1016/j.jadohealth.2018.08.015] [Medline: 30396723]

24. Carlisle JS, Shickle D, Cork M, McDonagh A. Concerns over confidentiality may deter adolescents from consulting their doctors. A qualitative exploration. J Med Ethics 2006 Mar;32(3):133-137 [FREE Full text] [doi: 10.1136/jme.2004.011262] [Medline: 16507655]

Abbreviations

EHR: electronic health record
IRB: Institutional Review Board
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 17.07.18; peer-reviewed by J Hefner, T McFarlane; comments to author 08.10.18; revised version received 04.02.19; accepted 24.03.19; published 26.04.19.

Please cite as:
Miklin DJ, Vangara SS, Delamater AM, Goodman KW
Understanding of and Barriers to Electronic Health Record Patient Portal Access in a Culturally Diverse Pediatric Population
JMIR Med Inform 2019;7(2):e11570
URL: http://medinform.jmir.org/2019/2/e11570/
doi:10.2196/11570
PMID:31066681

Review

Toward Standardized Monitoring of Patients With Chronic Diseases in Primary Care Using Electronic Medical Records: Systematic Review

Leandra Falck[1], MD; Marco Zoller[1], MD; Thomas Rosemann[1], MD, PhD; Nahara Anani Martínez-González[1], MSc; Corinne Chmiel[1], MD, PD

Institute of Primary Care, University of Zurich and University Hospital of Zurich, Zurich, Switzerland

Corresponding Author:
Leandra Falck, MD
Institute of Primary Care
University of Zurich and University Hospital of Zurich
Pestalozzistrasse 24
Zurich,
Switzerland
Phone: 41 44 255 98 55
Email: Leandra.Falck@usz.ch

Abstract

Background: Long-term care for patients with chronic diseases poses a huge challenge in primary care. In particular, there is a deficit regarding monitoring and structured follow-up. Appropriate electronic medical records (EMRs) could help improving this but, so far, there are no evidence-based specifications concerning the indicators that should be monitored at regular intervals.

Objective: The aim was to identify and collect a set of evidence-based indicators that could be used for monitoring chronic conditions at regular intervals in primary care using EMRs.

Methods: We searched MEDLINE (Ovid), Embase (Elsevier), the Cochrane Library (Wiley), the reference lists of included studies and relevant reviews, and the content of clinical guidelines. We included primary studies and guidelines reporting about indicators that allow for the assessment of care and help monitor the status and process of disease for five chronic conditions, including type 2 diabetes mellitus, asthma, arterial hypertension, chronic heart failure, and osteoarthritis.

Results: The use of the term "monitoring" in terms of disease management and long-term care for patients with chronic diseases is not widely used in the literature. Nevertheless, we identified a substantial number of disease-specific indicators that can be used for routine monitoring of chronic diseases in primary care by means of EMRs.

Conclusions: To our knowledge, this is the first systematic review summarizing the existing scientific evidence on the standardized long-term monitoring of chronic diseases using EMRs. In a second step, our extensive set of indicators will serve as a generic template for evaluating their usability by means of an adapted Delphi procedure. In a third step, the indicators will be summarized into a user-friendly EMR layout.

(JMIR Med Inform 2019;7(2):e10879) doi:10.2196/10879

KEYWORDS

monitoring of chronic diseases; indicators; primary care; systematic review; electronic medical record; diabetes mellitus type 2; arterial hypertension; asthma; osteoarthritis; chronic heart failure

Introduction

In 2016, the World Health Organization estimated that 71% of the overall deaths worldwide occurred due to noncommunicable diseases [1]. The majority of these diseases include cardiovascular diseases, chronic respiratory diseases, and diabetes. In particular, the prevalence of type 2 (non-insulin-dependent) diabetes mellitus, arterial hypertension, asthma, chronic heart failure, and musculoskeletal diseases is increasing rapidly around the world leading to increased multimorbidity and polypharmacy, especially in the older population [1,2]. The burden of these diseases consequently imposes a significant threat to health, quality of life, and economic status in the affected population. Moreover, the regular monitoring of chronic diseases poses huge challenges

and requires knowledge and communication skills, as well as the capability of organization and coordination. The chronic care model (CCM) was originally introduced to graphically picture the concept of disease management [3]. The eHealth enhanced chronic care model was subsequently introduced as the means to improve the CCM in view of the progress and development of information and communication technology [4]. This model shows the existing variety of technically well-advanced applications as part of the monitoring process. Too many clinical offices in Switzerland lack basic electronic devices since many general practitioners still use paper-based patient records.

In 2012, 31 European countries were ranked based on the usage of electronic medical records (EMRs) in primary care [5]. In this global ranking of EMR usage, Switzerland ranked number 24. In a Swiss study, only up to 44.8% of the participating primary care physicians reported the usage of EMRs [6]. Therefore, it is currently almost impossible to exchange data with digital applications that are increasingly available and used by patients [6]. To efficiently monitor patients with chronic diseases, a well-structured and organized EMR system is crucial to ensure that all necessary information can be easily entered and retrieved, while no essential information is missed. Surprisingly, there are no evidence-based specifications concerning the indicators that should be monitored at regular intervals. On one hand, there are currently no international standards for the monitoring of patients with chronic diseases by means of EMR in primary care. On the other hand, there are deficits regarding the actual monitoring and structured follow-up. Therefore, we aimed to identify and collect a set of evidence-based indicators that could be used for monitoring patients with chronic conditions at regular intervals in primary care using EMRs.

Methods

Systematic Identification and Assessment of Supporting Evidence

We followed the principles of systematic reviews [7] and developed a protocol a priori to guide the identification and assessment of the monitoring indicators.

Inclusion Criteria

We included clinical guidelines and primary peer-reviewed studies of any design, carried-out mainly in primary care (ie, family health care) patients aged 18 years and older, who were diagnosed with type 2 (non-insulin-dependent) diabetes mellitus, arterial hypertension, asthma, chronic heart failure, or osteoarthritis. The first four diseases are among the most common noninfectious diseases worldwide. Osteoarthritis, in particular, generates a large part of indirect costs [2]. In order to be included, studies must have also reported on indicators that allow the assessment of care and help monitor the status and process of disease for these five chronic conditions. Therefore, we considered disease indicators that help reduce the risk of exacerbation, such as intermediate outcome indicators (eg, hemoglobin A_{1c} [HbA_{1c}] for diabetics or blood pressure measurements for hypertensive patients) and process indicators (eg, regular foot care or nutrition counselling). We included studies regardless of whether specific interventions were evaluated. In addition, all studies and clinical guidelines should have been published in English or German.

Search Methods and Study Identification

We developed a comprehensive search strategy in collaboration with an expert librarian. The librarian conducted the search and produced a set of studies that matched the predefined search criteria. We identified studies published between 2000 and 2015 by applying this strategy in MEDLINE (Ovid), Embase (Elsevier), and the Cochrane Library (Wiley). No restrictions were made regarding the country of origin of the studies. The search strategy included a combination of the concepts and terminology, synonyms and related words for monitoring and for medical, health, electronic, patient, or file records. It also included primary, family, health care, or general practitioner, and the five chronic conditions (ie, type 2 [non-insulin-dependent] diabetes mellitus, arterial hypertension, asthma, chronic heart failure, and osteoarthritis). The focused search also included the terminology indicators, parameter, and management. An example of the full search strategy is available in Multimedia Appendix 1.

We identified additional publications by manually searching the reference lists of included studies and relevant reviews. We also searched for monitoring indicators in the clinical guidelines in order to identify as many indicators as possible and to enable a holistic management of chronic diseases. Given that most guidelines are not indexed in the former medical literature databases, and to identify the clinical guidelines related to any of the five chronic diseases, we searched World Wide Web-based databases, including the National Guideline Clearinghouse for US guidelines [8] and the Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften eV (AWMF) [9] for German guidelines.

Study Selection and Assessment

For study selection, we created a system to prioritize the studies. One reviewer identified eligible studies by first screening the titles and abstracts of all records retrieved by the searches based on the inclusion criteria. All potentially eligible abstracts were rated manually from one to five stars according to their relevance for this review. The stars were assigned based on whether or not the key terms were mentioned (ie, “indicator,” “monitoring,” “assessment,” “management,” and/or “guideline”). The ranking was assigned as follows:

1. One star: Remote reference to the key terms; no indicators expected in full text.
2. Two stars: Little reference to the key terms; indicators in full text unlikely.
3. Three stars: Reference of at least one key term; indicators in full text possible.
4. Four stars: Reference of at least one key term; indicators in full text very possible.
5. Five stars: Reference of indicators, monitoring, or interval of measuring indicators.

The full text of all studies with an abstract that was rated with at least two stars was obtained, if available, and further evaluated

based on the reporting of indicators. For studies where the full text was not available but were deemed important to inform our monitoring tool, we used the data reported in the abstract. When it was necessary, the study team was consulted throughout the evaluation process to confirm the eligibility of indicators.

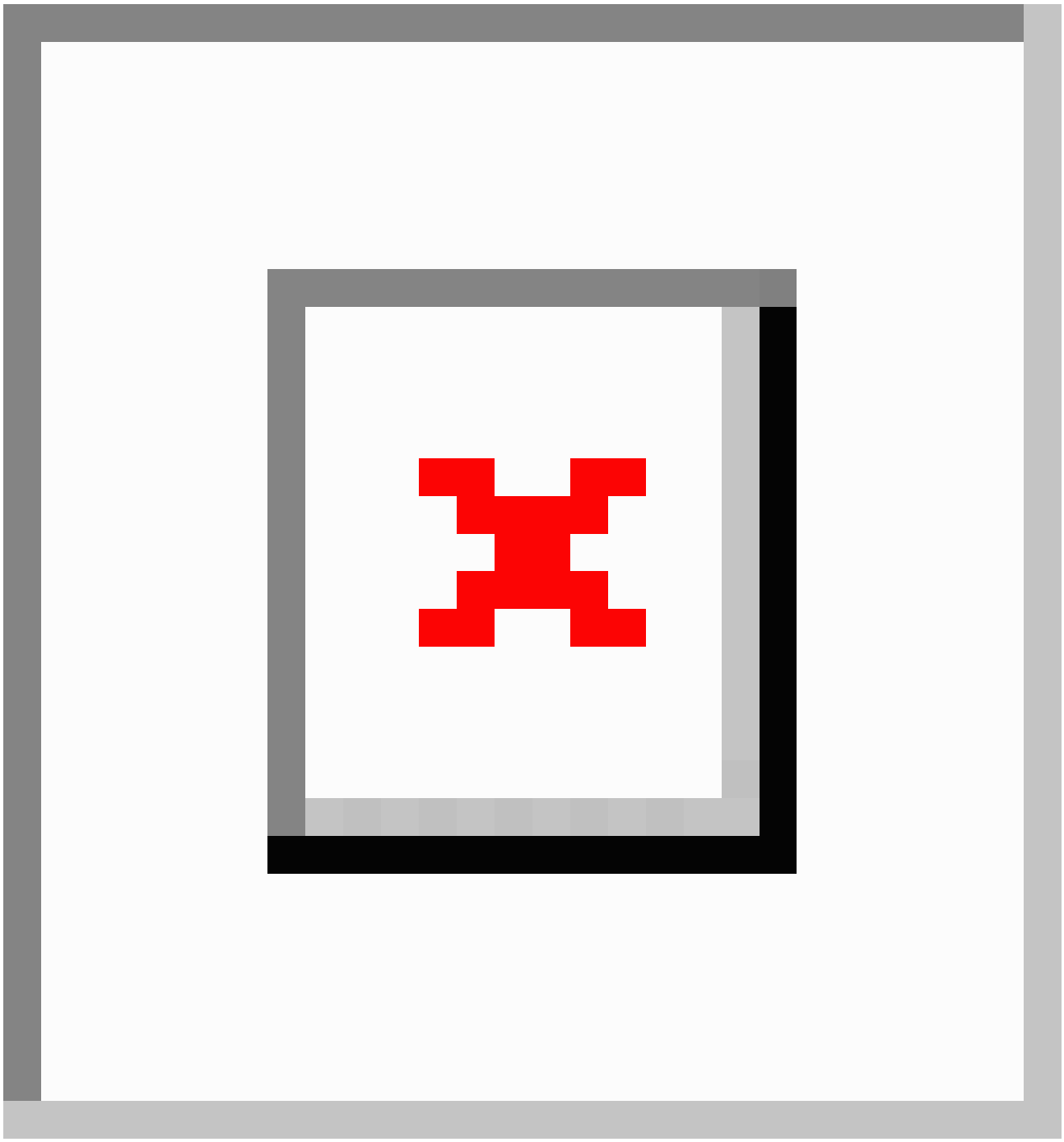
Data Extraction and Synthesis

For each included study, we extracted the bibliographic details (ie, author, year, and country of origin), all the monitoring indicators reported, the guideline on which the indicators were based, and the country of origin of the guidelines for each of the five chronic diseases. One reviewer extracted all data, and another reviewer verified the extracted data. We compiled a data profile for each study or guideline, and generated a set of indicators using Microsoft Excel. We report a descriptive summary of the indicators for each of the chronic conditions.

Results

Our literature searches identified 795 original records (see Figure 1). After deduplication and perusal of titles and abstracts, we screened 621 records (range by disease: 33-180) and excluded 408 records that did not meet our inclusion criteria (eg, focused on specific therapy or medication or did not cover the topic). We examined in detail the full text, where available, of 213 publications (range by disease: 13 to 82).

Figure 1. Flowchart demonstrating the identification and selection of evidence. a: type 2 diabetes mellitus; b: asthma; c: arterial hypertension; d: heart failure; e: osteoarthritis; *: 5 of 87 publications (6%) reported indicators for more than one disease of interest.



We included 87 original publications, 5 (6%) in abstract form only, reporting indicators for diabetes mellitus [10-63], asthma [60,64-70], arterial hypertension [10,35,39,71-81], heart failure [33,82-92], and osteoarthritis [93-96]. Multimedia Appendix 2 presents a list of all included studies that reported monitoring indicators for the five chronic conditions. A total of 5 publications (6%) reported indicators for more than one chronic disease [10,33,35,39,60]. The number of included publications by disease with at least one indicator ranged from 4 to 54. Most records (54/87, 62%) were published on type 2 diabetes mellitus, while osteoarthritis was the most underrepresented of the five diseases, with only 4 records (5%). A total of 74 of all 87 included studies (85%) contained process indicators, the most significant type of indicators. Concerning diabetes mellitus, a third of all publications (54/179, 30.2%) reported at least one indicator. For arterial hypertension and heart failure, only 8% (7/87) of all publications reported at least one indicator. Overall, most records used guidelines from the United States, followed by the United Kingdom. For diabetes mellitus, the American Diabetes Association and the National Institute for Health and Care Excellence were the most-used guidelines. The most frequently mentioned indicators for diabetes are presented in Table 1. The indicators for the other four diseases are presented in Multimedia Appendices 3-6.

Table 1. Diabetes mellitus indicators that are most frequently mentioned in guidelines and studies. The indicators are sorted first by guidelines and then by studies.

Indicators for diabetes mellitus	Number of guidelines where indicators are mentioned (guidelines)	Number of studies where indicators are mentioned
Fundoscopic examination	7 (a-g)[a]	20 [10,12,13,18,21,23,25,30-32,41,43-46,50-52,60,61]
Height, weight, and body mass index	7 (a-g)	33 [11-16,18,20,21,23-25,27-29,32,33,35,40,41,44,45,48,49,53-56,58-60,62,63]
Blood pressure measurement	7 (a-g)	45 [11,13,15-27,29-36,39-45,47-49,52,53,55,56,58-63]
10 g monofilament	7 (a-g)	N/A[b]
Hemoglobin A_{1c} (ie, glycated hemoglobin)	7 (a-g)	46 [10,12,13,15-23,26,28-37,39-45,47-54,56-63]
Foot inspection	7 (a-g)	17 [12,15,18,21,23,25,30-32,43-46,50-52,61]
Erectile dysfunction	7 (a-g)	N/A
Albuminuria	7 (a-g)	18 [12,13,18,22,23,25,31,32,35,41,43-46,51,55,61,62]
Lipid profile	7 (a-g)	8 [25,26,30,43,45,46,52,61]
Low-density lipoprotein	N/A	30 [11,12,15,18-20,22-24,29,31-37,41-44,47-49,52-54,63]
High-density lipoprotein	N/A	14 [11,20,23,28,29,33,37,39,49,51,53,54,62,63]
Triglyceride	N/A	15 [20,29,30,33,37,39,48,49,51,53-55,57,62,63]
Creatinine	7 (a-g)	18 [13,15,16,22,25-27,29,33,41,46,51,55,57-60,62]
Alcohol intake	7 (a-g)	2 [24,53]
Neuropathy and history of foot lesion	7 (a-g)	3 [18,20,55]
History of myocardial infarction (ie, cardiovascular disease)	6 (a-f)	2 [18,22]
Foot pulses	6 (a-f)	3 [18,32,60]
Smoking status	6 (a-f)	24 [11-15,18,20,22-24,26,28,29,31,35,41,44,48,50,53,58-61]
Orthostatic hypotension	5 (a, b, d, e, g)	N/A
Skin inspection	5 (a, b, d, f, g)	N/A
Vibration by 128 Hz tuning fork	5 (a-d, g)	1 [60]
Plasma glucosis	4 (b-d, g)	12 [11,21,24,33,39,45,51,54,55,57,59,63]
Onset of diabetes	3 (b, c, f)	9 [11,18,22,23,28,48,55,58,59]
Indicators appeared in fewer than five guidelines	225	N/A
Indicators appeared in fewer than 10 studies	N/A	76

[a]The letters a-g refer to the guidelines listed in Multimedia Appendices 7-11.

[b]N/A: not applicable.

In total, there were 249 indicators for type 2 diabetes mellitus, 183 for asthma, 335 for arterial hypertension, 231 for chronic heart failure, and 164 for osteoarthritis. The majority of indicators were identified by screening both peer-reviewed articles and clinical guidelines. A few extra indicators were reported only in peer-reviewed articles. That is, clinical guidelines on their own contributed to the great majority of all indicators identified. Surprisingly, only a few guidelines, such as the American guideline for asthma, included a section dedicated to monitoring or follow-up. Most of the guidelines that we screened did not specify the interval at which the indicators should be monitored. Also, in some guidelines, self-monitoring was a big topic for chronic heart disease (ie, weight control), asthma (ie, peak expiratory flow), and type 2 diabetes mellitus (ie, glucose monitoring).

Our systematic review also found that the term "monitoring," in the sense of long-term patient care, was not widely used. Although publications reported the actual monitoring indicators, the process of monitoring for the different diseases, including, for example, the potential risks associated with overmonitoring, was only scarcely addressed. The publication by Glasziou was the only one giving a broader overview on the topic [97]. Only a handful of publications reported a complete set of indicators that can be used for monitoring, but these were either not specific for primary care or not eligible for implementation in EMRs [98-101].

Discussion

Principal Findings

To our knowledge, this study represents the first summary of the existing scientific evidence about the indicators that help standardize the monitoring of chronically ill patients in primary care by the use of EMRs. Long-term care of patients with chronic diseases is challenging and there are deficits regarding their monitoring and structured follow-up. Chronic care often involves collaboration between several people involved in the treatment process. That is only one reason for its complexity. Interpersonal differences in monitoring can decrease the quality of monitoring processes. Surprisingly, there are currently no gold standards or consensus regarding the systematic monitoring of patients with chronic diseases, in particular by means of EMRs. To efficiently monitor patients with chronic diseases, a well-structured and organized EMR system is crucial to ensure that all necessary information can be easily entered and retrieved and that no essential information is missed. Our study is, thus, the first initiative toward the urgent need of standardization for monitoring patients with chronic diseases in primary care.

Our systematic literature review showed that the term "monitoring" in terms of disease management and long-term patient care is not widely used. There is a plethora of literature about quality indicators that might have the potential to improve the outcome of a disease. The Quality and Outcomes Framework (QOF) in the United Kingdom, for example, assesses indicators for such purposes [102]. Beyond identifying indicators that can be easily assessed, such as the indicators used by the QOF, our goal was to summarize the existing literature on all the indicators available for long-term monitoring.

So far, only a few authors have focused on the topic of the monitoring of chronic diseases. According to Glasziou, the process of monitoring aims to establish the response to treatment and to detect both adverse effects and the need to adjust treatment [97]. The process of monitoring can be divided into different phases (ie, pretreatment, during treatment, and after treatment). Each phase requires measurements at different intervals.

When analyzing different diseases, monitoring is probably most widely mentioned in blood pressure management. There are various publications reporting on the optimal way and interval of measuring blood pressure [76,103,104]. However, literature beyond the indicator of blood pressure measurement remains scarce. Regarding diabetes mellitus, there is an extended monitoring tool that was designed as a disease management tool for practice nurses [101]. The tool's design is based on a traffic light scheme to detect any deficit and need for action. In addition, a detailed guideline on how to monitor the diabetic foot is provided by the International Working Group on the Diabetic Foot [105]. As for bronchial asthma, two study groups have addressed the optimal way and potential problems of finding and evaluating indicators to monitor patients with asthma, including an overview of the most important indicators [98,100]. Similarly, Grypdonck presents a small set of indicators for monitoring patients with osteoarthritis of the knee [93]. Self-monitoring seems to be an important topic concerning osteoarthritis and asthma. An English study conducted by interviewing general practitioners about osteoarthritis showed that the majority of respondents thought monitoring of osteoarthritis is important, even though almost half did not monitor patients at all. Interestingly, more than half of the respondents felt that patients should do self-monitoring [106]. Patient involvement is crucial for monitoring. Particularly, in high-frequency monitoring situations such as chronic heart failure, telecardiological service, including transtelephonic monitoring, reduces the length of hospitalization and improves quality of life [91]. Surprisingly, publications concerning monitoring of chronic heart failure seem to be scarce [90]. The underrepresentation of osteoarthritis and chronic heart failure is also reflected in the number of indicators detected in the primary literature, compared to a large number of records reporting on indicators for type 2 diabetes mellitus. Another topic repeatedly found in the results was the involvement of a clinical practice nurse in monitoring [101,107-109]. The clinical nurse can, for example, fill out a monitoring questionnaire in face-to-face sessions with the patient, on the phone, or even electronically. This could counteract the problem of workload and time constraints as a frequent response to why monitoring is not conducted [106].

Strengths and Limitations

To our knowledge, this study represents the first scientifically founded recommendation for the standardized long-term monitoring of chronically ill patients in primary care. Usually, systematic reviews only concentrate on primary literature and do not include guidelines in their search strategy, since most guidelines are not indexed in databases. In our study, we explicitly searched for guideline programs such as the National Guideline Clearinghouse for American guidelines and the

AWMF for German guidelines. We added a substantial number of manual searches within reference lists and search engines in order to gain a maximal insight of the existing literature. This strategy was worth the extra effort, considering that most relevant indicators were found in guidelines and not in the primary literature. Possible confounders are that publications and guidelines reported in languages other than German and English were excluded.

Outlook

In a second step, our extensive set of indicators obtained from this work will serve as a generic template for a monitoring tool. By means of an adapted Delphi procedure, the indicators will be further evaluated in terms of their usability. In a third step, the indicators will be summarized into a user-friendly EMR layout.

Conclusion

This is the first study that systematically summarizes the existing scientific evidence about the standardized long-term monitoring of chronic diseases by means of EMRs. It aims to help improve care for patients with chronic diseases in primary care.

Acknowledgments

We are grateful to Martina Gosteli for her assistance with the search strategy. This project was funded by the Institut für Praxisinformatik, mandated by the Swiss Medical Association FMH.

Authors' Contributions

LF was involved in designing the search strategy, performing the systematic screening and review of the literature, and writing of the manuscript. MZ was involved in study design and revised the manuscript. NAMG contributed to the study design and search strategy; NAMG also revised and improved the manuscript. TR supervised the development and methodology of the study and helped improve the final version of the manuscript. CC was involved in the study design, study selection, and prioritization; CC also verified the extracted data, and supervised and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy for MEDLINE (Ovid), Embase (Elsevier), and the Cochrane Library (Wiley).

[DOCX File, 14KB - medinform_v7i2e10879_app1.docx]

Multimedia Appendix 2

List of all included studies, including monitoring indicators for the five chronic conditions.

[DOCX File, 27KB - medinform_v7i2e10879_app2.docx]

Multimedia Appendix 3

Asthma indicators mentioned in guidelines and studies.

[DOCX File, 15KB - medinform_v7i2e10879_app3.docx]

Multimedia Appendix 4

Arterial hypertension indicators mentioned in guidelines and studies.

[DOCX File, 16KB - medinform_v7i2e10879_app4.docx]

Multimedia Appendix 5

Chronic heart failure indicators most frequently mentioned in guidelines and studies.

[DOCX File, 14KB - medinform_v7i2e10879_app5.docx]

Multimedia Appendix 6

Osteoarthritis indicators most frequently mentioned in guidelines and studies.

[DOCX File, 14KB - medinform_v7i2e10879_app6.docx]

Multimedia Appendix 7

Guidelines screened for indicators for type 2 diabetes mellitus.

[DOCX File, 14KB - medinform_v7i2e10879_app7.docx]

Multimedia Appendix 8

Guidelines screened for indicators for asthma.

[DOCX File, 13KB - medinform_v7i2e10879_app8.docx]

Multimedia Appendix 9

Guidelines screened for indicators for arterial hypertension.

[DOCX File, 14KB - medinform_v7i2e10879_app9.docx]

Multimedia Appendix 10

Guidelines screened for indicators for chronic heart failure.

[DOCX File, 14KB - medinform_v7i2e10879_app10.docx]

Multimedia Appendix 11

Guidelines screened for indicators for osteoarthritis.

[DOCX File, 13KB - medinform_v7i2e10879_app11.docx]

References

1. World Health Statistics 2018: Monitoring Health for the Sustainable Development Goals (SDGs). Geneva, Switzerland: World Health Organization; 2018. URL: https://www.who.int/gho/publications/world_health_statistics/2016/Annex_B/en/ [accessed 2019-05-14] [WebCite Cache ID 78Mf6qAst]
2. Bachmann N, Burla L, Kohler D, Schweizerisches Gesundheitsobservatorium. Gesundheit in der Schweiz: Fokus Chronische Erkrankungen. Nationaler Gesundheitsbericht 2015. Bern, Switzerland: Hogrefe Verlag; 2015. URL: https://www.obsan.admin.ch/sites/default/files/publications/2015/gesundheitsbericht_2015_d.pdf [accessed 2019-05-13] [WebCite Cache ID 78LZfvcWG]
3. Wagner EH. Chronic disease management: What will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4. [Medline: 10345255]
4. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LM. The eHealth Enhanced Chronic Care Model: A theory derivation approach. J Med Internet Res 2015 Apr 01;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
5. De Rosis S, Seghieri C. Basic ICT adoption and use by general practitioners: An analysis of primary care systems in 31 European countries. BMC Med Inform Decis Mak 2015 Aug 22;15:70 [FREE Full text] [doi: 10.1186/s12911-015-0185-z] [Medline: 26296994]
6. Djalali S, Ursprung N, Rosemann T, Senn O, Tandjung R. Undirected health IT implementation in ambulatory care favors paper-based workarounds and limits health data exchange. Int J Med Inform 2015 Nov;84(11):920-932. [doi: 10.1016/j.ijmedinf.2015.08.001] [Medline: 26296433]
7. Egger M, Smith GD. Principles of and procedures for systematic reviews. In: Egger M, Smith GD, Altman DG, editors. Systematic Reviews in Health Care: Meta‐Analysis in Context. 2nd edition. London, UK: BMJ Publishing Group; 2008:23-42.
8. Agency for Healthcare Research and Quality. 2015. National Guideline Clearinghouse URL: https://www.guideline.gov/ [WebCite Cache ID 6yobWTV23]
9. AWMF Das Portal der wissenschaftlichen Medizin. Leitliniensuche URL: http://www.awmf.org [WebCite Cache ID 6yodPl7r6]
10. Suija K, Kivisto K, Sarria-Santamera A, Kokko S, Liseckiene I, Bredehorst M, et al. Challenges of audit of care on clinical quality indicators for hypertension and type 2 diabetes across four European countries. Fam Pract 2015 Feb;32(1):69-74. [doi: 10.1093/fampra/cmu078] [Medline: 25411423]
11. Shah AD, Langenberg C, Rapsomaniki E, Denaxas S, Pujades-Rodriguez M, Gale CP, et al. Type 2 diabetes and incidence of cardiovascular diseases: A cohort study in 1·9 million people. Lancet Diabetes Endocrinol 2015 Feb;3(2):105-113 [FREE Full text] [doi: 10.1016/S2213-8587(14)70219-0] [Medline: 25466521]
12. Devkota BP, Ansstas M, Scherrer JF, Salas J, Budhathoki C. Internal medicine resident training and provision of diabetes quality of care indicators. Can J Diabetes 2015 Apr;39(2):133-137. [doi: 10.1016/j.jcjd.2014.10.001] [Medline: 25499028]

13. Barkhuysen P, de Grauw W, Akkermans R, Donkers J, Schers H, Biermans M. Is the quality of data in an electronic medical record sufficient for assessing the quality of primary care? J Am Med Inform Assoc 2014;21(4):692-698 [FREE Full text] [doi: 10.1136/amiajnl-2012-001479] [Medline: 24145818]
14. Szczech LA, Stewart RC, Su H, DeLoskey RJ, Astor BC, Fox CH, et al. Primary care detection of chronic kidney disease in adults with type-2 diabetes: The ADD-CKD Study (awareness, detection and drug therapy in type 2 diabetes and chronic kidney disease). PLoS One 2014;9(11):e110535 [FREE Full text] [doi: 10.1371/journal.pone.0110535] [Medline: 25427285]
15. van Melle MA, Lamkaddem M, Stuiver MM, Gerritsen AA, Devillé WL, Essink-Bot ML. Quality of primary care for resettled refugees in the Netherlands with chronic mental and physical health problems: A cross-sectional analysis of medical records and interview data. BMC Fam Pract 2014 Sep 23;15:160 [FREE Full text] [doi: 10.1186/1471-2296-15-160] [Medline: 25249345]
16. Djalali S, Frei A, Tandjung R, Baltensperger A, Rosemann T. Swiss quality and outcomes framework: Quality indicators for diabetes management in Swiss primary care based on electronic medical records. Gerontology 2014;60(3):263-273. [doi: 10.1159/000357370] [Medline: 24603324]
17. Goff SL, Murphy L, Knee AB, Guhn-Knight H, Guhn A, Lindenauer PK. Effects of an enhanced primary care program on diabetes outcomes. Am J Manag Care 2017 Mar 01;23(3):e75-e81 [FREE Full text] [Medline: 28385028]
18. Vidal-Pardo JI, Pérez-Castro TR, López-Álvarez XL, Santiago-Pérez MI, García-Soidán FJ, Muñiz J. Effect of an educational intervention in primary care physicians on the compliance of indicators of good clinical practice in the treatment of type 2 diabetes mellitus [OBTEDIGA project]. Int J Clin Pract 2013 Aug;67(8):750-758. [doi: 10.1111/ijcp.12145] [Medline: 23668834]
19. Sidorenkov G, Voorham J, de Zeeuw D, Haaijer-Ruskamp FM, Denig P. Treatment quality indicators predict short-term outcomes in patients with diabetes: A prospective cohort study using the GIANTT database. BMJ Qual Saf 2013 Apr;22(4):339-347. [doi: 10.1136/bmjqs-2012-001203] [Medline: 23386729]
20. Winkley K, Thomas SM, Sivaprasad S, Chamley M, Stahl D, Ismail K, et al. The clinical characteristics at diagnosis of type 2 diabetes in a multi-ethnic population: The South London Diabetes cohort (SOUL-D). Diabetologia 2013 Jun;56(6):1272-1281. [doi: 10.1007/s00125-013-2873-5] [Medline: 23494447]
21. Gavran L, Zildzic M, Batic-Mujanovic O, Alic A, Gledo I, Prasko S. Auditing of medical chart among type 2 diabetic patient done by primary care physicians. Med Arch 2012;66(6):388-390. [Medline: 23409518]
22. Knudsen ST, Mosbech TH, Hansen B, Kønig E, Johnsen PC, Kamper A. Screening for microalbuminuria in patients with type 2 diabetes is incomplete in general practice. Dan Med J 2012 Sep;59(9):A4502. [Medline: 22951198]
23. Mata-Cases M, Roura-Olmeda P, Berengué-Iglesias M, Birulés-Pons M, Mundet-Tuduri X, Franch-Nadal J, Diabetes Study Group in Primary Health Care (GEDAPS: Grup d'Estudi de la Diabetis a l'Atenció Primària de Salut, Catalonian Society of Family and Community Medicine). Fifteen years of continuous improvement of quality care of type 2 diabetes mellitus in primary care in Catalonia, Spain. Int J Clin Pract 2012 Mar;66(3):289-298 [FREE Full text] [doi: 10.1111/j.1742-1241.2011.02872.x] [Medline: 22340449]
24. Nouwens E, van Lieshout J, Wensing M. Comorbidity complicates cardiovascular treatment: Is diabetes the exception? Neth J Med 2012 Sep;70(7):298-305 [FREE Full text] [Medline: 22961822]
25. Satman I, Imamoglu S, Yilmaz C, ADMIRE Study Group. A patient-based study on the adherence of physicians to guidelines for the management of type 2 diabetes in Turkey. Diabetes Res Clin Pract 2012 Oct;98(1):75-82. [doi: 10.1016/j.diabres.2012.05.003] [Medline: 22652276]
26. Marley JV, Nelson C, O'Donnell V, Atkinson D. Quality indicators of diabetes care: An example of remote-area Aboriginal primary health care over 10 years. Med J Aust 2012 Oct 01;197(7):404-408. [Medline: 23025738]
27. Patapas JM, Blanchard AC, Iqbal S, Vasilevsky M, Dannenbaum D. Management of aboriginal and nonaboriginal people with chronic kidney disease in Quebec: Quality-of-care indicators. Can Fam Physician 2012 Feb;58(2):e107-e111 [FREE Full text] [Medline: 22439172]
28. Staff M. Can data extraction from general practitioners' electronic records be used to predict clinical outcomes for patients with type 2 diabetes? Inform Prim Care 2012;20(2):95-102 [FREE Full text] [Medline: 23710774]
29. Hill F, Bradley C. A computer based, automated analysis of process and outcomes of diabetic care in 23 GP practices. Ir Med J 2012 Feb;105(2):45-47. [Medline: 22455238]
30. Alfadda AA, Bin-Abdulrahman KA, Saad HA, Mendoza CD, Angkaya-Bagayawa FF, Yale JF. Effect of an intervention to improve the management of patients with diabetes in primary care practice. Saudi Med J 2011 Jan;32(1):36-40. [Medline: 21212914]
31. Dickerson LM, Ables AZ, Everett CJ, Mainous AG, McCutcheon AM, Bazaldua OV, National Interdisciplinary Primary Care Practice-Based Research Network. Measuring diabetes care in the national interdisciplinary primary care practice-based research network (NIPC-PBRN). Pharmacotherapy 2011 Jan;31(1):23-30. [doi: 10.1592/phco.31.1.23] [Medline: 21182355]
32. Vidal Pardo JI, Pérez Castro TR, López Álvarez XL, García Soidán FJ, Santiago Pérez MI, Muñiz J. Quality of care of patients with type-2 diabetes in Galicia (NW Spain) [OBTEDIGA project]. Int J Clin Pract 2011 Oct;65(10):1067-1075. [doi: 10.1111/j.1742-1241.2011.02739.x] [Medline: 21801286]

33. Weenink J, van Lieshout J, Jung HP, Wensing M. Patient care teams in treatment of diabetes and chronic heart failure in primary care: An observational networks study. Implement Sci 2011 Jul 03;6:66 [FREE Full text] [doi: 10.1186/1748-5908-6-66] [Medline: 21722399]
34. Gladstone J, Howard M. Effect of advanced access scheduling on chronic health care in a Canadian practice. Can Fam Physician 2011 Jan;57(1):e21-e25 [FREE Full text] [Medline: 21252121]
35. Holbrook A, Pullenayegum E, Thabane L, Troyan S, Foster G, Keshavjee K, et al. Shared electronic vascular risk decision support in primary care: Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness (COMPETE III) randomized trial. Arch Intern Med 2011 Oct 24;171(19):1736-1744. [doi: 10.1001/archinternmed.2011.471] [Medline: 22025430]
36. O'Connor PJ, Sperl-Hillen JM, Rush WA, Johnson PE, Amundson GH, Asche SE, et al. Impact of electronic health record clinical decision support on diabetes care: A randomized trial. Ann Fam Med 2011;9(1):12-21 [FREE Full text] [doi: 10.1370/afm.1196] [Medline: 21242556]
37. Sundquist K, Chaikiat A, León VR, Johansson S, Sundquist J. Country of birth, socioeconomic factors, and risk factor control in patients with type 2 diabetes: A Swedish study from 25 primary health-care centres. Diabetes Metab Res Rev 2011 Mar;27(3):244-254. [doi: 10.1002/dmrr.1161] [Medline: 21309045]
38. Reddy P, Philpot B, Ford D, Dunbar JA. Identification of depression in diabetes: The efficacy of PHQ-9 and HADS-D. Br J Gen Pract 2010 Jun;60(575):e239-e245 [FREE Full text] [doi: 10.3399/bjgp10X502128] [Medline: 20529487]
39. Samoutis GA, Soteriades ES, Stoffers HE, Philalithis A, Delicha EM, Lionis C. A pilot quality improvement intervention in patients with diabetes and hypertension in primary care settings of Cyprus. Fam Pract 2010 Jun;27(3):263-270. [doi: 10.1093/fampra/cmq009] [Medline: 20348163]
40. Shah BR, Bhattacharyya O, Yu C, Mamdani M, Parsons JA, Straus SE, et al. Evaluation of a toolkit to improve cardiovascular disease screening and treatment for people with type 2 diabetes: Protocol for a cluster-randomized pragmatic trial. Trials 2010 Apr 23;11:44 [FREE Full text] [doi: 10.1186/1745-6215-11-44] [Medline: 20416080]
41. Petrazzuoli F, Soler JK, Buono N, Dobbs F. Quality of care for hypertensive patients with type 2 diabetes in a rural area of Southern Italy: Is the recording of patient data and the achievement of quality indicators targets satisfactory? Rural Remote Health 2010;10(3):1258 [FREE Full text] [Medline: 20843159]
42. Sperl-Hillen JM, O'Connor PJ, Rush WA, Johnson PE, Gilmer T, Biltz G, et al. Simulated physician learning program improves glucose control in adults with diabetes. Diabetes Care 2010 Aug;33(8):1727-1733 [FREE Full text] [doi: 10.2337/dc10-0439] [Medline: 20668151]
43. Pedersen ML. Management of type 2 diabetes mellitus in Greenland, 2008: Examining the quality and organization of diabetes care. Int J Circumpolar Health 2009 Apr;68(2):123-132. [Medline: 19517872]
44. Holbrook A, Thabane L, Keshavjee K, Dolovich L, Bernstein B, Chan D, COMPETE II Investigators. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. CMAJ 2009 Jul 07;181(1-2):37-44 [FREE Full text] [doi: 10.1503/cmaj.081272] [Medline: 19581618]
45. Moharram MM, Farahat FM. Quality improvement of diabetes care using flow sheets in family health practice. Saudi Med J 2008 Jan;29(1):98-101. [Medline: 18176682]
46. Novo A, Jokić I. Medical audit of diabetes mellitus in primary care setting in Bosnia and Herzegovina. Croat Med J 2008 Dec;49(6):757-762 [FREE Full text] [Medline: 19090600]
47. Samuels TA, Bolen S, Yeh HC, Abuid M, Marinopoulos SS, Weiner JP, et al. Missed opportunities in diabetes management: A longitudinal assessment of factors associated with sub-optimal quality. J Gen Intern Med 2008 Nov;23(11):1770-1777 [FREE Full text] [doi: 10.1007/s11606-008-0757-z] [Medline: 18787908]
48. Smith SA, Shah ND, Bryant SC, Christianson TJ, Bjornsen SS, Giesler PD, Evidens Research Group. Chronic care model and shared care in diabetes: Randomized trial of an electronic decision support system. Mayo Clin Proc 2008 Jul;83(7):747-757. [doi: 10.4065/83.7.747] [Medline: 18613991]
49. Voorham J, Denig P, Wolffenbuttel BH, Haaijer-Ruskamp FM. Cross-sectional versus sequential quality indicators of risk factor management in patients with type 2 diabetes. Med Care 2008 Feb;46(2):133-141. [doi: 10.1097/MLR.0b013e31815b9da0] [Medline: 18219241]
50. Wens J, Dirven K, Mathieu C, Paulus D, Van Royen P, Belgian Diabetes Project Group. Quality indicators for type-2 diabetes care in practice guidelines: An example from six European countries. Prim Care Diabetes 2007 Feb;1(1):17-23. [doi: 10.1016/j.pcd.2006.07.001] [Medline: 18632015]
51. Nitiyanant W, Chetthakul T, Sang-A-kad P, Therakiatkumjorn C, Kunsuikmengrai K, Yeo JP. A survey study on diabetes management and complication status in primary care setting in Thailand. J Med Assoc Thai 2007 Jan;90(1):65-71. [Medline: 17621735]
52. Herrin J, Nicewander DA, Hollander PA, Couch CE, Winter FD, Haydar ZR, et al. Effectiveness of diabetes resource nurse case management and physician profiling in a fee-for-service setting: A cluster randomized trial. Proc (Bayl Univ Med Cent) 2006 Apr;19(2):95-102 [FREE Full text] [Medline: 16609732]
53. Wan Q, Harris MF, Jayasinghe UW, Flack J, Georgiou A, Penn DL, et al. Quality of diabetes care and coronary heart disease absolute risk in patients with type 2 diabetes mellitus in Australian general practice. Qual Saf Health Care 2006 Apr;15(2):131-135 [FREE Full text] [doi: 10.1136/qshc.2005.014845] [Medline: 16585115]

54. Al Khaja KA, Sequeira RP, Damanhori AH. Comparison of the quality of diabetes care in primary care diabetic clinics and general practice clinics. Diabetes Res Clin Pract 2005 Nov;70(2):174-182. [doi: 10.1016/j.diabres.2005.03.029] [Medline: 15890429]
55. Cueto-Manzano AM, Cortes-Sanabria L, Martinez-Ramirez HR, Rojas-Campos E, Barragan G, Alfaro G, et al. Detection of early nephropathy in Mexican patients with type 2 diabetes mellitus. Kidney Int Suppl 2005 Aug(97):S40-S45 [FREE Full text] [doi: 10.1111/j.1523-1755.2005.09707.x] [Medline: 16014099]
56. Lusignan S, Sismanidis C, Carey IM, DeWilde S, Richards N, Cook DG. Trends in the prevalence and management of diagnosed type 2 diabetes 1994-2001 in England and Wales. BMC Fam Pract 2005 Mar 22;6(1):13 [FREE Full text] [doi: 10.1186/1471-2296-6-13] [Medline: 15784133]
57. Sequeira RP, Al Khaja KA, Damanhori AH. Evaluating the treatment of hypertension in diabetes mellitus: A need for better control? J Eval Clin Pract 2004 Feb;10(1):107-116. [Medline: 14731158]
58. Wermeille J, Bennie M, Brown I, McKnight J. Pharmaceutical care model for patients with type 2 diabetes: Integration of the community pharmacist into the diabetes team--A pilot study. Pharm World Sci 2004 Feb;26(1):18-25. [Medline: 15018255]
59. Goudswaard AN, Lam K, Stolk RP, Rutten GE. Quality of recording of data from patients with type 2 diabetes is not a valid indicator of quality of care: A cross-sectional study. Fam Pract 2003 Apr;20(2):173-177. [Medline: 12651792]
60. Campbell SM, Hann M, Hacker J, Durie A, Thapar A, Roland MO. Quality assessment for three common conditions in primary care: Validity and reliability of review criteria developed by expert panels for angina, asthma and type 2 diabetes. Qual Saf Health Care 2002 Jun;11(2):125-130 [FREE Full text] [doi: 10.1136/qhc.11.2.125] [Medline: 12448803]
61. Parchman ML, Burge SK, Residency Research Network of South Texas Investigators. Continuity and quality of care in type 2 diabetes: A Residency Research Network of South Texas study. J Fam Pract 2002 Jul;51(7):619-624. [Medline: 12160500]
62. Renders CM, Valk GD, Franse LV, Schellevis FG, van Eijk JT, van der Wal G. Long-term effectiveness of a quality improvement program for patients with type 2 diabetes in general practice. Diabetes Care 2001 Aug;24(8):1365-1370. [Medline: 11473071]
63. Linmans JJ, Spigt MG, Deneer L, Lucas AE, de Bakker M, Gidding LG, et al. Effect of lifestyle intervention for people with diabetes or prediabetes in real-world primary care: Propensity score analysis. BMC Fam Pract 2011 Sep 13;12:95 [FREE Full text] [doi: 10.1186/1471-2296-12-95] [Medline: 21914190]
64. Minard JP, Dostaler SM, Taite AK, Olajos-Clow JG, Sands TW, Licskai CJ, et al. Development and implementation of an electronic asthma record for primary care: Integrating guidelines into practice. J Asthma 2014 Feb;51(1):58-68. [doi: 10.3109/02770903.2013.845206] [Medline: 24083321]
65. Lim KG, Rank MA, Cabanela RL, Furst JW, Rohrer JE, Liesinger J, et al. The asthma ePrompt: A novel electronic solution for chronic disease management. J Asthma 2012 Mar;49(2):213-218. [doi: 10.3109/02770903.2012.654419] [Medline: 22304226]
66. Lougheed MD, Minard J, Dworkin S, Juurlink M, Temple WJ, To T, et al. Pan-Canadian REspiratory STandards INitiative for Electronic Health Records (PRESTINE): 2011 national forum proceedings. Can Respir J 2012;19(2):117-126 [FREE Full text] [doi: 10.1155/2012/870357] [Medline: 22536581]
67. Oei SM, Thien FC, Schattner RL, Sulaiman ND, Birch K, Simpson P, et al. Effect of spirometry and medical review on asthma control in patients in general practice: A randomized controlled trial. Respirology 2011 Jul;16(5):803-810. [doi: 10.1111/j.1440-1843.2011.01969.x] [Medline: 21401801]
68. Nokela M, Arnlind MH, Ehrs P, Krakau I, Forslund L, Jonsson EW. The influence of structured information and monitoring on the outcome of asthma treatment in primary care: A cluster randomized study. Respiration 2010;79(5):388-394 [FREE Full text] [doi: 10.1159/000235548] [Medline: 19672055]
69. Yawn BP, Bertram S, Wollan P. Introduction of Asthma APGAR tools improve asthma management in primary care practices. J Asthma Allergy 2008 Aug 31;1:1-10 [FREE Full text] [Medline: 21436980]
70. Baddar S, Worthing EA, Al-Rawas OA, Osman Y, Al-Riyami BM. Compliance of physicians with documentation of an asthma management protocol. Respir Care 2006 Dec;51(12):1432-1440 [FREE Full text] [Medline: 17134524]
71. Hasselström J, Zarrinkoub R, Holmquist C, Hjerpe P, Ljungman C, Qvarnström M, et al. The Swedish Primary Care Cardiovascular Database (SPCCD): 74 751 hypertensive primary care patients. Blood Press 2014 Apr;23(2):116-125. [doi: 10.3109/08037051.2013.814829] [Medline: 23914944]
72. Tong SF, Khoo EM, Nordin S, Teng C, Lee VK, Zailinawati AH, et al. Process of care and prescribing practices for hypertension in public and private primary care clinics in Malaysia. Asia Pac J Public Health 2012 Sep;24(5):764-775. [doi: 10.1177/1010539511402190] [Medline: 21659332]
73. Pavlik VN, Greisinger AJ, Pool J, Haidet P, Hyman DJ. Does reducing physician uncertainty improve hypertension control?: Rationale and methods. Circ Cardiovasc Qual Outcomes 2009 May;2(3):257-263 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.109.849984] [Medline: 20031846]
74. Chan PF, Chao DVK. The Hong Kong Practictioner. 2006 Dec. An audit on management of hypertension in a hospital authority general outpatient clinic URL: http://www.hkcfp.org.hk/Upload/HK_Practitioner/2006/hkp2006vol28dec/original_article.html [accessed 2019-05-13] [WebCite Cache ID 78LcpOKYm]

75. Asnani M, Brown P, O'Connor D, Lewis T, Win S, Reid M. A clinical audit of the quality of care of hypertension in general practice. West Indian Med J 2005 Jun;54(3):176-180. [Medline: 16209222]
76. Rabinowitz I, Tamir A. The SaM (Screening and Monitoring) approach to cardiovascular risk-reduction in primary care: Cyclic monitoring and individual treatment of patients at cardiovascular risk using the electronic medical record. Eur J Cardiovasc Prev Rehabil 2005 Feb;12(1):56-62. [Medline: 15703507]
77. Mitchell E, Sullivan F, Grimshaw JM, Donnan PT, Watt G. Improving management of hypertension in general practice: A randomised controlled trial of feedback derived from electronic patient data. Br J Gen Pract 2005 Feb;55(511):94-101 [FREE Full text] [Medline: 15720929]
78. Alli C, Mariotti G, Avanzini F, Colombo F, Barlera S, Tognoni G, Studio sulla Pressione Arteriosa nell'Anziano (SPAA). Long-term prognostic impact of repeated measurements over 1 year of pulse pressure and systolic blood pressure in the elderly. J Hum Hypertens 2005 May;19(5):355-363. [doi: 10.1038/sj.jhh.1001827] [Medline: 15772693]
79. Tierney WM, Brunt M, Kesterson J, Zhou X, L'Italien G, Lapuerta P. Quantifying risk of adverse clinical events with one set of vital signs among primary care patients with hypertension. Ann Fam Med 2004;2(3):209-217 [FREE Full text] [Medline: 15209196]
80. Lackland DT, Lin Y, Tilley BC, Egan BM. An assessment of racial differences in clinical practices for hypertension at primary care sites for medically underserved patients. J Clin Hypertens (Greenwich) 2004 Jan;6(1):26-31; quiz 32 [FREE Full text] [Medline: 14724421]
81. Frijling BD, Lobo CM, Hulscher ME, Akkermans RP, van Drenth BB, Prins A, et al. Intensive support to improve clinical decision making in cardiovascular care: A randomised controlled trial in general practice. Qual Saf Health Care 2003 Jun;12(3):181-187 [FREE Full text] [doi: 10.1136/qhc.12.3.181] [Medline: 12792007]
82. Amarasingham R, Patel PC, Toto K, Nelson LL, Swanson TS, Moore BJ, et al. Allocating scarce resources in real-time to reduce heart failure readmissions: A prospective, controlled study. BMJ Qual Saf 2013 Dec;22(12):998-1005 [FREE Full text] [doi: 10.1136/bmjqs-2013-001901] [Medline: 23904506]
83. Logeart D, Isnard R, Resche-Rigon M, Seronde M, de Groote P, Jondeau G, Heart Failure of the French Society of Cardiology. Current aspects of the spectrum of acute heart failure syndromes in a real-life setting: The OFICA study. Eur J Heart Fail 2013 Apr;15(4):465-476 [FREE Full text] [doi: 10.1093/eurjhf/hfs189] [Medline: 23186936]
84. Lind L, Klompstra L, Jaarsma T, Strömberg A. Implementation and testing of a digital pen and paper tool to support patients with heart failure and their health care providers in detecting early signs of deterioration and monitor adherence. Int J Integr Care 2011 Jul 21;11(7):e115. [doi: 10.5334/ijic.787]
85. Maddocks H, Marshall JN, Stewart M, Terry AL, Cejic S, Hammond J, et al. Quality of congestive heart failure care: Assessing measurement of care using electronic medical records. Can Fam Physician 2010 Dec;56(12):e432-e437 [FREE Full text] [Medline: 21156884]
86. Korb K, Hummers-Pradier E, Stich K, Chenot J, Scherer M. Implementation of recommendations for the diagnosis of heart failure. [corrected] [Article in German]. Dtsch Med Wochenschr 2010 Jan;135(4):120-124. [doi: 10.1055/s-0029-1244827] [Medline: 20101554]
87. Fonarow GC, Albert NM, Curtis AB, Stough WG, Gheorghiade M, Heywood JT, et al. Improving evidence-based care for heart failure in outpatient cardiology practices: Primary results of the Registry to Improve the Use of Evidence-Based Heart Failure Therapies in the Outpatient Setting (IMPROVE HF). Circulation 2010 Aug 10;122(6):585-596. [doi: 10.1161/CIRCULATIONAHA.109.934471] [Medline: 20660805]
88. Vercauteren S, Castadot M, Vanhalewyn M, Boulanger S, Robert A, Bouvy C, et al. BELGIUM-HF (Better Efficacy in Lowering events by General practitioner's Intervention Using remote Monitoring in Heart Failure): Concept and feasibility. In: Proceedings of the Belgian Society of Cardiology, 28th Annual Scientific Meeting. 2009 Feb Presented at: Belgian Society of Cardiology, 28th Annual Scientific Meeting; January 29-31, 2009; Brussels, Belgium p. 117. [doi: 10.2143/AC.64.1.2034372]
89. Majeed A, Williams J, de Lusignan S, Chan T. Management of heart failure in primary care after implementation of the National Service Framework for Coronary Heart Disease: A cross-sectional study. Public Health 2005 Feb;119(2):105-111. [doi: 10.1016/j.puhe.2004.06.006] [Medline: 15694957]
90. Subramanian U, Fihn SD, Weinberger M, Plue L, Smith FE, Udris EM, et al. A controlled trial of including symptom data in computer-based care suggestions for managing patients with chronic heart failure. Am J Med 2004 Mar 15;116(6):375-384. [doi: 10.1016/j.amjmed.2003.11.021] [Medline: 15006586]
91. Roth A, Kajiloti I, Elkayam I, Sander J, Kehati M, Golovner M. Telecardiology for patients with chronic heart failure: The 'SHL' experience in Israel. Int J Cardiol 2004 Oct;97(1):49-55. [doi: 10.1016/j.ijcard.2003.07.030] [Medline: 15336806]
92. Gnani S, Gray J, Khunti K, Majeed A. Managing heart failure in primary care: First steps in implementing the National Service Framework. J Public Health (Oxf) 2004 Mar;26(1):42-47. [Medline: 15044573]
93. Grypdonck L, Aertgeerts B, Luyten F, Wollersheim H, Bellemans J, Peers K, et al. Development of quality indicators for an integrated approach of knee osteoarthritis. J Rheumatol 2014 Jun;41(6):1155-1162. [doi: 10.3899/jrheum.130680] [Medline: 24737907]

94. Jansen MJ, Hendriks EJ, Oostendorp RA, Dekker J, De Bie RA. Quality indicators indicate good adherence to the clinical practice guideline on "Osteoarthritis of the hip and knee" and few prognostic factors influence outcome indicators: A prospective cohort study. Eur J Phys Rehabil Med 2010 Sep;46(3):337-345 [FREE Full text] [Medline: 20926999]
95. MacLean CH, Saag KG, Solomon DH, Morton SC, Sampsel S, Klippel JH. Measuring quality in arthritis care: Methods for developing the Arthritis Foundation's quality indicator set. Arthritis Rheum 2004 Apr 15;51(2):193-202 [FREE Full text] [doi: 10.1002/art.20248] [Medline: 15077259]
96. Peat G, Lawton H, Hay E, Greig J, Thomas E, KNE-SCI Study Group. Development of the Knee Standardized Clinical Interview: A research tool for studying the primary care clinical epidemiology of knee problems in older adults. Rheumatology (Oxford) 2002 Oct;41(10):1101-1108. [Medline: 12364627]
97. Glasziou P, Irwig L, Mant D. Monitoring in chronic disease: A rational approach. BMJ 2005 Mar 19;330(7492):644-648 [FREE Full text] [doi: 10.1136/bmj.330.7492.644] [Medline: 15774996]
98. van Steenkiste BC, Jacobs JE, Verheijen NM, Levelink JH, Bottema BJ. A Delphi technique as a method for selecting the content of an electronic patient record for asthma. Int J Med Inform 2002 Apr;65(1):7-16. [Medline: 11904244]
99. Gensichen J, Peitz M, Torge M, Mosig-Frey J, Wendt-Hermainski H, Rosemann T, et al. The "Depression Monitoring list" (DeMoL) with integrated PHQ-D-Rationale and design of a tool for the case management for depression in primary care [Article in German]. Z Arztl Fortbild Qualitatssich 2006;100(5):375-382. [Medline: 16955623]
100. Lougheed MD, Taite A, ten Hove J, Morra A, Van Dam A, Ducharme FM, et al. Pan-Canadian asthma and COPD standards for electronic health records: A Canadian Thoracic Society Expert Working Group Report. Can J Respir Crit Care Sleep Med 2018 Nov 05;2(4):244-250. [doi: 10.1080/24745332.2018.1517623]
101. Chmiel C, Birnbaum B, Gensichen J, Rosemann T, Frei A. The diabetes traffic light scheme: Development of an instrument for the case management in patients with diabetes mellitus in primary care [Article in German]. Praxis (Bern 1994) 2011 Nov 30;100(24):1457-1473. [doi: 10.1024/1661-8157/a000751] [Medline: 22124957]
102. Roland M. Linking physicians' pay to the quality of care: A major experiment in the United Kingdom. N Engl J Med 2004 Sep 30;351(14):1448-1454. [doi: 10.1056/NEJMhpr041294] [Medline: 15459308]
103. Chmiel C, Senn O, Rosemann T, Del Prete V, Steurer-Stey C. CoCo trial: Color-coded blood pressure Control, a randomized controlled study. Patient Prefer Adherence 2014;8:1383-1392 [FREE Full text] [doi: 10.2147/PPA.S68213] [Medline: 25346595]
104. Schectman G, Barnas G, Laud P, Cantwell L, Horton M, Zarling EJ. Prolonging the return visit interval in primary care. Am J Med 2005 Apr;118(4):393-399. [doi: 10.1016/j.amjmed.2005.01.003] [Medline: 15808137]
105. Bakker K, Apelqvist J, Schaper NC, International Working Group on Diabetic Foot Editorial Board. Practical guidelines on the management and prevention of the diabetic foot 2011. Diabetes Metab Res Rev 2012 Feb;28 Suppl 1:225-231. [doi: 10.1002/dmrr.2253] [Medline: 22271742]
106. Clarson LE, Nicholl BI, Bishop A, Edwards JJ, Daniel R, Mallen CD. Monitoring osteoarthritis: A cross-sectional survey in general practice. Clin Med Insights Arthritis Musculoskelet Disord 2013;6:85-91 [FREE Full text] [doi: 10.4137/CMAMD.S12606] [Medline: 24324351]
107. Scalvini S, Martinelli G, Baratti D, Domenighini D, Benigno M, Paletta L, et al. Telecardiology: One-lead electrocardiogram monitoring and nurse triage in chronic heart failure. J Telemed Telecare 2005;11 Suppl 1:18-20. [doi: 10.1258/1357633054461750] [Medline: 16035981]
108. Morilla-Herrera JC, Garcia-Mayor S, Martín-Santos FJ, Kaknani Uttumchandani S, Leon Campos Á, Caro Bautista J, et al. A systematic review of the effectiveness and roles of advanced practice nursing in older people. Int J Nurs Stud 2016 Jan;53:290-307. [doi: 10.1016/j.ijnurstu.2015.10.010] [Medline: 26542652]
109. Agrinier N, Altieri C, Alla F, Jay N, Dobre D, Thilly N, et al. The ICALOR home-nurse led disease management programme in heart failure prevents hospital readmission: A French nationwide time-series comparison. In: Proceedings of the European Society of Cardiology Congress 2012. 2012 Presented at: European Society of Cardiology Congress 2012; August 25-29, 2012; Munich, Germany p. 159-160 URL: https://academic.oup.com/eurheartj/article-pdf/33/suppl_1/19/1188353/ehs281.pdf

Abbreviations

AWMF: Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften eV
CCM: chronic care model
EMR: electronic medical record
HbA_{1c}: hemoglobin A_{1c}
QOF: Quality and Outcomes Framework

Edited by G Eysenbach; submitted 06.05.18; peer-reviewed by R Bush, M Reading, W Zhang; comments to author 01.09.18; revised version received 20.03.19; accepted 07.04.19; published 24.05.19.

Please cite as:
Falck L, Zoller M, Rosemann T, Martínez-González NA, Chmiel C
Toward Standardized Monitoring of Patients With Chronic Diseases in Primary Care Using Electronic Medical Records: Systematic Review
JMIR Med Inform 2019;7(2):e10879
URL: http://medinform.jmir.org/2019/2/e10879/
doi:10.2196/10879
PMID:31127717

Original Paper

Incorporating Social Determinants of Health in Electronic Health Records: Qualitative Study of Current Practices Among Top Vendors

Maysoun Freij[1], MPH, PhD; Prashila Dullabh[1], MD; Sarah Lewis[2], MSPH, PhD; Scott R Smith[3], MPH, PhD; Lauren Hovey[1], MA; Rina Dhopeshwarkar[1], MPH

[1]NORC at the University of Chicago, Bethesda, MD, United States

[2]Centers for Medicare and Medicaid Innovation, Baltimore, MD, United States

[3]Office of the Assistant Secretary for Planning and Evaluation, Washington, DC, United States

Corresponding Author:
Maysoun Freij, MPH, PhD
NORC at the University of Chicago
4350 East West Highway
8th Floor
Bethesda, MD, 20814
United States
Phone: 1 301 634 9539
Email: freij-maysoun@norc.org

Abstract

Background: Social determinants of health (SDH) are increasingly seen as important to understanding patient health and identifying appropriate interventions to improve health outcomes in what is a complex interplay between health system-, community-, and individual-level factors.

Objective: The objective of the paper was to investigate the development of electronic health record (EHR) software products that allow health care providers to identify and address patients' SDH in health care settings.

Methods: We conducted interviews with six EHR vendors with large market shares in both ambulatory and inpatient settings. We conducted thematic analysis of the interviews to (1) identify their motivations to develop such software products, (2) describe their products and uses, and (3) identify facilitators and challenges to collection and use of SDH data—through their products or otherwise—either at the point of care or in population health interventions.

Results: Our findings indicate that vendor systems and their functionalities are influenced by client demand and initiative, federal initiatives, and the vendors' strategic vision about opportunities in the health care system. Among the small sample of vendors with large market shares, SDH is a new area for growth, and the vendors range in the number and sophistication of their SDH-related products. To enable better data analytics, population health management, and interoperability of SDH data, vendors recognized the need for more standardization of SDH performance measures across various federal and state programs, better mapping of SDH measures to multiple types of codes, and development of more codes for all SDH measures of interest.

Conclusions: Vendors indicate they are actively developing products to facilitate the collection and use of SDH data for their clients and are seeking solutions to data standardization and interoperability challenges through internal product decisions and collaboration with policymakers. Due to a lack of policy standards around SDH data, product-specific decisions may end up being de facto policies given the market shares of particular vendors. However, commercial vendors appear ready to collaboratively discuss policy solutions such as standards or guidelines with each other, health care systems, and government agencies in order to further promote integration of SDH data into the standard of care for all health systems.

(JMIR Med Inform 2019;7(2):e13849) doi:10.2196/13849

KEYWORDS

electronic health records; social determinants of health

Introduction

Health care reform initiatives over the past decade have incentivized value-based care payment models and the adoption and development of electronic health records (EHRs) [1,2]. Emphasis on value over volume has drawn attention to the importance of social determinants of health (SDH) in potentially affecting health outcomes. SDH include a wide range of social, economic, and environmental factors that contribute to the health of individuals (Figure 1) [3].

A 2014 report by the National Academies of Medicine (NAM) argued that the integration of SDH into EHRs would better enable health providers to address health inequities and support research into how social and environmental factors influence health [4]. Federal initiatives have spurred SDH data collection through EHRs, including the Comprehensive Primary Care Plus (CPC+) model and Medicare Accountable Care Organizations (ACOs) and Accountable Health Communities (AHCs) [5]. The Centers for Medicare & Medicaid Services (CMS) 2016 Medicaid Managed Care rule has encouraged states to include more community-based, nonclinical services that may address SDH [6,7]. At the local level, health care providers, health departments, universities, legal aid, and social service organizations are developing health improvement interventions that rely on the collection and use of SDH data [8].

Numerous screening tools and approaches have been developed to screen and address SDH [9,10]. Three widely recognized SDH screening tools in the United States are (1) the NAM (2014) set of social and behavioral measures [11]; (2) the National Association of Community Health Center (NACHC) Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE) tool [12]; and (3) the Center for Medicare and Medicaid Innovation's Accountable Health Communities tool [13]. These tools vary in terms of the overall number of domains or questions, and health care organizations may choose to include additional SDH domains or measures to meet all the needs of their patients. A recent study of six health systems found they all included domains in their SDH screening tools that are not among NAM's recommended domains, including housing, food insecurity, and transportation [14]. By adapting screening tool questions and domains, organizations have effectively created many different SDH screening tools. Lack of standardization for incorporating data from various screening tools and measures has limited the usefulness of the data within and across EHR systems [15].

With expanded government interest in value-based care (VBC) and quality, health information technology companies that serve as EHR vendors have had both indirect and direct roles in working with policymakers and health care systems. Their indirect role in policymaking has occurred through partnerships with the federal government, health care systems, and other technology companies [16,17]. In forging these relationships, policymakers have directly contributed to the evolution of EHR vendors' interest in actively engaging in population health as opposed to only developing medical record-keeping products [18,19].

Figure 1. Social determinants of health. Adapted from: Healthy People 2020: Social Determinants of Health [3].

Further, vendors are increasingly incorporating SDH into their EHRs as a way to help their clients respond to the anticipated quality demands of value-based purchasing [20]. Some have dubbed this as a shift from EHRs to comprehensive health records [21]. While health care systems may influence the development of EHR features, there are concerns that the large market shares of relatively few EHR vendors may make vendors less responsive to designing EHRs to meet patients' and clinicians' needs, particularly while controlling costs and promoting interoperability [22].

Because of their unique position at the nexus of health systems and health policies and the significant impact their organizational decisions have on EHR-based data capture and clinical practice, we conducted key informant interviews with top EHR vendors focused on SDH. This paper describes vendor perspectives on current challenges and promising opportunities to improve the capture and usability of SDH data in EHRs.

Methods

We began with a scan of PubMed for peer-reviewed literature and grey literature involving EHRs, SDH, and/or health disparities. Results were limited to articles published in English between January 2012 and June 2018. Through a preliminary review of over 250 articles, we identified 52 for in-depth review and thematic analysis of current practices for collecting and using SDH data through EHRs, uses of SDH data in EHRs for clinical care, and promising opportunities for improving such data collection.

Building on this information, we conducted key informant interviews with research and product development staff at EHR vendor companies to learn more about their current activities related to the integration of SDH in EHRs. To draw a purposive sample, we identified 10 vendors with the largest market shares in hospital and ambulatory settings (a total of 17 vendors) and selected the three vendors that held the largest shares in both settings. We then included the three other vendors among the five vendors with the largest shares in either inpatient or ambulatory settings, for a total of nine vendors [23,24]. Through email solicitation, we gained participation from six vendors but were unable to reach appropriate staff for three vendors during the study period. One to three representatives for each vendor joined the phone interviews in March and April 2018. The vendor and participant names have been kept confidential. The interviews were 60 minutes in length and were audio-recorded and transcribed for the purposes of analysis. We explored motivators, successes and facilitators, challenges and barriers, and lessons learned from SDH product development and solicited feedback for policymakers to consider that would improve the collection and use of SDH data for patient care. This study was reviewed and approved by the University of Chicago's Institutional Review Board.

We conducted a thematic analysis of interview transcripts using NVivo software (QSR International Pty Ltd). To conduct this analysis, we developed a code book based upon topics discussed during the interviews and also a conceptual model that emerged from the interviews. In the conceptual model, vendors' clients (ie, health care systems or providers) have their own interests and preferences in relation to the policy environment, needs of their patients, resources in their community, and their own models of health care. Health care system clients provide sites for implementation and testing of SDH tools and are often part of the development of the vendors' SDH products themselves. Our analysis explored the intersection of health policy and health systems in vendor perspectives on SDH product development.

The code book included definitions of individual codes related to policy demands, client demands, vendor's motivators' and experiences, SDH data sources and products, research and development of SDH products, implementation experiences, and vendor requests in terms of policies or strategies to facilitate the collection and use of SDH data through EHRs. A senior researcher developed the codebook and trained three research analysts to each code two to three interviews that they had observed and transcribed. The team met to review and discuss the coding process. Testing of intercoder reliability involved multiple staff coding samples of the same text using an initial codebook. We revised the codebook and refined code definitions as needed to assure consistency across staff coding styles. The senior staff also reviewed coded transcripts to assure accuracy and consistency in coded material. Once transcripts were coded, the authors integrated and interpreted findings across codes to understand current practices in the development of SDH-related products in EHRs and the challenges and opportunities for using these products to address patients' nonmedical needs in health care settings.

Results

Motivators of Social Determinants of Health Product Development

All vendors in our sample stressed the importance of meeting their clients' needs and demands. One of the main drivers of their clients' interests in collecting and using SDH in the course of health care delivery is the expansion of VBC programs. Vendors cited Patient-Centered Medical Homes, CPC+, and ACOs as motivating their clients to ask for SDH products within their EHRs. Two vendors noticed the most demand came from federally qualified health centers (FQHCs) or community health centers, whereas another observed more widespread interest from academic medical centers, integrated delivery systems, and pediatric and/or specialty groups, stating, "there is interest, not only in utilizing [SDH] from a workflow standpoint, but also making sure that [SDH] becomes an integral part of the patient's story over different settings, so that it's becoming more [of a] norm as part of the handoff between care settings."

Additionally, four vendors identified the Promoting Interoperability (formerly Meaningful Use) incentives for EHR use and Office of the National Coordinator for Health Information Technology (ONC) health IT certification requirements as main drivers for the integration of SDH in EHRs. As a result, all providers using certified EHRs are collecting some SDH data (ie, race, ethnicity, gender identification, and sexual orientation), although they may not necessarily view it or act upon it as such. One vendor explained, "with Meaningful Use, every practice has access to EHRs and there is an immense amount of data that is available [that] has

not widely been used for outcomes data research," such as research or interventions on SDH.

Types of Social Determinants of Health Products and Their Use Cases

Vendors in the study sample varied in their level of investment and development of SDH products. Our findings affirm that the types of SDH products created and used by vendors varies greatly based on their client needs and input and their own strategic planning. In general, vendors have or are in the process of incorporating SDH data in screening tools, population health management tools, tools to improve referral management, and analytic tools (Figure 2).

Figure 2. Types of available social determinants of health (SDH) tools and products among sample of vendors (n=6).

Screening Tools Are the Most Common Type of Social Determinants of Health Product

All vendors mentioned the use of screening tools as a part of their platform to collect SDH data. Among the types of screening tools, most vendors described using a configuration of the NACHC's PRAPARE tool due to their clients' current demands and its use of structured data and distinct outcomes. PRAPARE EHR templates exist on most top vendor platforms, and the tool is also free as part of a publicly available toolkit [12]. Beyond enabling clients to use the PRAPARE tool or collect whatever other SDH data they choose through their EHRs, one vendor has developed a fully integrated screening product that includes eight NAM recommended measures and two from PRAPARE. Another is working through the intellectual property rights to fully integrate PRAPARE into applications to make it more usable for all of its clients and not just community health centers. Most vendors also described the use of standardized tools to capture data on behavioral health—a common SDH domain—including the Columbia-Suicide Severity Rating Scale [25] and 9-Item Patient Health Questionnaire [26].

All vendors described offering clients multiple or customizable screening tools to focus on fewer or additional measures as needed. As one vendor explained, "Our overarching strategy is to collect SDH data at the individual level in a structured way that is flexible for clients." Another vendor described the multiplicity of screening tools its clients use and the back-and-forth dynamic with clients that ultimately leads to the development and tailoring of tools:

> *A number of organizations were using [our social history] form that has been there for a long time. They were creating their own forms to be able to collect this data in a variety of different ways. In some cases they were using other tools, such as the PRAPARE tool, that a number in our group liked and adopted, and it made sense... And so, in some cases it really is...customers being innovative and using different tools and giving us feedback that is determining the best way for us to standardize this on a go-forward basis. We certainly never want to restrict customers from doing what they think they need to be successful or to be innovative.*

Variation in screening tools was attributed to the variable demands of particular patient populations (ie, pediatrics) and to the lack of common screening requirements across different federal or state programs. One vendor explained the challenges of developing screening tools that account for federal and state requirements and clients' preferences and integrate into providers' workflows:

> *Mostly what I've seen is each state has a different set of requirements in terms of content, questionnaires, screening tools.... There is variation in requirements from state to state or even in a state depending on the practice size or if they are an FQHC [Federally Qualified Health Center]... [Also] some things [may be] a standard [measure] when it comes to a federal requirement but [how] some [measures] are [collected may be] more specific to [a client's] workflow. In which case we have to make [the measures] go into different sections [of the EHR rather than be in one form that matches the federal requirements]. [The requirements] break the flow sometimes. The customers just want ease of documentation so the challenge is how we can bring everything together into one place. Some being structured data that is standard and some being nonstandard customer specific data.*

While screening tools are a common way of capturing SDH data, vendors also described a number of places where SDH data could be collected or found. These include EHR-specific data sets or forms, problem tables, free-text fields located in various places (eg, social history section, clinical notes and assessments section, details section of structured screening tools), the demographic section of the patient's health record, and the patient portal.

Population Health Management is a Common Use Case for Social Determinants of Health Data

Three vendors described the development of proprietary population health management tools capable of using algorithms,

extracting data, and/or researching community-level patient needs. Although there is not widespread use of SDH data in population health initiatives, one vendor expected that they could be used for diabetes management and food security or medication adherence and utilities. Another described analysis of opioid use, pain tolerance, and pain medication abuse mapped to SDH in areas of opioid addiction. One vendor also described a common request from clients to use secondary survey data to identify "hot spots" or areas of high social need in the communities they serve. It uses data from the CDC Social Vulnerability Index to improve providers' understanding of community-level social health needs [27]. All vendors recognized growing demand from clients to, as one put it, "move the needle in population health."

For Most Vendors, the Use of Referral Products is Still in Early Development or Newly Integrated Into Their Platform

For the five vendors with products capable of making referrals for community services, the common methods are (1) the use of a third-party tool like Aunt Bertha [28], (2) using an EHR-integrated tool like order forms, or (3) using a proprietary tool that allows information exchange among health care systems and outside service providers. These tools are capable of improving care transitions, finding community resources available within a specified radius of a patient's home address, providing a list of requests or interventions that have been recommended for a patient or assigning a patient to a certain referral program, and providing direct messaging between clinical providers and community-based social service providers for a warm handoff and coordination of complex cases. One vendor describes options that clients have in creating and using referral tools:

> *One tool that [we] developed is a search tool that finds community resources given the SDH factors that are at the highest risk. For example, using the patient's home address, we can look within say a 5-mile radius and show all of the transportation services or all of the food pantries. In order to do so our customers can build a list [themselves] or use a third-party vendor that can compile a list that helps them manage the rapidly changing community landscape. Relying on a [third-party] vendor in this space is a strategy that makes sense.*

Further, the vendor has created a portal so that the health system and the community service provider can communicate about shared clients. One interviewee explained:

> *The portal was really to close that loop from a community referral perspective so that they could be on the same care team, they could share parts of the record as appropriate, and they could even contribute feedback by way of notes or simple assessments to really round out the whole picture of someone's care.*

The vendor views such tools as a way of connecting to community-based service providers that historically have not used EHR products but that are integral to addressing the whole health of a patient.

Other vendors also want to close the feedback loop with information on whether patients followed through or benefited from the referral and to have that information reflected in the EHR. Typically, this is done by someone on the clinical care team documenting that the referral has been fulfilled. However, as one vendor observed, among community health clinics, referrals are often made to a service offered within a clinic's facility or by phone to known community-based service providers; as such, these interactions are not commonly documented in the EHR. Vendors also recognized that there is a lack of consistency in how referrals are documented or managed across EHR systems due to variations in standards implementation, proprietary designs, and also challenges with simply making electronic referrals from health systems to community service providers.

Vendors Varied in Their Ability to Provide Data Analytics and Reporting

Similar to the use of referral products and capabilities, vendors are still in the early stages of developing mechanisms for analytics and reporting related to SDH. Three vendors interviewed reported using SDH data from the EHRs for risk stratification and outcome assessment. One mentioned the specific use of SDH for reporting to Medicaid for VBC incentives. Another described the use of analytics and reports for following a patient's progression but was unsure if there is a specific mechanism for reporting SDH. Specifically, the vendor noted concerns with maintaining flexibility in screening tools available to clients and mapping those tools to the same field for analysis. One vendor described strategic development efforts to allow SDH to be included in existing report functions with the goal of better enabling the identification of gaps in care and population management.

In terms of assessing health outcomes, vendors report that measuring both short-term outcomes, such as the completion of the referral, and long-term outcomes, such as changes in costs, utilization, and health outcomes, are difficult both technically and due to challenges addressing SDHs. One vendor has observed clients defining impacts in terms of quality metrics such as reducing readmission rates or reducing emergency department use; it reported that one client assessed outcomes from the person's perspective of their wellness.

To develop better analytic tools, one vendor has developed a proprietary value set which it is analyzing for the development of risk algorithms that incorporate SDH. It has found that SDH indicators are highly concentrated among a third of the clients or that 30% of clients have collected 90% of the SDH that have been found in the data set. Further, it reports that 90% of what is being collected is only for 13 types of SDH measures, namely separation or divorce, death in the family, unemployment, problems living alone, addiction in family, and caregiver roles; less common are issues like homelessness or child abuse.

Coding Standards and Interoperability

Data standards are codes for the capture and exchange of electronic health data that govern and ease their integration with other data sets for analysis and use. Specifically, vendors report the use of *International Statistical Classification of Diseases*

and Related Health Problems, 10th Revision (ICD-10) and accompanying Z-codes, Logical Observation Identifiers Names and Codes (LOINC) and Systematized Nomenclature of Medicine (SNOMED), and current procedural terminology (CPT) codes, which are necessary for the standardized coding of multiple aspects of the patient record [29-31]. To screen for SDH, there are LOINC and SNOMED codes that cover the same SDH domains; to assess or diagnose SDH, there are SNOMED and ICD-10 codes that cover the same diagnosis; to document an intervention on an SDH (ie, making a referral), there are SNOMED and CPT codes that cover the same procedures [32]. In addition to having multiple terminologies of codes, there are also multiple codes within the same SDH domain.

Due to a lack of standardization, vendors described challenges with the multiplicity and ambiguity of coding SDH measures. One vendor explained, "When looking at the ICD-9 codes, there are about 45 codes that can be used for SDH and when you look at cross-walking those there are about 127 codes in SNOMED that link back to a SDH." Another vendor described challenges that emerge from the absence of standard terminology. For example, since LOINC and SNOMED do not provide codes for transportation assistance, practices may use a dummy CPT code to track it.

Vendors reported that even with well-known tools like PRAPARE, vendors must sometimes make idiosyncratic coding decisions. In general, the PRAPARE tool has very specific questions and answers—for example, a click list of options for level of education that can link to LOINC terms for each of the responses. Where the ambiguity arises is mapping questions like, "What is the highest level of school you've finished?" Although the LOINC and SNOMED answer options might be the same, the vendor would not feel comfortable making the decision to code to one terminology over the other. From the vendor perspective, ideally PRAPARE would be hard coded to a single standard to ensure consistency and interoperability.

Further, not all SDH information can be coded, and free-text fields are frequently used. In spite of the tens of thousands of codes among ICD, LOINC, and SNOMED, some vendors commented that a lot of information that is collected cannot be characterized by a given code and falls into free text. One vendor explained:

> *In an ideal world all of this [SDH] information would be collected in a codified way, and there would be a table where they can see all of this information. However, in the world today all of the information can be variable in terms of where and how it is collected. It sometimes comes up in the problem table, but we have not begun to even look at the free-text physician notes section, where they anticipate even more information may be collected.*

Three vendors reported that some depression surveys are challenging to analyze because they combine yes/no questions with free-text fields intended to capture more detailed information about the patient. Clients appreciate being able to capture these explanations from patients via the free text, in spite of the challenges with codifying them.

In terms of interoperability, lack of standards in both what SDH data is collected and how it is coded also makes its exchange among health care providers difficult. While vendors can use the Consolidated Clinical Document Architecture (C-CDA) to make electronic referrals to community service providers and support system-to-system exchange, one vendor explained that the C-CDA does not codify specific SDH data elements.

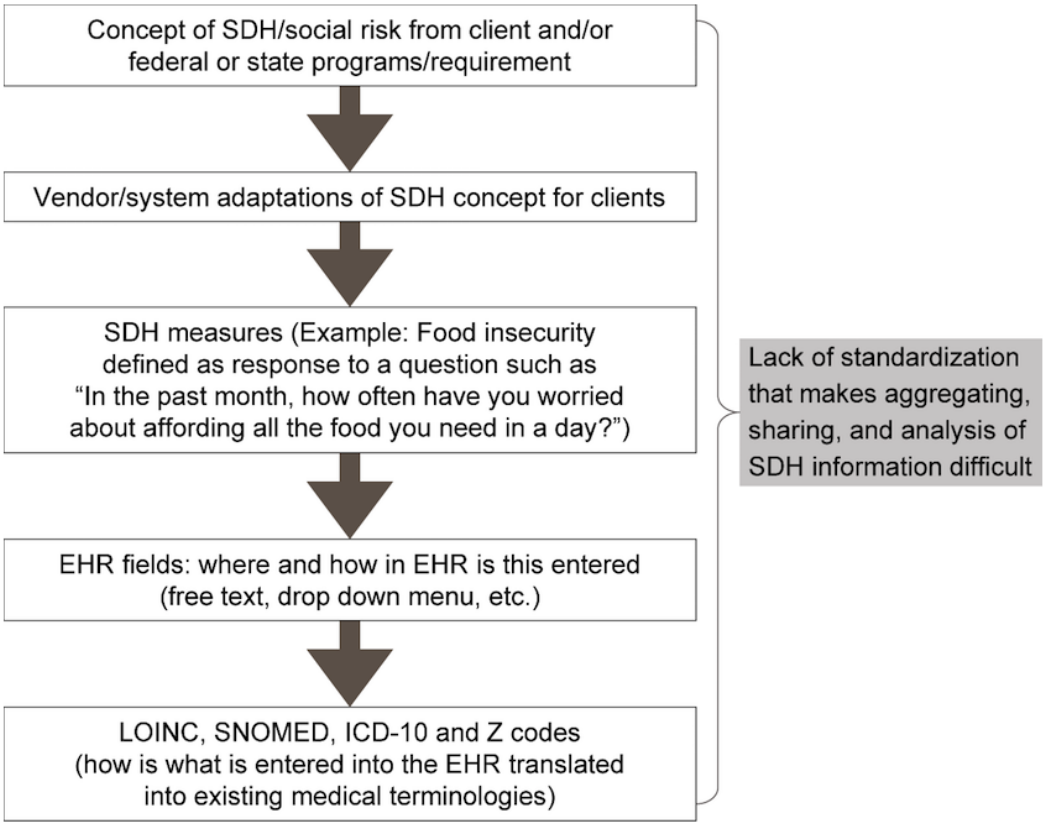
Another vendor reported working on a project with some Regional Health Information Exchange Organizations (RHIOs) interested in receiving SDH data. They are starting with race and ethnicity with the intent of sending additional information as the project develops and anticipate other RHIOs will express similar interest.

Finally, vendors described challenges with analysis of SDH data due to lack of standardization. One vendor spoke of the need to standardize or structure SDH data while preserving client flexibility in its collection. An interviewee explained:

> *If the data is more structured, the analysis is easier. If we have to scale to many clients, with many different screening tools, our job is not to force into one screening tool, but is to normalize the results of the screening tools, so we can map food insecurity tools A and B to the same field that can then be used for analysis. As an IT vendor that kind of data structure is very important.*

Although clearly the benefit of standardization was viewed from the perspective of the potential benefit to the vendor itself, it is understood that generally better standardization would allow health systems to better analyze and interpret SDH data in clinical decision-making. Figure 3 depicts the chain reaction of variability that leads to the lack of standardization and its limits on the use of SDH data in patient care and population health planning.

Figure 3. Systemic variability leading to lack of standardization and usability of social determinants of health data. SDH: social determinants of health; EHR: electronic health record; LOINC: Logical Observation Identifiers Names and Codes; SNOMED: Systematized Nomenclature of Medicine; ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th Revision.



Vendor Recommendations on Standardization

Vendors in this sample indicated that in the absence of national standards, clients are getting "pretty creative" in the collection and use of SDH data. Vendors showed support for discussions among vendors, standards bodies, and government organizations to reduce ambiguity in the code sets, as well as to ensure all voices are heard. Ultimately, several emphasized that vendors must follow the recommendations that public agencies outline and sought direction on standardized tools to collect SDH data, standards for SDH data coding and interoperability, and incentives for SDH data collection and use.

Standardized Tools to Collect Social Determinants of Health Data

Vendors generally agreed that having standardized definitions of SDH across all government programs would improve the field from a research and analytics perspective. It would also help vendors build tools that are more interoperable. Specifically, if different federal programs can agree on a set of measures, it would facilitate more standardization. For example, one individual commented that, "The PRAPARE tool is great, but the private sector does not seem to be open to it, and it is not an exact match to some of the other national programs already, so there is some disconnect there" that leads to the implementation of differing SDH tools across health systems.

Standards for Social Determinants of Health Data Coding and Interoperability

Vendors encouraged the use of standard terminology to enable interoperable exchange of SDH-related data. In some cases, more than one standard is assigned to a particular data element. Vendors would appreciate guidance on the preferred standard to be used for a minimum set of data elements. However, they also caution that not all elements can be codified, and how a specific tool is implemented in the EHR should be at the client's discretion. In particular, this relates to making determinations about the tools that are most useful to their practices, with the recognition that the data they capture must roll up to meet federal reporting standards.

Vendors are involved in discussions and workgroups related to SDH standards that promote data capture and interoperability (Textbox 1). Some participated in national standards development organization activities like the Health Level-7 International C-CDA standards workgroup. Some were involved with nongovernmental initiatives such as one led by the Social Interventions Research & Evaluation Network (SIREN) to improve interoperability of SDH data in EHRs [33]. Finally, vendors continue to engage in industry efforts focused on health information exchange. One vendor reported participation in an industry-wide interoperability initiative called Carequality that grew out of the Sequoia Project [34].

Textbox 1. Two vendors' views on their role in creating coding standards.

> *Yes, we have a role to play [in developing standards for coding social determinants of health data], but we also want to be cognizant of the optics and want other vendors to participate. We don't want to be perceived as commandeering the narrative.*
>
> *It's hard as an [information technology] company to push a standard, because others may perceive it as bias. When an open standard for social determinants is pushed from a national group it is better and that's something we support.*

Incentives for Social Determinants of Health Data Collection and Use

From the demand side, clients drive demand, investment, and more development, as do policies, including incentives and VBC programs. However, vendors wonder whether the incentives will be fair and whether SDH collection is a fad versus a priority with longevity. One vendor posed the question of whether SDH will come to be as large a movement as quality improvement was for health care.

Discussion

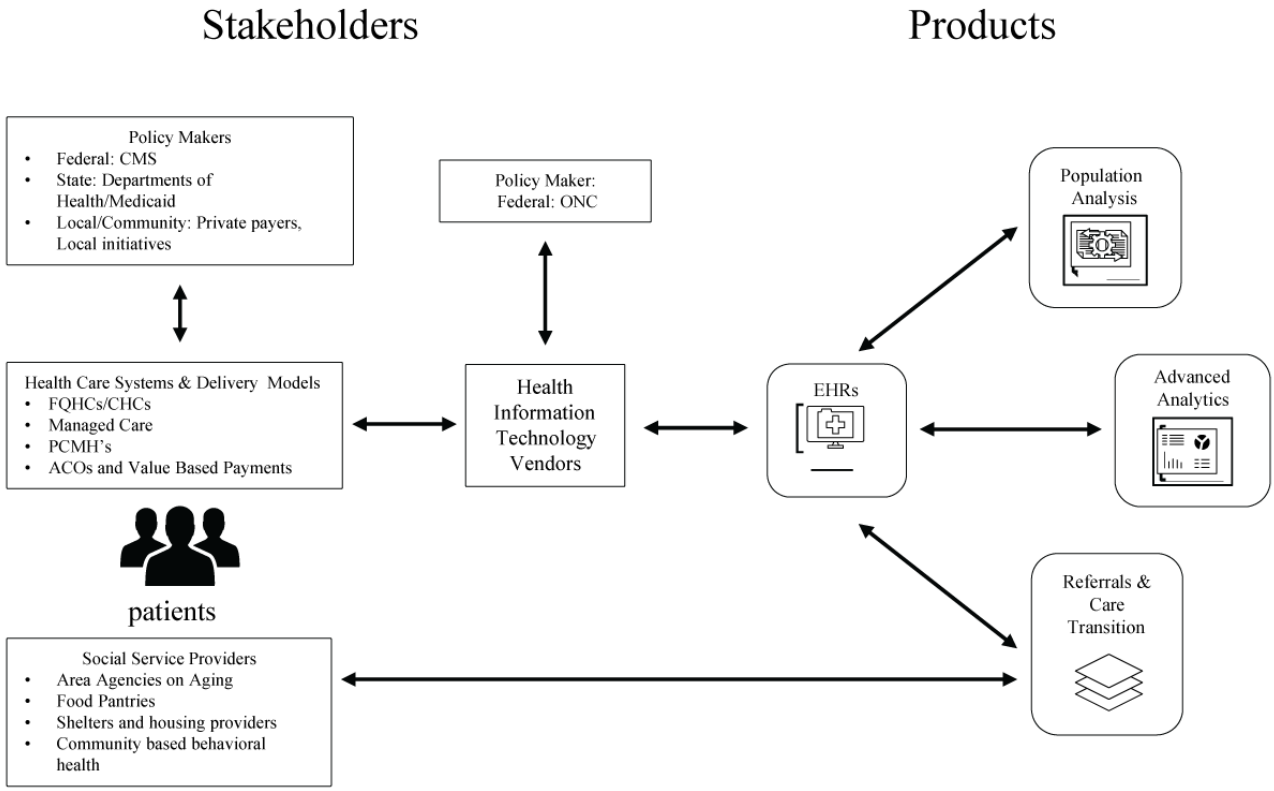
Principal Findings

Vendor systems and their functionalities are the result of the multiple, interrelated forces of federal policy and regulation, client demand, and the vendors' own strategic vision for opportunities in the health care system. Through interviews with vendors, we explored the roles of client demand and federal policies related to SDH capture and use. We also explored issues related to use of standards and interoperable information sharing, use cases for SDH to improve clinical care and processes, and potential avenues for growth in use of SDH data. In doing so, we see the influence of numerous stakeholders—federal, state, and local policy makers; health systems; social services systems; health information technology vendors; and patients—on the development of SDH-related products in EHRs (Figure 4). Health information technology companies that serve as EHR vendors must adhere to federal policies set out by ONC, and health care systems and the delivery models they use must adhere to federal policies set out by CMS (among others) as well as state and local health-related policies. Both types of stakeholders may also have some influence on such policies as well. The SDH-related products that vendors make to enable population analysis, advanced analytics, and referrals and care transitions seek to better integrate care delivered in health care settings with social services outside of those settings, thereby addressing patients' nonmedical needs. Yet many interests, policies, and products need to align in order for this to happen.

In this study, we found that even among vendors with large market shares in both ambulatory and inpatient settings, SDH is a new area for investment, and there is room for growth in terms of product development and analytic capacity. While all vendors interviewed use or have enabled some SDH data collection screening instruments or measures in their EHR platform, they vary in terms of capacity to track referrals and analyze data. Vendors activities also ranged from simply seeking to help clients meet regulatory obligations to those engaged in research to develop products that will help clients better target and address needs, including those related to SDH, of their patients.

Vendors identified a number of challenges primarily with analyzing SDH data and sharing them among health systems. This includes challenges with multiple overlapping but distinct performance metrics and indicators across various federal and state programs, lack of agreement on mapping SDH measures to codes, and lack of codes for all measures. Finally, there is a general problem with interoperability among different health care systems that makes sharing and using SDH data difficult. Vendors appear to have taken a role in resolving these challenges through participation in policy development, standardizing bodies, and vendor-specific solutions and decisions. With the lack of policy regulations around SDH data, product-specific decisions may end up being de facto policies given the market share of particular vendors. However, vendors appear ready for formal policymaking discussions to seek solutions that may further promote the integration of SDH data into mainstream health care delivery.

Figure 4. Stakeholders that inform vendors' social determinants of health–related products in electronic health records. CMS: Centers for Medicare and Medicaid Services; ONC: Office of the National Coordinator for Health Information Technology; FQHC: Federally Qualified Health Center; CHC: community health center; PCMH: patient-centered medical home; ACO: Accountable Care Organization; EHR: electronic health record.



Limitations

The findings from this study are based on a purposive, qualitative sample with a small number of vendors. They are not intended to represent the state of the EHR field at large but rather to help identify trends in the development and use of SDH screening tools and data among vendors with considerable stake in this area given their market shares in inpatient and outpatient settings. The study was also limited to vendors we could reach during a limited study period. With more time, we would have sought more input from representatives working on this increasingly commercialized component of health care systems.

Conclusions

In order to advance the collection and use of SDH data in health care settings through EHRs, the findings from this study suggest at least three next steps:

- Identify core SDH measures where standard development is still needed. For example, since LOINC and SNOMED do not provide codes for transportation assistance, additional code development may be needed.
- Provide guidance on preferred terminology standards for some SDH measures. For example, since education and bereavements have several codes that can be used, providing guidance on preferred terminology would eliminate vendors and health care organizations making idiosyncratic coding choices.
- Identify standards for a subset of SDH measures that health systems can routinely collect through EHRs. Building upon earlier work by ONC to require certified EHRs to collect SDH measures such as race and ethnicity, initiatives to develop standards around specific SDH domains may help encourage their widespread use in EHRs. SIREN and its Gravity Project are a current example of such an effort. This national collaborative seeks to promote interoperable documentation of three priority SDH domains: food security, housing stability and quality, and transportation [35].

This study has shown that in the absence of standardization of SDH screening instruments, measurements, and codification, EHR vendors will provide their clients multiple options and flexible tools to meet their varying needs and interests. We were limited to a small number of vendors that we could reach in a short time frame, but the vendors have large market shares and were consistent in the need to remain adaptable and responsive to client needs and federal and state requirements. They appreciated the potential for standardized SDH data to identify patients with high social need, improve care coordination between health care providers and community service providers, and build further evidence on the connections between SDH and health outcomes through better data analytics and population health management. Vendors and providers seek approaches that balance the use of existing data with the need to collect standardized new data in order to streamline the integration of SDH data in providers' workflow and create a holistic picture of patients that may ultimately reduce health disparities.

Acknowledgments

This report was prepared under contract #HHS23320160020I. The task order number for the current Cost Plus Fixed Fee umbrella contract is HHSP23337001T between the Department of Health and Human Services' Office of the Assistant Secretary for Planning and Evaluation Office of Health Policy and NORC at the University of Chicago. The views expressed in this article are the authors' and do not necessarily represent the views of the US Department of Health and Human Services or the United States Government. The authors would like to thank Andrew Wang at Northwestern Medicine Health Sciences Integrated PhD Program for his input on this project and Jasmine Leonard, Andrew Card, and Krysta Heaney-Huls at NORC at the University of Chicago for their input on the preparation of this manuscript.

Conflicts of Interest

None declared.

References

1. Burwell SM. Setting value-based payment goals—HHS efforts to improve U.S. health care. N Engl J Med 2015 Mar 5;372(10):897-899. [doi: 10.1056/NEJMp1500445] [Medline: 25622024]
2. Azar A. Remarks on Value-Based Transformation and Innovation. 2018 May 02. URL: https://www.hhs.gov/about/leadership/secretary/speeches/2018-speeches/remarks-on-value-based-transformation-and-innovation.html [accessed 2019-05-20] [WebCite Cache ID 78VEOHUJj]
3. Healthy People 2020: social determinants of health.: Office of Disease Prevention and Health Promotion URL: https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health [accessed 2019-04-22] [WebCite Cache ID 77piRKRhT]
4. Institute of Medicine of the National Academies. Capturing Social And Behavioral Domains In Electronic Health Records: Phase 1. Washington, DC: National Academies Press; Jun 2014.
5. Health IT in the Quality Payment Program.: National Partnership for Women & Families; 2016. URL: http://www.nationalpartnership.org/research-library/health-care/HIT/health-it-in-the-quality-payment-program.pdf [accessed 2018-11-09] [WebCite Cache ID 73o8f5J6N]
6. Machledt D. Addressing the social determinants of health through Medicaid managed care.: Commonwealth Fund URL: http://www.commonwealthfund.org/~/media/files/publications/issue-brief/2017/nov/machledt_social_determinants_medicaid_managed_care_ib_v2.pdf [accessed 2019-05-20] [WebCite Cache ID 73oI0RywS]
7. Spencer A, Freda B, Mcginnis T, Gottlieb L. Measuring social determinants of health among Medicaid beneficiaries: early state lessons. 2016. URL: http://www.chcs.org/media/CHCS-SDOH-Measures-Brief_120716_FINAL.pdf [accessed 2018-11-09] [WebCite Cache ID 73oJPGsCB]
8. De Milto L, Nakashian M. Using social determinants of health data to improve health care and health: a learning report.: Robert Wood Johnson Foundation; 2016. URL: https://www.rwjf.org/content/dam/farm/reports/reports/2016/rwjf428872 [accessed 2019-05-20] [WebCite Cache ID 73oJemUbx]
9. Gottlieb L, Cottrell EK, Park B, Clark KD, Gold R, Fichtenberg C. Advancing social prescribing with implementation science. J Am Board Fam Med 2018;31(3):315-321 [FREE Full text] [doi: 10.3122/jabfm.2018.03.170249] [Medline: 29743213]
10. Social need screening tools comparison table. University of California, San Francisco: Social Interventions Research & Evaluation Network URL: https://sirenetwork.ucsf.edu/tools-resources/mmi/screening-tools-comparison/adult-nonspecific [accessed 2018-11-09] [WebCite Cache ID 73oJm2YGR]
11. Institute of Medicine of the National Academies. Capturing Social And Behavioral Domains And Measures In Electronic Health Records: Phase 2. Washington, DC: National Academies Press; Jan 2015.
12. The protocol for responding to and assessing patients' assets, risks, and experiences (PRAPARE).: National Association of Community Health Centers, Association of Asian Pacific Community Health Organizations, Oregon Primary Care Association; 2017. URL: http://www.nachc.org/research-and-data/prapare/ [accessed 2018-11-09] [WebCite Cache ID 73oK8ZjBs]
13. Smith-Bindman R. Is computed tomography safe? N Engl J Med 2010 Jul 01;363(1):1-4 [FREE Full text] [doi: 10.1056/NEJMp1002530] [Medline: 20573919]
14. LaForge K, Gold R, Cottrell E, Bunce AE, Proser M, Hollombe C, et al. How 6 Organizations developed tools and processes for social determinants of health screening in primary care: an overview. J Ambul Care Manage 2018;41(1):2-14 [FREE Full text] [doi: 10.1097/JAC.0000000000000221] [Medline: 28990990]
15. Cantor MN, Thorpe L. Integrating data on social determinants of health into electronic health records. Health Aff (Millwood) 2018 Dec;37(4):585-590. [doi: 10.1377/hlthaff.2017.1252] [Medline: 29608369]
16. Meaningful Use: public health–EHR vendors collaboration initiative.: Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/ehrmeaningfuluse/Public-Health-EHR-Vendors-Collaboration-Initiative.html [accessed 2018-11-09] [WebCite Cache ID 73oKmJqle]

17. Integrated health model initiative.: American Medical Association; 2018. URL: https://ama-ihmi.org [accessed 2018-11-13] [WebCite Cache ID 73uLB5SKZ]
18. Tripathi M. EHR evolution: policy and legislation forces changing the EHR. J AHIMA 2012 Oct;83(10):24-29. [Medline: 23061349]
19. Glaser J. From the electronic health record to the electronic health plan.: H&HN; 2015. URL: https://www.hhnmag.com/articles/3272-from-the-electronic-health-record-to-the-electronic-health-plan [accessed 2018-11-09] [WebCite Cache ID 73oL2hwx3]
20. Bresnick J. Faulkner guides Epic Systems from EHR to comprehensive health record. 2018. URL: https://healthitanalytics.com/features/faulkner-guides-epic-systems-from-ehr-to-comprehensive-health-record [accessed 2018-11-09] [WebCite Cache ID 73oLVTN0Z]
21. Sullivan T. Epic's rival EHR vendors say they too are making the CHR switch. 2017. URL: https://www.healthcareitnews.com/news/epics-rival-ehr-vendors-say-they-too-are-making-chr-switch [accessed 2018-11-09] [WebCite Cache ID 73oM9DrHC]
22. Koppel R, Lehmann CU. Implications of an emerging EHR monoculture for hospitals and healthcare systems. J Am Med Inform Assoc 2015 Mar;22(2):465-471. [doi: 10.1136/amiajnl-2014-003023] [Medline: 25342181]
23. Certified Health IT developers and editions reported by hospitals participating in the Medicare EHR Incentive Program.: Office of the National Coordinator for Health Information Technology; 2017. URL: https://dashboard.healthit.gov/quickstats/pages/FIG-Vendors-of-EHRs-to-Participating-Professionals.php [accessed 2018-11-09] [WebCite Cache ID 73oMVauoc]
24. Certified Health IT developers and editions reported by ambulatory primary care physicians, medical and surgial specialists, podiatrists, optometrists, dentists, and chiropractors participating in the Medicare EHR Incentive Program.: Office of the National Coordinator for Health Information Technology; 2017. URL: https://dashboard.healthit.gov/quickstats/pages/FIG-Vendors-of-EHRs-to-Participating-Hospitals.php [accessed 2018-11-09] [WebCite Cache ID 73oNLXVYH]
25. The Columbia-Suicide Severity Rating Scale (C-SSRS).: Columbia Light Project URL: http://cssrs.columbia.edu/the-columbia-scale-c-ssrs/about-the-scale/ [accessed 2018-11-10] [WebCite Cache ID 73oOVcGmq]
26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
27. Social vulnerablity index.: Centers for Disease Control and Prevention URL: https://svi.cdc.gov/ [accessed 2018-11-13] [WebCite Cache ID 73uNeAmwU]
28. Aunt Bertha. URL: https://company.auntbertha.com/ [accessed 2018-11-10] [WebCite Cache ID 73oOlJf8Y]
29. Factors influencing health status and contact with health services Z00-Z99. URL: https://www.icd10data.com/ICD10CM/Codes/Z00-Z99 [accessed 2018-11-10] [WebCite Cache ID 73oOrejki]
30. International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM).: Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/nchs/icd/icd10cm.htm [accessed 2018-11-10] [WebCite Cache ID 73oP3Hhh1]
31. Supporting interoperability: terminology, subsets, and other resources from NLM.: US National Library of Medicine; 2018. URL: https://www.nlm.nih.gov/hit_interoperability.html [accessed 2018-11-10] [WebCite Cache ID 73oPKQxxa]
32. Arons A, DeSilvey S, Fichtenberg C, Gottlieb L. Compendium of Medical Terminology Codes for Social Risk Factors. San Francisco, CA: Social Interventions Research and Evaluation Network; 2018. URL: https://sirenetwork.ucsf.edu/tools-resources/mmi/compendium-medical-terminology-codes-social-risk-factors [accessed 2019-04-23] [WebCite Cache ID 77qt5EeAg]
33. Arons A, Desilvey S, Fichtenberg C, Gottlieb L. Improving the interoperability of social determinants data in electronic health records: working paper for November 2017 expert panel convening. Social Interventions Research & Evaluation Network Stakeholder Meeting 2017 Nov 9:1-26.
34. The Sequoia Project. URL: https://sequoiaproject.org/ [accessed 2018-11-10] [WebCite Cache ID 73oPQTMe5]
35. The Gravity Project: a national collaborative to advance interoperable social risk and protective factors documentation.: Social Interventions Research and Evaluation Network URL: https://sirenetwork.ucsf.edu/TheGravityProject [accessed 2019-05-20] [WebCite Cache ID 77qtNtl0z]

Abbreviations

ACO: Accountable Care Organization
AHC: Accountable Health Community
C-CDA: Consolidated Clinical Document Architecture
CMS: Centers for Medicare & Medicaid Services
CPC+: Comprehensive Primary Care Plus
CPT: current procedural terminology
EHR: electronic health record
FQHC: federally qualified health centers
ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th Revision
LOINC: Logical Observation Identifiers Names and Codes

NACHC: National Association of Community Health Center
NAM: National Academies of Medicine
ONC: Office of the National Coordinator for Health Information Technology
PRAPARE: Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences
RHIO: Regional Health Information Exchange Organization
SDH: social determinants of health
SIREN: Social Interventions Research & Evaluation Network
SNOMED: Systematized Nomenclature of Medicine
VBC: value-based care

Edited by J Hefner; submitted 28.02.19; peer-reviewed by N Kashyap, G Sanders; comments to author 21.03.19; revised version received 30.04.19; accepted 01.05.19; published 07.06.19.

Please cite as:
Freij M, Dullabh P, Lewis S, Smith SR, Hovey L, Dhopeshwarkar R
Incorporating Social Determinants of Health in Electronic Health Records: Qualitative Study of Current Practices Among Top Vendors
JMIR Med Inform 2019;7(2):e13849
URL: http://medinform.jmir.org/2019/2/e13849/
doi:10.2196/13849
PMID:31199345

Original Paper

Patient-Sharing Relations in the Treatment of Diabetes and Their Implications for Health Information Exchange: Claims-Based Analysis

Georg Duftschmid[1], PhD; Christoph Rinner[1], PhD; Simone Katja Sauter[1], MSc; Gottfried Endel[2], MD; Peter Klimek[3,4], PhD; Christoph Mitsch[5], MD; Harald Heinzl[6], PhD

[1]Section for Medical Information Management, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria

[2]Main Association of Austrian Social Security Institutions, Vienna, Austria

[3]Section for Science of Complex Systems, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria

[4]Complexity Science Hub Vienna, Vienna, Austria

[5]Department of Ophthalmology and Optometry, Medical University of Vienna, Vienna, Austria

[6]Section for Clinical Biometrics, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria

Corresponding Author:
Georg Duftschmid, PhD
Section for Medical Information Management
Center for Medical Statistics, Informatics, and Intelligent Systems
Medical University of Vienna
Spitalgasse 23
Vienna, A-1090
Austria
Phone: 43 1 40400 ext 66960
Fax: 43 1 40400 66970
Email: georg.duftschmid@meduniwien.ac.at

Abstract

Background: Health information exchange (HIE) among care providers who cooperate in the treatment of patients with diabetes mellitus (DM) has been rated as an important aspect of successful care. Patient-sharing relations among care providers permit inferences about corresponding information-sharing relations.

Objectives: This study aimed to obtain information for an effective HIE platform design to be used in DM care by analyzing patient-sharing relations among various types of care providers (ToCPs), such as hospitals, pharmacies, and different outpatient specialists, within a nationwide claims dataset of Austrian DM patients. We focus on 2 parameters derived from patient-sharing networks: (1) the principal HIE partners of the different ToCPs involved in the treatment of DM and (2) the required participation rate of ToCPs in HIE platforms for the purpose of effective communication.

Methods: The claims data of 7.9 million Austrian patients from 2006 to 2007 served as our data source. DM patients were identified by their medication. We established metrics for the quantification of our 2 parameters of interest. The principal HIE partners were derived from the portions of a care provider's patient-sharing relations with different ToCPs. For the required participation rate of ToCPs in an HIE platform, we determine the concentration of patient-sharing relations among ToCPs. Our corresponding metrics are derived in analogy from existing work for the quantification of the continuity of care.

Results: We identified 324,703 DM patients treated by 12,226 care providers; the latter were members of 16 ToCPs. On the basis of their score for 2 of our parameters, we categorized the ToCPs into *low*, *medium*, and *high*. For the *most important HIE partner* parameter, pharmacies, general practitioners (GPs), and laboratories were the representatives of the top group, that is, our care providers shared the highest numbers of DM patients with these ToCPs. For the *required participation rate of type of care provide (ToCP) in HIE platform* parameter, the concentration of DM patient-sharing relations with a ToCP tended to be inversely related to the ToCPs member count.

Conclusions: We conclude that GPs, pharmacies, and laboratories should be core members of any HIE platform that supports DM care, as they are the most important DM patient-sharing partners. We further conclude that, for implementing HIE with

ToCPs who have many members (in Austria, particularly GPs and pharmacies), an HIE solution with high participation rates from these ToCPs (ideally a nationwide HIE platform with obligatory participation of the concerned ToCPs) seems essential. This will raise the probability of HIE being achieved with any care provider of these ToCPs. As chronic diseases are rising because of aging societies, we believe that our quantification of HIE requirements in the treatment of DM can provide valuable insights for many industrial countries.

(JMIR Med Inform 2019;7(2):e12172) doi:10.2196/12172

KEYWORDS

health information exchange; professional-patient relations; diabetes mellitus; Austria

Introduction

Background

Health information exchange (HIE) has been found to improve quality of care in general [1], and it constitutes an important aspect of the successful treatment of diabetes mellitus (DM) in particular [2-6]. Common HIE solutions are disease-specific DM information systems [2,6], as well as general HIE platforms at the regional [7] or national [8] level.

The characteristics of DM patient-sharing relations among care providers can provide useful information for the design and selection of an efficient HIE solution for the treatment of DM. According to Barnett et al, the likelihood of 2 care providers having an information-sharing relation increases with the number of patients they share [9]. Generalizing this finding from individual care providers to *types of care providers* (ToCPs), such as general practitioners (GPs), pharmacies, hospitals, and different types of specialists, we reason that those ToCPs with whom care providers share most of their DM patients should be integral parts of HIE solutions for DM treatment. Furthermore, the concentration of DM patient-sharing relations among different ToCPs allows inferences about the required participation rates of ToCPs in HIE solutions. As we will explain in the section entitled *Measurement of the required participation rates of ToCPs in an HIE solution*, concentration values and required participation rates are inversely related.

Earlier work addressed the associations of patient-sharing relations with health care expenditure, utilization, quality of care [10], interacting drug prescriptions [11], as well as medication costs and patient health status [12]. To our knowledge, patient-sharing networks (PSNs) have not yet been analyzed to gain insights for HIE solutions, except for an earlier study we performed on the use of the Austrian electronic health record system ELGA (acronym for German *Elektronische Gesundheitsakte*) [13].

Objectives

This study spans across a wider range of aspects than the study by Sauter et al [13]: we now consider all ToCPs who provided health services to our patient cohort rather than considering primary care physicians alone. We also added an analysis of the concentration of patient-sharing relations among different ToCPs, and we suggest a corresponding metric for this purpose. We now focus on DM to the extent that DM care teams depend on HIE [2-6], and thus minimize the inclusion of data concerning random relations among care providers who treat the same patient for unrelated reasons.

We present, for the first time, a comprehensive quantification of the requirements for HIE in the treatment of DM patients, on the basis of a nationwide dataset. The aims of this report were the following:

1. To identify the most important HIE partners of the different ToCPs involved in DM treatment on a large scale by analyzing DM patient-sharing relations among care providers on a nationwide basis in Austria.
2. To identify the required participation rate of ToCPs in HIE solutions to achieve effective communication among care providers in the context of DM management, by analyzing the concentration of DM patient-sharing relations among ToCPs.

These characteristics of DM patient-sharing relations can serve as input for the design of HIE solutions, and they can serve as a decision-making aid for care providers in selecting the most suitable HIE platform when several competing platforms exist in their area. We aim to obtain information concerning the required participants and participation rates of HIE solutions for the treatment of DM. Details of implementation, such as system architecture, interface design, or security mechanisms are not included in the scope of this work.

Methods

Data Source

Deidentified claims data of the Main Association of Austrian Social Security Institutions constituted our data source. These included outpatient (GPs, specialists, and pharmacies) as well as inpatient care (see Table 1 for numbers of care providers per type of care provider, ToCP) of 7.9 million persons from all age groups, who were insured by one of the public Social Security Institutions in Austria and had one or more contacts with a care provider between 2006 and 2007. Around 95% of the Austrian population at the time are covered by the database. The small gap results from a few insurance carriers not covered by the database and patients excluded because of inconsistent data for gender or year of birth.

This study was approved by the ethics committee of the Medical University of Vienna (#1903/2017).

All database queries and calculations were implemented using PostgreSQL version 9.4 (PostgreSQL Global Development Group). In total, our SQL script had 858 lines. The calculation of our concentration metrics usual provider cooperation (UPCo) and concentration of cooperation (COCo) is shown in Multimedia Appendix 1.

Identification of Study Patients

The Austrian health care system does not prescribe the documentation of outpatient diagnoses for reimbursement purposes. Therefore, we identified DM patients on the basis of their medication. In other words, we focused on DM patients undergoing pharmaceutical treatment.

Patients were eligible when at least two diabetes-specific medications had been dispensed to them between 2006 and 2007. In accordance with Chini et al, 9 ATC (Anatomical Therapeutic Chemical Classification System with Defined Daily Doses) codes of the groups *A10A: insulins and analogues* and *A10B: oral antidiabetics* were considered diabetes-specific [14].

Patients below 20 years of age in 2006 were excluded (0.90% (2964/329,313) of our cohort) as, in contrast to older patients, this age group could not be validated well in comparison with a reference population [4]. Patients with missing data about their age (0.49% [1646/329,313] of our cohort) were also excluded.

Identification of Study Care Providers

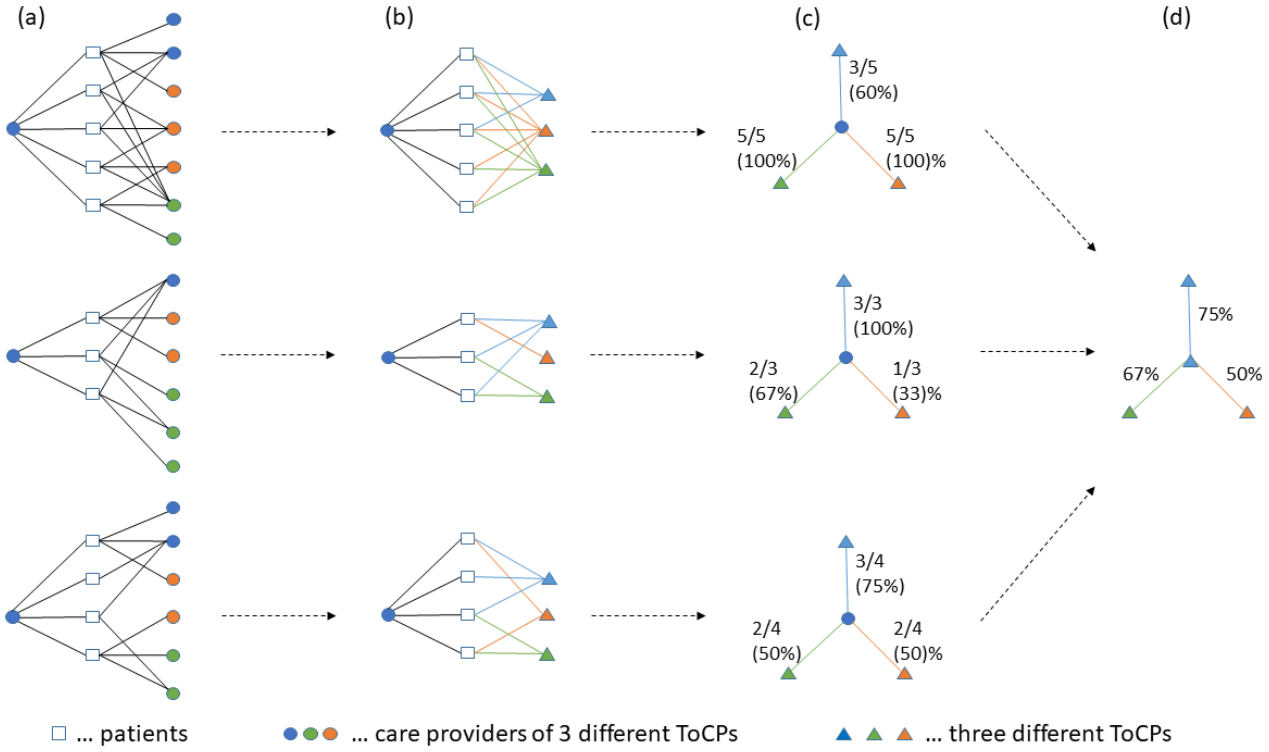
We considered all public care providers (those having a contract with a public Austrian Social Security Institution) who provided one or more services to this study's patients between 2006 and 2007. The ToCP was known in each case. Hospital data were available for inpatient visits but not for walk-in clinics (such as DM ambulances). Dentists were not considered as a substantial part of their services is paid privately by patients, and we lacked the corresponding data.

Measuring the Need for Health Information Exchange Among Types of Care Providers

Our basic assumption is that the need for HIE with a ToCP is directly related to the number of patients shared with the ToCP. The underlying rationale is that the more patients a care provider shares with other care providers of a particular ToCP, the more external information is generated by this ToCP for the care provider's patients, and the more important it becomes for the care provider to establish HIE with the respective ToCP.

We originate from the PSNs [15] of each care provider. Compared with the study by Landon et al [15], the PSNs are *reduced* to the patient-sharing relations between the observed *index* care provider and other *linked* care providers, as the patient-sharing relations among the linked care providers are not relevant for our metrics. Figure 1 explains how we derived patient-sharing portions among ToCPs from the PSNs.

Figure 1. A total of 3 reduced patient-sharing networks (PSNs) are presented. (a) The care providers on the left of each PSN are referred to as the index care providers, whereas the care providers on the right of each PSN are termed linked care providers. An edge between a patient and a linked care provider means that this patient is shared between the index care provider and the linked care provider. Hence, the PSNs show 3 index health care providers of the same ToCP (type of care provider; blue) sharing patients with linked care providers from 3 ToCPs (blue, orange, green). (b) Linked care providers are aggregated per ToCP (shown as colored triangles); an edge between a patient and a linked ToCP means that this patient is shared among the index care provider and at least one linked care provider of this ToCP. Patient-sharing relations are colored according to the linked care providers' ToCPs. (c) Shared patients per linked ToCP are calculated; edges depict the proportions (percentages) of patients shared by the index care provider with each linked ToCP. (d) Index care providers are aggregated per ToCP (here only blue is available); edges depict the typical (median) percentages of patients shared with linked ToCPs (here blue, orange, and green).



Measuring the Required Participation Rates of Types of Care Providers in a Health Information Exchange Solution

We see the concentration of DM patient-sharing relations among ToCPs as an indicator that is inversely related to the ToCPs' required participation rates in an HIE solution. If, for example, each DM patient of a care provider gets her medication at a different pharmacy, a broad participation of pharmacies in the HIE will be necessary to allow the exchange of medication data with practically any pharmacy that a DM patient might visit. However, if a care provider's DM patients are referred to a small number of laboratories, the care provider might get by with an HIE solution with restricted participation of laboratories as long as it covers the laboratories typically visited by her DM patients.

In this study, 2 care providers who share 1 or more patients are referred to as *cooperating* care providers. For measuring the COCo among care providers, we were inspired by the existing metrics for the quantification of the *continuity of care* [16,17]. From the *usual provider continuity (UPC)* [17] that is the most frequently applied continuity of care (COC) index in literature [18], we derived our new metric *UPCo* by exchanging UPC's patient contacts with patient-sharing relations. The new metric is then defined in the following manner:

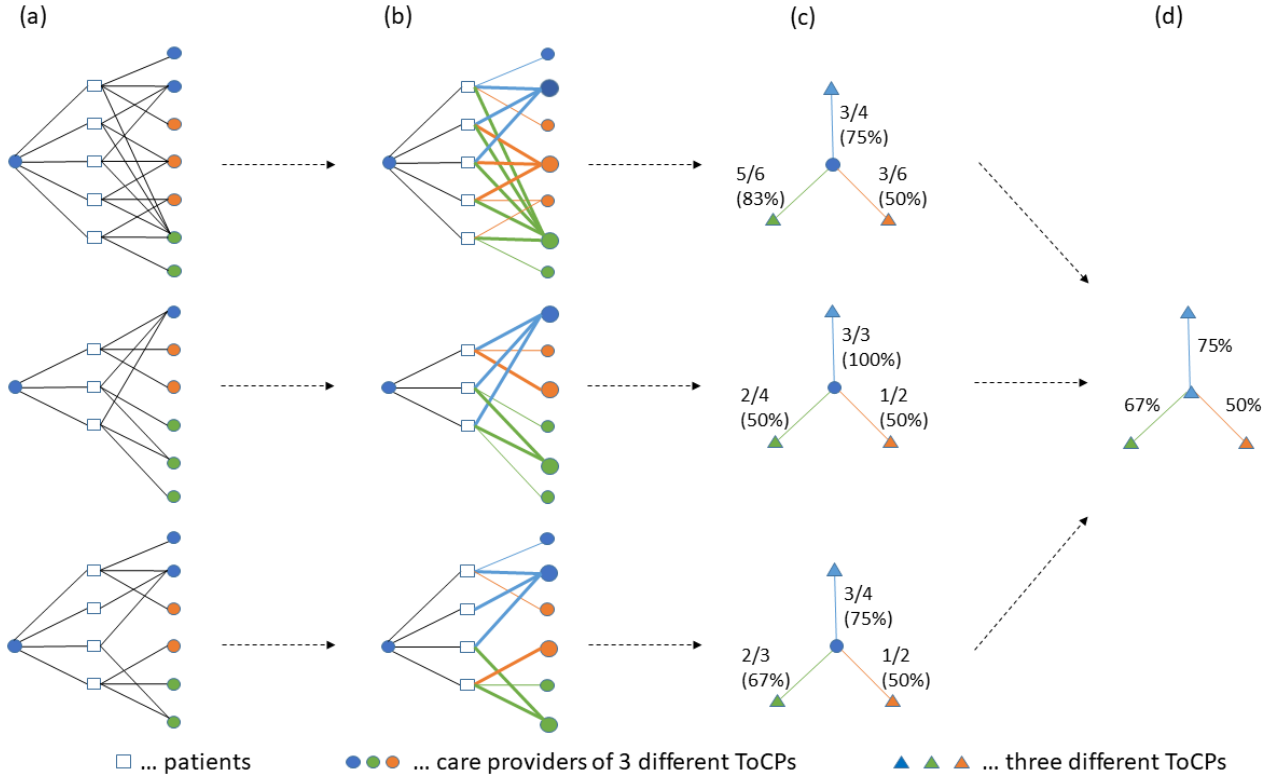
$$UPCo = n_i / N$$

n_i is the number of patients the index care provider shares with the *usual linked care provider*, and N is the total number of patient-sharing relations between the index care provider and all linked care providers in a specific time period. We determine a usual linked care provider for each ToCP and define the usual linked care provider for ToCP X (following a typical UPC procedure) as the linked care provider of ToCP X with whom the index care provider shares the highest number of her patients. If 2 or more linked care providers share the same maximum number of patients with the index care provider, 1 of them is arbitrarily chosen as the usual linked care provider, whereas the others are treated as regular linked care providers. The UPCo for ToCP X then reflects the percentage of an index care provider's patient-sharing relations within ToCP X that are associated with the usual linked care provider of ToCP X.

Figure 2 explains how we derived typical UPCo values among ToCPs from the PSNs.

As a crosscheck of the robustness of our UPCo measurements, we additionally calculated a second metric derived from the frequently used COC [16]. The corresponding results were very similar to the UPCo measurements (see Multimedia Appendix 2 for details).

Figure 2. Calculation of typical usual provider cooperation (UPCo) values. (a) We originate from the same patient-sharing network as shown in Figure 1 a. (b) Index care providers' usual linked care providers (shown as enlarged dots) are determined separately for each ToCP (type of care provider), thereby ties are broken randomly. Patient-sharing relations are colored according to the linked care providers' ToCPs; relations with the usual linked care provider are emphasized additionally. (c) UPCo values (depicted on edges both as proportions and percentages) are calculated for each ToCP. By way of an example, the UPCo value of 75% in the topmost network indicates that 3 out of 4 patient-sharing relations between the first index care provider and the linked blue care providers are shared with the usual linked blue care provider. (d) Index care providers are aggregated per ToCP; edges depict the typical (median) UPCo values with linked ToCPs. Trivial UPCo values of 100% in case of an index care provider who shared only 1 single patient with only 1 single-linked care provider within a particular ToCP were not considered in our analysis. This was the case in 5.7% of the patient-sharing relations between index care providers and linked ToCPs.



Results

Identified Study Patients

We identified 324,703 DM patients (3.92% (324,703/8,280,711) of the mean Austrian population between 2006 and 2007) who satisfied our inclusion and exclusion criteria (see section *Identification of study patients*).

Identified Study Care Providers

We identified 12,226 care providers who satisfied our inclusion and exclusion criteria (see section entitled *Identification of study care providers*). Table 1 shows the distribution of these providers across the different ToCPs.

Need for Health Information Exchange Among Types of Care Providers

Figure 3 shows the results of the procedure described in Figure 1, that is, the median portions of DM patients shared by care providers of any 2 ToCPs.

The rows of Figure 3 show median portions of patients *shared by an index care provider of a particular ToCP*. As an example, the row entitled *GP* shows how many of her patients a GP typically shares with linked care providers of each ToCP. The columns of Figure 3 show median portions of patients *shared with a linked care provider of a particular ToCP*. For instance, the column entitled *GP* shows how many of their patients the index care providers of each ToCP typically share with a GP.

The bottom right cell of Figure 3 shows that the average median percentage of DM patients shared by our index care providers with any single-linked ToCP was 41%. According to the right-most column entitled *Mean*, the corresponding range was between 35% (GPs, pharmacies) and 46% (physical medicine).

According to the bottom row entitled *Mean*, linked ToCPs differed strongly in the average median percentages of patients that index care providers shared with them. We grouped the linked ToCPs according to these portions and assigned each of them to 1 of the 3 categories: (0%-33%), (33%-66%), and (66%-99%), on the basis of the observed range. Pharmacies (99%), GPs (97%), and laboratories (86%) were located in the top category. Radiologists (62%), ophthalmologists (58%), hospitals (56%), and internal medicine specialists (46%) were assigned to the middle category. All other linked ToCPs belonged to the bottom category.

We noted similar values within each column of Figure 3, except for the main diagonal. This means that the number of patients shared with any single-linked ToCP was similar for all index ToCPs. In most cases, cells along the main diagonal of Figure 3 contain low values compared with the other values in their column. In other words, care providers usually share more of their patients with care providers of other ToCPs than with their own ToCP.

Required Participation Rates of Types of Care Providers

Figure 4 shows the results of the procedure described in Figure 2, that is, the median UPCo values among care providers of any 2 ToCPs.

The bottom right cell of Figure 4 shows that the average median concentration of our index care providers' shared patients on the usual linked care provider of any single-linked ToCP was 34%. The strongest deviations (compare right-most column *mean*) from this average occurred for laboratories (17%) and pharmacies (44%).

Table 1. Numbers of study care providers per type of care provider.

Type of care provider	n
General Practitioner	4892
Pharmacy	2240
Internal medicine	949
Gynecology	778
Ophthalmology	510
Surgery	441
Orthopedics	407
Neurology/psychiatry	391
Dermatology	329
Otolaryngology	299
Radiology	277
Urology	258
Pulmology	165
Hospital	132
Laboratory	102
Physical medicine	56

Figure 3. Typical (median) portion of patients shared by an index care provider of the ToCP shown in the leftmost column, with linked care providers of the ToCP shown in the topmost row. ToCPs are sorted by the number of care providers per ToCP (compare Table 1). Cells are color-coded with colors ranging from dark green for high values to dark red for low values. Intermediate values are shown in graded color intensities. The bottom row and the right-most column show the mean value of the corresponding column's respectively row's values. As a measure of variance, interquartile ranges are given in Multimedia Appendix 3. GP: general practitioner; pharm: pharmacy; int med: internal medicine; gynaec: gynecology; ophthalm: ophthalmology; surg: surgery; orthop: orthopedics; neurol/psych: neurology/psychiatry; dermat: dermatology; otolaryng: otolaryngology; radiol: radiology; urol: urology; pulm: pulmology; hosp: hospital; lab: laboratory; phys med: physical medicine.

Shared patients																	
ToCP	GP	pharm.	int. med.	gynaec.	ophthalm.	surg.	orthop.	neurol./psych.	dermat.	otolaryng.	radiol.	urol.	pulm. spec.	hosp.	lab	phys. med.	Mean
GP	76%	100%	33%	12%	51%	5%	15%	14%	20%	18%	46%	17%	11%	58%	77%	0%	35%
pharm.	95%	91%	33%	12%	51%	6%	14%	13%	18%	18%	45%	18%	11%	56%	78%	0%	35%
int. med.	97%	100%	24%	15%	63%	7%	22%	16%	27%	26%	63%	25%	15%	57%	91%	0%	41%
gynaec.	100%	100%	46%	4%	68%	8%	25%	16%	29%	22%	83%	8%	12%	50%	92%	0%	41%
ophthalm.	98%	100%	44%	16%	12%	8%	19%	14%	24%	25%	55%	22%	12%	54%	84%	0%	37%
surg.	100%	100%	53%	17%	65%	3%	25%	14%	29%	27%	78%	25%	9%	62%	93%	0%	44%
orthop.	99%	100%	55%	20%	65%	11%	16%	22%	33%	29%	85%	25%	17%	58%	87%	0%	45%
neurol./psych.	100%	100%	50%	17%	59%	8%	25%	16%	26%	27%	61%	23%	16%	64%	86%	0%	42%
dermat.	98%	100%	49%	19%	64%	11%	25%	17%	14%	29%	60%	26%	16%	54%	86%	0%	42%
otolaryng.	98%	100%	50%	16%	65%	11%	24%	20%	33%	11%	62%	27%	19%	57%	86%	0%	42%
radiol.	98%	100%	49%	24%	62%	12%	30%	18%	28%	26%	40%	23%	17%	56%	89%	2%	42%
urol.	98%	100%	51%	6%	62%	11%	19%	17%	29%	28%	58%	6%	16%	57%	92%	0%	41%
pulm. spec.	99%	100%	57%	17%	63%	12%	26%	20%	30%	35%	67%	28%	6%	62%	88%	0%	44%
hosp.	98%	100%	36%	11%	52%	6%	14%	17%	19%	21%	44%	19%	12%	47%	76%	0%	36%
lab	98%	100%	47%	15%	57%	10%	24%	15%	28%	24%	62%	24%	15%	53%	85%	2%	41%
phys. med.	98%	100%	52%	23%	63%	13%	42%	20%	34%	29%	83%	25%	16%	55%	84%	2%	46%
Mean	**97%**	**99%**	**46%**	**15%**	**58%**	**9%**	**23%**	**17%**	**26%**	**25%**	**62%**	**21%**	**14%**	**56%**	**86%**	**0%**	**41%**

Figure 4. Typical (median) UPCo values of an index care provider of the ToCP shown in the left-most column with a linked care provider of the ToCP shown in the topmost row. ToCPs are sorted by the number of care providers per ToCP (compare Table 1). Cells are color-coded with colors ranging from dark green for high values to dark red for low values. Intermediate values are shown in graded color intensities. The bottom row and the right-most column show the mean value of the corresponding column's respectively row's values. As a measure of variance, interquartile ranges are given in Multimedia Appendix 3. GP: general practitioner; pharm: pharmacy; int med: internal medicine; gynaec: gynecology; ophthalm: ophthalmology; surg: surgery; orthop: orthopedics; neurol/psych: neurology/psychiatry; dermat: dermatology; otolaryng: otolaryngology; radiol: radiology; urol: urology; pulm: pulmology; hosp: hospital; lab: laboratory; phys med: physical medicine.

UPCo																	
ToCP	GP	pharm.	int. med.	gynaec.	ophthalm.	surg.	orthop.	neurol./psych.	dermat.	otolaryng.	radiol.	urol.	pulm. spec.	hosp.	lab	phys. med.	mean
GP	21%	24%	36%	36%	38%	39%	44%	34%	43%	40%	48%	42%	44%	48%	48%	67%	**41%**
pharm.	19%	19%	36%	33%	37%	49%	47%	50%	48%	50%	45%	51%	57%	47%	48%	60%	**44%**
int. med.	10%	17%	33%	33%	33%	46%	40%	49%	39%	44%	42%	50%	50%	45%	47%	58%	**40%**
gynaec.	9%	16%	33%	33%	33%	48%	41%	45%	40%	47%	41%	50%	52%	43%	40%	56%	**39%**
ophthalm.	6%	14%	27%	25%	30%	41%	36%	43%	33%	41%	38%	44%	47%	41%	41%	62%	**36%**
surg.	14%	17%	33%	33%	33%	35%	33%	39%	34%	35%	38%	43%	50%	45%	40%	67%	**37%**
orthop.	6%	14%	25%	25%	25%	39%	30%	38%	31%	36%	38%	40%	44%	38%	34%	54%	**32%**
neurol./psych.	8%	15%	31%	28%	33%	40%	40%	34%	36%	37%	40%	42%	40%	40%	39%	56%	**35%**
dermat.	6%	13%	25%	25%	26%	43%	33%	40%	31%	39%	35%	42%	46%	40%	37%	64%	**34%**
otolaryng.	6%	14%	27%	26%	30%	40%	38%	40%	39%	33%	38%	45%	50%	40%	38%	67%	**36%**
radiol.	6%	13%	28%	24%	27%	43%	39%	45%	35%	43%	41%	48%	51%	40%	38%	73%	**37%**
urol.	5%	13%	24%	30%	27%	38%	33%	36%	33%	36%	37%	38%	44%	40%	38%	58%	**33%**
pulm. spec.	5%	12%	27%	24%	25%	37%	37%	33%	34%	38%	35%	40%	39%	39%	38%	63%	**33%**
hosp.	3%	8%	22%	20%	24%	33%	30%	28%	28%	29%	26%	35%	35%	26%	37%	37%	**26%**
lab	2%	4%	11%	11%	10%	25%	14%	16%	15%	14%	16%	23%	20%	24%	29%	39%	**17%**
phys. med.	7%	13%	22%	23%	24%	39%	32%	44%	33%	40%	33%	41%	49%	33%	33%	50%	**32%**
mean	**8%**	**14%**	**27%**	**27%**	**28%**	**40%**	**35%**	**38%**	**34%**	**38%**	**37%**	**42%**	**45%**	**39%**	**39%**	**58%**	**34%**

The bottom row entitled *mean* shows the range of average median UPCo values with the different linked ToCPs. We grouped the linked ToCPs according to these values and assigned them to 1 of the 3 categories: (8%-25%), (25%-42%), and (42%-58%), on the basis of the observed range.

Physical medicine (58%), pulmology (45%), and urology (42%) were located in the top category. GPs (8%) and pharmacies (14%) were the only ToCPs in the bottom category. All other ToCPs were assigned to the middle category.

Overall, Figure 4 shows a rather smooth increase of UPCo values from left to right. This seems to indicate that the concentration of patient-sharing relations is associated with the number of care providers per linked ToCP. The row entitled *lab* shows unusual values in this respect; they indicate that the laboratories' concentration of patient-sharing relations with other ToCPs is rather constantly low, regardless of the number of care providers per linked ToCP.

Discussion

Material and Methods

We relied on an established research database as our main data source. The database has been successfully used in various earlier research projects related to HIE [4,13,19]. This study's population of DM patients was shown to be plausible, on the basis of a comparison with the Austrian Health Survey of 2006/07 [4].

The UPCo and COCo (see Multimedia Appendix 2) metrics used to measure the concentration of patient-sharing relations

were simple to calculate and could thus be easily implemented in PostgreSQL. Selecting the linked care provider with whom the index care provider shared the highest number of her patients as the *usual provider* in the calculation of the UPCo seems reasonable to us when aiming for a measure for the concentration of patient-sharing relations.

Barnett and coworkers showed that the likelihood of a professional relationship between 2 care providers increases with the number of patients shared between them [9]. According to the authors, 9 or more shared patients indicate an actual professional relationship with a likelihood of more than 80%. We did not apply a minimum number of shared patients to consider a relationship between 2 care providers as our focus was on HIE with ToCPs. If, for example, a care provider shares each of her patients with a different pharmacy, pharmacies will still be an important HIE partner in our context, and it will be important to know that the COCo with pharmacies is low for this care provider.

We did not perform a stratified analysis for different patient attributes such as age and gender, as it would not provide useful conclusions for our research question. Each care provider typically treats a variety of patients with different characteristics and should be covered by a single HIE solution that is most suitable for all her patients. For HIE solutions that are optimized for different patient groups, it would not be realistic for a care provider to use multiple HIE solutions in parallel.

Need for Health Information Exchange Among Types of Care Providers

Ideally, each care provider involved in the treatment of DM patients would have access to an HIE platform that would allow her to exchange any health information with any care provider of any type. However, in current practical settings, this is usually not the case. As current platforms typically provide only partial HIE (ie, the content of selected ToCPs is shared or a subset of care providers participates in the HIE platform), it seems reasonable to categorize ToCPs according to their relevance for HIE. This knowledge could help to determine priorities for integrating ToCPs in an HIE platform.

According to Figure 3, we interpret pharmacies, GPs, and laboratories as HIE partners of high priority for any care provider involved in DM treatment insofar as the highest portions of patients are shared with them (compare the columns *GP*, *lab*, and *pharm.* in Figure 3). It should be noted that patient-sharing portions with laboratories are highly variable (compare interquartile ranges for the column entitled *lab* in Multimedia Appendix 3). In other words, laboratories might not be top priority HIE partners for all care providers.

Radiologists, ophthalmologists, hospitals, and internal medicine specialists might be classified as HIE partners of middle-rate priority in DM treatment. Variations are rather high among radiologists and internal medicine specialists (compare interquartile ranges for the corresponding columns in Multimedia Appendix 3), which means that they might be less important HIE partners for some care providers. Rather, a few DM patients are shared with all other ToCPs compared with the ToCPs rated as high or middle-rate priority (see above); thus, HIE will be needed less frequently here.

For an exemplary application of this study's results, we analyzed the recently introduced Austrian electronic health record (EHR) system ELGA [20] for its suitability in the treatment of DM. ELGA currently supports the exchange of (1) medication data prescribed by any ToCP and dispensed by pharmacies, (2) lab reports generated by hospital-based and outpatient laboratories, (3) radiology reports generated by hospitals and outpatient radiologists, and (4) hospital discharge letters. In other words, ELGA covers all ToCPs whom we rated as high-priority or medium-priority HIE partners, although GPs, ophthalmologists, and internal medicine specialists currently might only feed medication data into ELGA.

The fact that care providers typically share more of their patients with care providers of other types than with their own type (compare main diagonal of Figure 3) might primarily be explained by the fact that the payment systems of most Austrian social health insurances limit the access to care providers per ToCP (only 1 care provider per ToCP might be accessed during 1 accounting period).

Required Participation Rates of Types of Care Providers

In view of the postulated inverse relation of UPCo values with required participation, GPs and pharmacies would require high participation rates in the HIE platform. High participation rates might also be recommended for members of the middle UPCo category (int. med., gynec., ophthalm., dermat., orthop., radiol., neurol./psych., otolaryng., hosp., lab., and surg.). These desired participation rates could be reliably achieved by means of a national HIE platform with obligatory participation of the aforementioned ToCPs. In Austria, ELGA ensures this condition through obligatory participation of all public care providers.

The apparent association of the concentration of patient-sharing relations with the number of care providers per linked ToCP seems intuitive. For instance, if 1 of our index care providers shared DM patients with a linked ToCP of *physical medicine*, the patients could choose among only 56 physical medicine specialists. If we assume that care providers of a ToCP are distributed rather evenly in geographic terms (public care providers should reasonably be distributed in a way that allows them to be accessed homogeneously by the entire population), the 56 physical medicine specialists will usually be located farther apart from each other than the 4892 GPs. It would thus be obvious that many of the index care providers' patients concentrated on the physical medicine specialist who is located closest to the index care provider. This would explain the high UPCo values in the column entitled *phys. med.* in Figure 4. In contrast, DM patients shared with GPs would naturally be divided among several different GPs located in close vicinity to the index care provider. This would explain the low UPCo values in the column entitled *GP* in Figure 4.

Another explanation for the characteristics of Figure 4 could be the *patient sending/receiving role* of the ToCPs. For instance, GPs and internal medicine specialists have a *patient sending role*. They act as *gatekeepers* and are the first to be visited by

a DM patient in a treatment chain. When the patients of a GP have to visit other care providers in due course, similar patterns in selecting these care providers (and thus high UPCo values) could result from recommendations of the GP. In contrast, a care provider with a *patient receiving role* (such as laboratories and pharmacies) has less influence on whose patients are sent to her, resulting in low UPCo values. This explanation would be in accordance with the high UPCo values in the rows entitled *GP* and *int. med.* in Figure 4, and it would be in accordance with the low UPCo values in the row entitled *lab*. However, it would be in conflict with the high UPCo values in the row entitled *pharm.*.

Related Work

In the context of diabetes-specific HIE, several authors concentrated on the patients' role in information sharing [21,22]. HIE between DM patients and care providers was examined with a focus on sharing medication data [23], email communication [24], and patient preferences [25].

Koopman and coworkers name a set of data elements that are relevant for outpatient family physicians and general internal medicine physicians in the treatment of DM patients [26]. However, they neither address how these data elements were identified nor address which ToCPs should deliver the corresponding values.

Huebner-Bloder and coworkers identified 446 relevant data elements in the treatment of DM and grouped these in 9 categories [27]. They used a triangulation design that was mainly based on documentation analysis in 3 DM outpatient clinics and interviews with 6 internists specialized in DM. The identified data elements originate from GPs, internal medicine physicians, ophthalmologists, nephrologists, neurologists, gynecologists, psychiatrists, dermatologists, hospitals, laboratories, and from the patient's self-monitoring. The ToCPs identified by them as being relevant in the treatment of DM thus constitute a subset of our ToCPs, except for nephrologists (who are a part of the ToCP *internal medicine* in our claims data) and patient-reported data (not considered in our claims data).

According to Joshy and Simmons, HIE between systems of GPs and hospitals are crucial factors for the success of DM information systems [2]. They further state that "pharmacy data, lab measurements, retinal screening, and home blood glucose monitoring data are increasingly being linked into diabetes information systems." This fits with this study's results insofar as we identified GPs, pharmacies, and laboratories as high-priority HIE partners in the treatment of DM, as well as hospitals and ophthalmologists as middle-priority HIE partners. Patient-reported data were not considered in this study.

Existing HIE platforms only partly cover the information needs of care providers. According to a recent study, only 58% of the analyzed DM information systems provided HIE with hospitals, 22% provided HIE with primary care, and only 3% provided HIE with hospitals and primary care [28]. In their review of regional HIE platforms, Mäenpää et al conclude that the latter provide inadequate access to patient-relevant clinical data [29]. Nationwide EHR systems, which are operated as national HIE platforms in 59% of the European World Health Organization member states [30], are typically restricted to the exchange of patient summaries or selected document types [8].

Limitations

One of our basic assumptions was that those ToCPs with whom care providers share most of their DM patients should be considered high-priority HIE partners. Even though this assumption seems intuitive, in the individual case, there might also be care providers from ToCPs with low patient-sharing portions, who possess patient information of high importance. Furthermore, the need for HIE might differ among certain combinations of ToCPs and thus not be naturally reflected by patient-sharing rates. As an example, laboratories typically share 98% of their DM patients with GPs, whereas GPs *only* share 77% of their patients with laboratories (compare Figure 3). Nevertheless, it might be more important for the GP to receive a lab result (and then add it to the patient's local EHR) than for the laboratory to electronically receive the request for a particular test (this could probably also be solved conventionally without serious detriment).

Hospital visits could only be considered from the inpatient domain in our analysis; data from hospital outpatient departments were not included in our data source. As hospital outpatient departments are frequently visited by Austrian DM patients, hospitals are probably underrepresented as ToCPs in this study. Furthermore, visits to private care providers were not taken into account as the corresponding data were incomplete in our data source. This might have led to an underrepresentation of those ToCPs with large numbers of private care providers in Austria, such as physical medicine, surgery, and neurology/psychiatry.

We only focused on DM patients; therefore, our insights concerning HIE can only be applied to the treatment of DM. However, as a next step, we intend to repeat the analysis for other chronic diseases to see whether there are general *patterns* related to the design and selection of HIE platforms in the treatment of chronically ill patients.

Conclusions

The results of this study provide insights for 2 different types of actors in the HIE of DM patients.

First, implementers and providers of HIE platforms who strive to support DM treatment should ensure that they integrate GPs, pharmacies, and laboratories from the start, as these constitute HIE partners of high priority for all ToCPs. Radiologists, ophthalmologists, hospitals, and internal medicine specialists should be integrated into the HIE platforms in the second step. The remaining ToCPs seem to be HIE partners of lower priority in the treatment of DM and could thus be integrated in the final step, if resources permit. Furthermore, this study's results seem to suggest that DM patients shared with ToCPs who have many members (in our case, particularly GPs and pharmacies) are divided among many different care providers of these ToCPs. We conclude that, for implementing HIE with ToCPs who have many members, it would be essential to have an HIE solution with high participation rates of these ToCPs (ideally a nationwide HIE platform with obligatory participation of the

concerned ToCPs). This would increase the chances of HIE with any cooperating care provider of these ToCPs.

In Austria, ELGA satisfies these demands and thus serves as a suitable HIE platform for DM treatment. It could become even more useful if, besides medication data, further information registered by GPs, ophthalmologists, and internal medicine specialists was provided.

Second, care providers using a DM-specific HIE platform might gain insights from this study's results for *their* index ToCP (ie, *their* row of Figure 3 and Figure 4). This study's results provide information about a care provider's general DM-related HIE characteristics with respect to all linked ToCPs. For instance, radiologists seem to be important HIE partners for orthopedists as they typically share 85% of their DM patients with radiologists, according to the row entitled *orthop.* in Figure 3. However, for hospitals, radiologists appear to be HIE partners of rather low priority (only 44% shared DM patients). Furthermore, the rather constantly low UPCo values of laboratories indicate that a laboratory's shared DM patients are usually divided among several care providers for each linked ToCP. This suggests that, in the context of DM treatment in Austria, laboratories benefit from being connected to an HIE solution with high participation rates of other ToCPs. Being able to transmit test results to practically any requesting care provider via the HIE solution will widen the laboratory's catchment area of cooperating care providers and thus be commercially useful.

Acknowledgments

The authors would like to thank Florian Endel for his support in accessing the database that we used for the study. This report builds on earlier work of the authors that was supported financially by the Main Association of the Austrian Social Security Institutions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PostgreSQL script for the calculation of the metrics usual provider cooperation and concentration of cooperation.

[TXT File, 2KB - medinform_v7i2e12172_app1.txt]

Multimedia Appendix 2

The concentration of cooperation metric.

[PDF File (Adobe PDF File), 351KB - medinform_v7i2e12172_app2.pdf]

Multimedia Appendix 3

Interquartile ranges for portions of shared patients and usual provider cooperation values.

[PDF File (Adobe PDF File), 321KB - medinform_v7i2e12172_app3.pdf]

References

1. Reis ZS, Maia TA, Marcolino MS, Becerra-Posada F, Novillo-Ortiz D, Ribeiro AL. Is there evidence of cost benefits of electronic medical records, standards, or interoperability in hospital information systems? Overview of systematic reviews. JMIR Med Inform 2017 Aug 29;5(3):e26 [FREE Full text] [doi: 10.2196/medinform.7400] [Medline: 28851681]
2. Joshy G, Simmons D. Diabetes information systems: a rapidly emerging support for diabetes surveillance and care. Diabetes Technol Ther 2006 Oct;8(5):587-597. [doi: 10.1089/dia.2006.8.587] [Medline: 17037973]
3. Raaijmakers LG, Hamers FJ, Martens MK, Bagchus C, de Vries NK, Kremers SP. Perceived facilitators and barriers in diabetes care: a qualitative study among health care professionals in the Netherlands. BMC Fam Pract 2013 Aug 10;14:114 [FREE Full text] [doi: 10.1186/1471-2296-14-114] [Medline: 23937325]
4. Rinner C, Sauter SK, Endel G, Heinze G, Thurner S, Klimek P, et al. Improving the informational continuity of care in diabetes mellitus treatment with a nationwide shared EHR system: estimates from Austrian claims data. Int J Med Inform 2016 Aug;92:44-53 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.05.001] [Medline: 27318070]
5. Duftschmid G, Dorda W, Endel G, Fröschl K, Gall W, Grossmann W, et al. Fragmentation of diabetes treatment in Austria - an indicator for the need for shared electronic health record systems. Stud Health Technol Inform 2012;180:667-671. [Medline: 22874275]
6. Weiß J, Welzel T, Hartmann BJ, Hübner U, Teuteberg F. Towards designing a secure exchange platform for diabetes monitoring and therapy. Stud Health Technol Inform 2018;248:239-246. [Medline: 29726443]

7. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc 2010;17(3):288-294 [FREE Full text] [doi: 10.1136/jamia.2010.003673] [Medline: 20442146]
8. European Commission. eHealth Strategies. European Countries on their Journey towards National eHealth Infrastructures URL: http://www.ehealth-strategies.eu/ [accessed 2018-07-16] [WebCite Cache ID 70xBjh8fQ]
9. Barnett ML, Landon BE, O'Malley AJ, Keating NL, Christakis NA. Mapping physician networks with self-reported and administrative data. Health Serv Res 2011 Oct;46(5):1592-1609 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01262.x] [Medline: 21521213]
10. Landon BE, Keating NL, Onnela J, Zaslavsky AM, Christakis NA, O'Malley AJ. Patient-sharing networks of physicians and health care utilization and spending among Medicare beneficiaries. JAMA Intern Med 2018 Jan 1;178(1):66-73. [doi: 10.1001/jamainternmed.2017.5034] [Medline: 29181504]
11. Ong M, Olson KL, Chadwick L, Liu C, Mandl KD. The impact of provider networks on the co-prescriptions of interacting drugs: a claims-based analysis. Drug Saf 2017 Dec;40(3):263-272 [FREE Full text] [doi: 10.1007/s40264-016-0490-1] [Medline: 28000151]
12. Lublóy A, Keresztúri JL, Benedek G. Formal professional relationships between general practitioners and specialists in shared care: possible associations with patient health and pharmacy costs. Appl Health Econ Health Policy 2016 Apr;14(2):217-227. [doi: 10.1007/s40258-015-0206-1] [Medline: 26476734]
13. Sauter S, Neuhofer L, Endel G, Klimek P, Duftschmid G. Analyzing healthcare provider centric networks through secondary use of health claims data. 2014 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); June 1-4, 2014; Valencia, Spain p. 522-525. [doi: 10.1109/BHI.2014.6864417]
14. Chini F, Pezzotti P, Orzella L, Borgia P, Guasticchi G. Can we use the pharmacy data to estimate the prevalence of chronic conditions? A comparison of multiple data sources. BMC Public Health 2011 Sep 5;11:688 [FREE Full text] [doi: 10.1186/1471-2458-11-688] [Medline: 21892946]
15. Landon BE, Keating NL, Barnett ML, Onnela J, Paul S, O'Malley AJ, et al. Variation in patient-sharing networks of physicians across the United States. J Am Med Assoc 2012 Jul 18;308(3):265-273 [FREE Full text] [doi: 10.1001/jama.2012.7615] [Medline: 22797644]
16. Bice TW, Boxerman SB. A quantitative measure of continuity of care. Med Care 1977 Apr;15(4):347-349. [Medline: 859364]
17. Breslau N, Reeb KG. Continuity of care in a university-based practice. J Med Educ 1975 Oct;50(10):965-969. [Medline: 1159765]
18. Jee SH, Cabana MD. Indices for continuity of care: a systematic review of the literature. Med Care Res Rev 2006 Apr;63(2):158-188. [doi: 10.1177/1077558705285294] [Medline: 16595410]
19. Rinner C, Grossmann W, Sauter SK, Wolzt M, Gall W. Effects of shared electronic health record systems on drug-drug interaction and duplication warning detection. Biomed Res Int 2015;2015:380497 [FREE Full text] [doi: 10.1155/2015/380497] [Medline: 26682218]
20. Herbek S, Eisl HA, Hurch M, Schator A, Sabutsch S, Rauchegger G, et al. The Electronic Health Record in Austria: a strong network between health care and patients. Eur Surg 2012 Jun;44(3):155-163. [doi: 10.1007/s10353-012-0092-9]
21. Bond CS, Ahmed OH. Can I help you? Information sharing in online discussion forums by people living with a long-term condition. J Innov Health Inform 2016 Dec 10;23(3):620-626. [doi: 10.14236/jhi.v23i3.853] [Medline: 28059696]
22. Weymann N, Härter M, Dirmaier J. Information and decision support needs in patients with type 2 diabetes. Health Informatics J 2016 Mar;22(1):46-59. [doi: 10.1177/1460458214534090] [Medline: 24916569]
23. Bernhard G, Mahler C, Seidling HM, Stützle M, Ose D, Baudendistel I, et al. Developing a shared patient-centered, web-based medication platform for type 2 diabetes patients and their health care providers: qualitative study on user requirements. J Med Internet Res 2018 Mar 27;20(3):e105 [FREE Full text] [doi: 10.2196/jmir.8666] [Medline: 29588269]
24. Devkota B, Salas J, Sayavong S, Scherrer JF. Use of an online patient portal and glucose control in primary care patients with diabetes. Popul Health Manag 2016 Apr;19(2):125-131. [doi: 10.1089/pop.2015.0034] [Medline: 26237200]
25. Asan O, Cooper Ii F, Nagavally S, Walker RJ, Williams JS, Ozieh MN, et al. Preferences for health information technologies among us adults: analysis of the Health Information National Trends Survey. J Med Internet Res 2018 Oct 18;20(10):e277 [FREE Full text] [doi: 10.2196/jmir.9436] [Medline: 30341048]
26. Koopman RJ, Kochendorfer KM, Moore JL, Mehr DR, Wakefield DS, Yadamsuren B, et al. A diabetes dashboard and physician efficiency and accuracy in accessing data needed for high-quality diabetes care. Ann Fam Med 2011;9(5):398-405 [FREE Full text] [doi: 10.1370/afm.1286] [Medline: 21911758]
27. Hübner-Bloder G, Duftschmid G, Kohler M, Rinner C, Saboor S, Ammenwerth E. Clinical situations and information needs of physicians during treatment of diabetes mellitus patients: a triangulation study. Stud Health Technol Inform 2011;169:369-373. [Medline: 21893775]
28. Parker VE, Simmons D. Diabetes information technology in UK diabetes secondary care: a snapshot. Br J Diabetes Vasc Dis 2013 Jul 19;13(3):150-154. [doi: 10.1177/1474651413493019]
29. Mäenpää T, Suominen T, Asikainen P, Maass M, Rostila I. The outcomes of regional healthcare information systems in health care: a review of the research literature. Int J Med Inform 2009 Nov;78(11):757-771. [doi: 10.1016/j.ijmedinf.2009.07.001] [Medline: 19656719]

30. World Health Organization: Regional Office for Europe. 2016. From innovation to implementation – eHealth in the WHO European Region (2016) URL: http://www.euro.who.int/en/health-topics/Health-systems/e-health/publications/2016/from-innovation-to-implementation-ehealth-in-the-who-european-region-2016 [accessed 2018-07-16] [WebCite Cache ID 70xDwsrkE]

Abbreviations

COC: continuity of care index
COCo: concentration of cooperation
DM: diabetes mellitus
EHR: electronic health record
ELGA: Elektronische Gesundheitsakte (electronic health record)
GP: general practitioner
HIE: health information exchange
PSN: patient-sharing network
ToCP: type of care provider
UPC: usual provider continuity
UPCo: usual provider cooperation

Edited by G Eysenbach; submitted 11.09.18; peer-reviewed by M Sharma, S Barteit, E Frontoni; comments to author 13.10.18; revised version received 08.12.18; accepted 20.01.19; published 12.04.19.

Please cite as:
Duftschmid G, Rinner C, Sauter SK, Endel G, Klimek P, Mitsch C, Heinzl H
Patient-Sharing Relations in the Treatment of Diabetes and Their Implications for Health Information Exchange: Claims-Based Analysis
JMIR Med Inform 2019;7(2):e12172
URL: http://medinform.jmir.org/2019/2/e12172/
doi:10.2196/12172
PMID:30977733

Review

Interoperable Electronic Health Records and Health Information Exchanges: Systematic Review

Mark J Dobrow[1], MSc, PhD; Jessica P Bytautas[1], MSc; Sukirtha Tharmalingam[2], MHSc; Simon Hagens[2], MBA

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Canada Health Infoway - Inforoute Santé du Canada, Toronto, ON, Canada

Corresponding Author:
Mark J Dobrow, MSc, PhD
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
4th Floor, 155 College St
Toronto, ON,
Canada
Phone: 1 4169780558
Email: mark.dobrow@utoronto.ca

Abstract

Background: As the availability of interoperable electronic health records (iEHRs) or health information exchanges (HIEs) continues to increase, there is greater need and opportunity to assess the current evidence base on what works and what does not regarding the adoption, use, and impact of iEHRs.

Objective: The purpose of this project is to assess the international evidence base on the adoption, use, and impact of iEHRs.

Methods: We conducted a systematic review, searching multiple databases—MEDLINE, Embase, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL)—with supplemental searches conducted in Google Scholar and grey literature sources (ie, Google, Grey Literature Report, and OpenGrey). All searches were conducted in January and February 2017. Articles were eligible for inclusion if they were published in English, were published from 2006 to 2017, and were either an original research study or a literature review. In order to be included, articles needed to focus on iEHRs and HIEs across multiple health care settings, as well as on the impact and effectiveness of iEHR adoption and use.

Results: We included 130 articles in the synthesis (113 primary studies, 86.9%; 17 reviews, 13.1%), with the majority focused on the United States (88/130, 67.7%). The primary studies focused on a wide range of health care settings; the three most prevalent settings studied included acute care (59/113, 52.2%), primary care (44/113, 38.9%), and emergency departments (34/113, 30.1%). We identified 29 distinct measurement items in the 113 primary studies that were linked to 522 specific measurement outcomes. Productivity and quality were the two evaluation dimensions that received the most attention, accounting for 14 of 29 (48%) measurement items and 306 of 522 (58.6%) measurement outcomes identified. Overall, the majority of the 522 measurement outcomes were positive (298/522, 57.1%). We also identified 17 reviews on iEHR use and impact, 6 (35%) that focused on barriers and facilitators to adoption and implementation and 11 (65%) that focused on benefits and impacts, with the more recent reviews finding little generalizable evidence of benefit and impact.

Conclusions: This review captures the status of an evolving and active field focused on the use and impact of iEHRs. While the overall findings suggest many positive impacts, the quality of the primary studies were not evaluated systematically. When broken down by specific measurement item, the results directed attention both to measurement outcomes that were consistently positive and others that were mostly negative or equivocal.

(JMIR Med Inform 2019;7(2):e12607) doi:10.2196/12607

KEYWORDS

health information exchange; electronic health record; interoperability; use; impact; systematic review

Introduction

Interoperable electronic health records (iEHRs) in Canada increasingly provide individual patients with a secure and private record of their health history and care within their health system [1]. The iEHR draws on core systems that collect information electronically, including client and provider demographic registries, diagnostic imaging systems, drug information systems, laboratory information systems, public health systems, and clinical reporting systems [2]. This record is designed to facilitate the sharing of data across the continuum of care, across health care delivery organizations, and across geographical areas. In Canada, 42% of nurses and 42% of primary care physicians report having access to provincial and territorial patient information systems [3,4]. However, the method to access information, the availability of information in care settings, and the user information needed to access information differs across provinces and territories.

While different in architecture, iEHR solutions are analogous to health information exchange (HIE) initiatives in the United States. Health professionals in Canada, as in most high-income countries, also receive patient information across settings through other digital health solutions, such as hospital information systems and laboratory systems. Other countries have developed systems similar to Canada's iEHRs. The common element of interest for this project is the provision of information across care settings and health professionals to improve care for patients.

As the availability of iEHRs continues to increase [2], there is greater need and opportunity to assess and understand the current evidence base on what works and what does not work regarding the adoption, use, and impact of iEHRs. This evidence is important to guide progress and improve iEHR capabilities but also to identify gaps in the evidence base where more targeted investment and evaluation is needed. Therefore, the purpose of this project is to conduct a systematic review of the international evidence base on the adoption, use, and impact of iEHRs or HIEs. The findings will also contribute to a national study to value the contribution of iEHRs and other connected information, which is part of a series of studies to value key digital health benefits [5].

Methods

We conducted a systematic review, documenting the key elements of the review, including search strategy, eligibility criteria, article selection process, analysis, and synthesis, using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.

Search Strategy

We developed and tested multiple search strategies that incorporated subject heading and keyword terms for "interoperable electronic health record" (including "iEHR"), "health information exchange" or "HIE"; "interoperability"; "adoption" or "use"; and "effectiveness," "impact," or "value." We consulted and sought feedback on the search strategies with health informatics experts, including from academic and government agencies each focused on health informatics, and a specialist librarian who performed an abbreviated Peer Review of Electronic Search Strategies (PRESS) assessment of a preliminary MEDLINE search strategy [6]. The health informatics experts identified five articles that reflected the targeted aims of the review, which we used to test the effectiveness of candidate search strategies to identify relevant articles. Ultimately, a final search strategy was selected and translated for use in several traditional databases—MEDLINE, Embase, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL)—with supplemental searches conducted in Google Scholar and grey literature sources (ie, Google, Grey Literature Report, and OpenGrey); see Multimedia Appendix 1 for database-specific search strategies. All searches were conducted in January and February 2017. Additionally, reference lists of all included articles were reviewed to identify additional articles that the search strategy missed.

Eligibility Criteria

Articles were eligible for inclusion in the review if they were published in English; published between 2006 and 2017, although, only a small number of articles published in 2017 were available at the time of our review; and were either an original research study, inclusive of both quantitative and qualitative study types, or a literature review. For inclusion in the review, articles needed to focus on iEHRs and HIEs, excluding electronic medical records, across two or more health care settings; they also needed to focus on the impact or effectiveness of iEHR or HIE adoption or use.

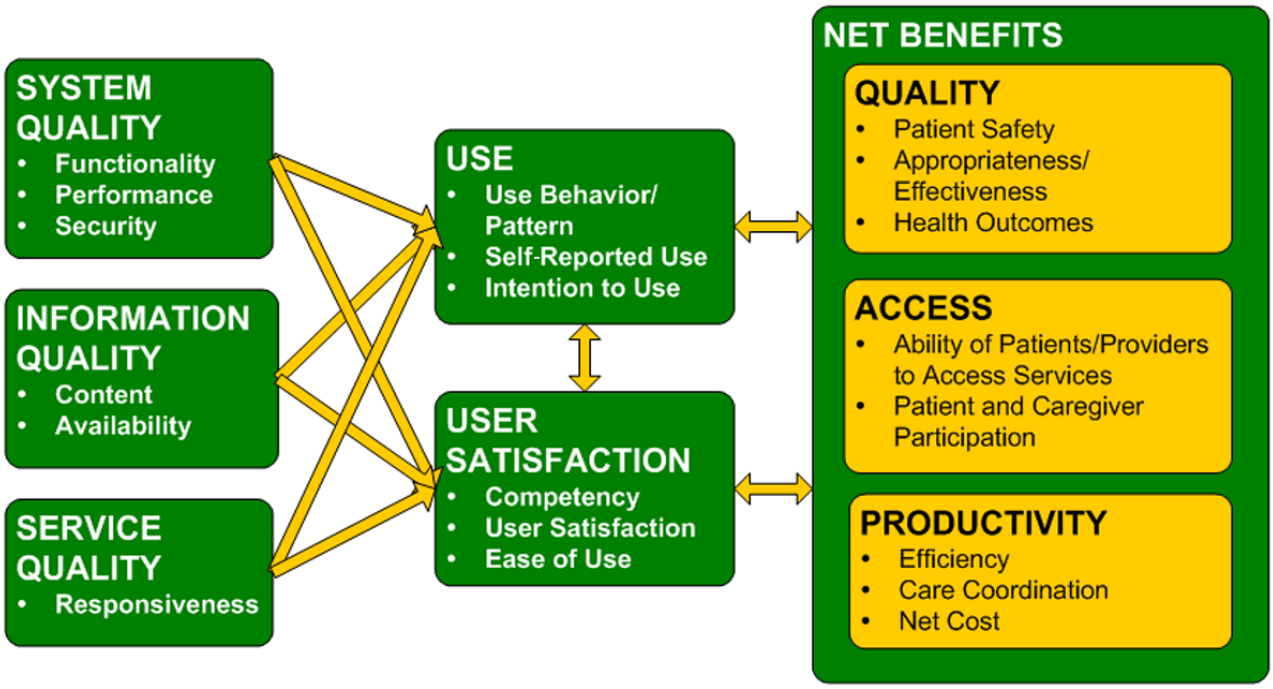
Study Selection

Two reviewers (JPB and one of three research assistants) independently screened all titles and abstracts or, in the case of Google and Google Scholar search results, titles and excerpts. Discrepancies were resolved through discussion and/or review by a third reviewer (MJD). The same process was used for review of full-text articles.

Analysis and Synthesis

Data from all articles identified for inclusion were extracted independently by two reviewers (JPB and one of three research assistants) using a predeveloped data extraction template that included the year of publication, article type, approach, methodology, jurisdiction, and health care setting. Given the anticipated heterogeneity of study types and the intention to capture the breadth of iEHR evaluation activity, we prioritized evaluation relevance over quality and, therefore, did not use available quality criteria to exclude primary studies [7-9].

Figure 1. Canada Health Infoway benefits evaluation framework.



For all primary studies, two reviewers (JPB and one of three research assistants) independently extracted distinct measurement items and measurement outcomes verbatim. Measurement items were then coded and recoded inductively over three iterations into thematic categories by one reviewer (JPB) in discussion with the review team. All measurement outcomes were categorized as positive, negative, or mixed or neutral by one reviewer (JPB) and reviewed by the review team. All thematic measurement item categories were classified as one of the eight evaluation dimensions of benefit—system quality, service quality, information quality, user satisfaction, use, productivity, quality, and access—based on the Infoway benefits evaluation framework (see Figure 1). This framework, based on the Delone and MacLean Information Systems Success Model [10], details the measurement item and outcome categories and has been used extensively across Canada and internationally since it was first published in 2007 [11]. This classification approach has also been used in a relevant recent review of project evaluations from electronic health record (EHR) implementations across Canada [12].

We employed a separate analysis approach for the included literature review articles, conducting a descriptive analysis of each review article that assessed the (1) main focus of the review, (2) main findings, and (3) recommendations for future research.

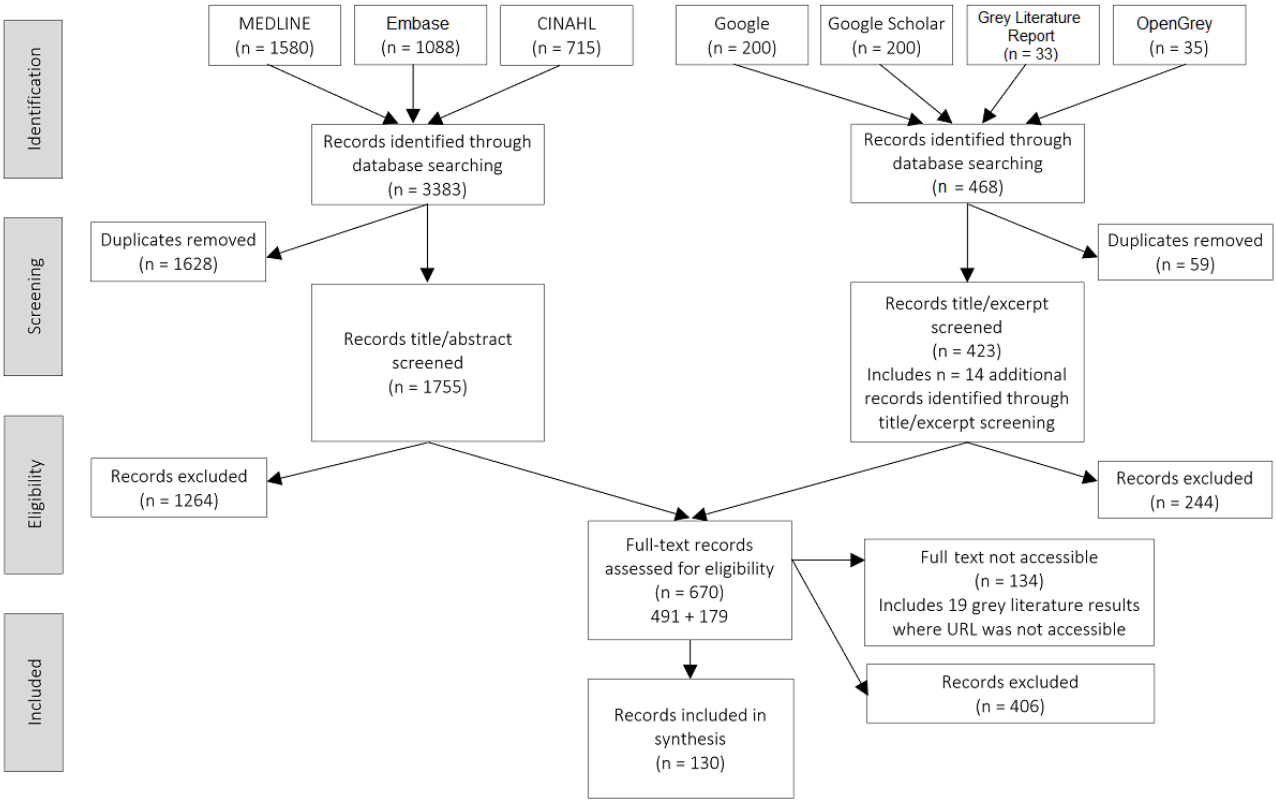
Results

Overview

Our search of seven data sources identified 3851 records. After deduplication; title, abstract, and excerpt screening; and full-text review, 130 articles were included in the synthesis; Figure 2 presents the PRISMA flow diagram for the review. These 130 articles included 113 journal articles (86.9%), 11 reports (8.5%), and 6 documents of other types (4.6%). Of the 130 articles, 113 were primary studies (86.9%) and 17 were various types of reviews (13.1%). The vast majority of the articles focused on the United States (88/130, 67.7%); 6 (4.6%) focused on Israel; 3 (2.3%) each focused on Canada, Finland, and the United Kingdom; 2 (1.5%) focused on South Korea; 7 (5.4%) focused on another single country, including Australia, Austria, Brazil, Greece, Kenya, the Netherlands, and Switzerland; and another 18 (13.8%) had a multi-country focus. Of the 130 articles, 95 (73.1%) were published in the 6-year period from 2012 to 2017.

For the 113 primary studies, the majority employed quantitative methodologies exclusively (71/113, 62.8%) or in combination with qualitative methods (ie, mixed-method approaches) (20/113, 17.7%). The primary studies focused on a wide range of health care settings; the three most prevalent settings studied included acute care (59/113, 52.2%), primary care (44/113, 38.9%), and emergency departments (34/113, 30.1%). Other settings included laboratories (18/113, 15.9%), ambulatory care (17/113, 15.0%), pharmacies (13/113, 11.5%), public health departments (14/113, 12.4%), long-term care (9/113, 8.0%), payer or purchaser organizations (9/113, 8.0%), and home and community care (3/113, 2.7%).

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. CINAHL: Cumulative Index to Nursing and Allied Health Literature.



Summary of Findings From Primary Studies

The primary studies focused on six general dimensions. We identified 29 distinct measurement items, representing evaluation themes, in the 113 primary studies that were linked to 522 specific measurement outcomes (see Table 1). Productivity and quality were the two evaluation dimensions that received the most attention in the articles we reviewed, accounting for 14 of 29 (48%) measurement items identified and 306 of 522 (58.6%) measurement outcomes identified. Of the six evaluation dimensions assessed, service quality and system quality received the least attention, accounting for only 5 of 29 (17%) measurement items and 79 of 522 (15.1%) measurement outcomes documented from the articles reviewed. Measurement items were not assigned to either the *use* or *access* evaluation dimensions.

Overall, the majority of the 522 measurement outcomes were positive (298/522, 57.1%), with the remaining measurement outcomes reported as neutral or mixed results (107/522, 20.5%) or negative findings (117/522, 22.4%). When examining each of the 29 measurement items separately, the majority (22/29, 76%) had more positive than negative measurement outcomes, with the most frequently studied measurement items having a larger proportion of positive over negative outcomes. The 5 measurement items (5/29, 17%) with more negative than positive measurement outcomes were (1) stakeholder engagement, (2) performance and reliability, (3) security and privacy, (4) overall quality of information, and (5) ease of use; 2 measurement items (2/29, 7%) had equal positive and negative measurement outcomes: (1) layout and format and (2) community-based care, public or population health, or preventive services.

To provide more details on the findings, we consider each of the six dimensions separately. A total of 2 measurement items (2/29, 7%) were aligned with the service quality dimension. These included stakeholder engagement, which has mostly negative results, and training and support, which has mostly positive results. For the system quality dimension, 3 of 29 (10%) measurement items applied, including performance and reliability, security and privacy, and assessment and planning. Of these 3, only the latter had positive measurement outcomes. The information quality dimension had 6 of 29 (21%) measurement items, with the 2 most frequently measured items—data accuracy and completeness, as well as information availability—mostly positive, while the 4 less frequently measured items each revealed equivocal results. A total of 4 of 29 measurement items (14%) aligned with the user satisfaction dimension, with 1 showing mostly positive measurement outcomes—perceived usefulness or value and trust or confidence in the system—while the remaining 3 measurement items showed measurement outcomes that were either equivocal or negative. The productivity and quality dimensions each had 7 of 29 (24%) measurement items. All 7 measurement items for productivity had positive measurement outcomes, while 6 of the 7 measurement items for quality also had positive measurement outcomes. As noted above, the productivity and quality dimensions have received the majority of focus for measurement and have yielded mostly positive outcomes.

Table 1. Classification of iEHR[a] and HIE[b] measurement outcomes from primary studies. Measurement items are ordered by dimension and then by total number of measurement outcomes. Identified measurement items are only reported once in the table.

Evaluation dimension, measurement item	N	Positive, n (%)	Mixed or neutral, n (%)	Negative, n (%)	Total, n (%)
Productivity					
Financial costs	58	28 (48)	16 (28)	14 (24)	58 (100)
Efficiency in ordering and accessing tests, exams, results, or other clinical info	53	37 (70)	14 (26)	2 (4)	53 (100)
Time savings in general	19	16 (84)	0 (0)	3 (16)	19 (100)
Reduced hospital admissions and readmissions; shorter length of stay	19	10 (53)	5 (26)	4 (21)	19 (100)
General productivity	13	10 (77)	2 (15)	1 (8)	13 (100)
Efficiency due to improved organizational and managerial effectiveness	6	6 (100)	0 (0)	0 (0)	6 (100)
Return on investment	2	2 (100)	0 (0)	0 (0)	2 (100)
Subtotal	170	109 (64)	37 (22)	24 (14)	170 (100)
Quality					
Enhanced ability to communicate, collaborate, and coordinate care	41	29 (71)	8 (20)	4 (10)	41 (100)
Overall quality of care	26	18 (69)	6 (23)	2 (8)	26 (100)
Clinical decision support	24	18 (75)	3 (13)	3 (13)	24 (100)
Prescribing behavior, medication monitoring, or support	19	9 (47)	8 (42)	2 (11)	19 (100)
Patient health outcomes	9	6 (67)	3 (33)	0 (0)	9 (100)
Patient safety	9	5 (56)	4 (44)	0 (0)	9 (100)
Community-based care, public or population health, or preventive services	8	3 (38)	2 (25)	3 (38)	8 (100)
Subtotal	136	88 (65)	34 (25)	14 (10)	136 (100)
Information quality					
Accuracy and completeness of data	22	12 (55)	5 (23)	5 (23)	22 (100)
Provided quickly or is available when needed	19	12 (63)	1 (5)	6 (32)	19 (100)
Enables access to information previously unavailable or accessed through another process	9	4 (44)	2 (22)	3 (33)	9 (100)
Overall quality of information	8	3 (38)	1 (13)	4 (50)	8 (100)
Standards, coding, or documentation for data storage and retrieval	8	4 (50)	1 (13)	3 (38)	8 (100)
Layout and format	6	3 (50)	0 (0)	3 (50)	6 (100)
Subtotal	72	38 (53)	10 (14)	24 (33)	72 (100)
User satisfaction					
Perceived usefulness or value and trust or confidence in system	22	17 (77)	2 (9)	3 (14)	22 (100)
Integrated into workflow	19	8 (42)	5 (26)	6 (32)	19 (100)
Ease of use	13	3 (23)	4 (31)	6 (46)	13 (100)
Overall satisfaction	11	4 (36)	5 (45)	2 (18)	11 (100)
Subtotal	65	32 (49)	16 (25)	17 (26)	65 (100)
System quality					
Performance and reliability	22	6 (27)	3 (14)	13 (59)	22 (100)
Security and privacy	16	6 (38)	1 (6)	9 (56)	16 (100)
Assessment and planning	6	5 (83)	0 (0)	1 (17)	6 (100)

Evaluation dimension, measurement item	N	Positive, n (%)	Mixed or neutral, n (%)	Negative, n (%)	Total, n (%)
Subtotal	44	17 (39)	4 (9)	23 (52)	44 (100)
Service quality					
Training and support	18	9 (50)	4 (22)	5 (29)	18 (100)
Stakeholder engagement	17	5 (29)	2 (12)	10 (59)	17 (100)
Subtotal	35	14 (40)	6 (17)	15 (43)	35 (100)
Total	522	298 (57)	107 (20)	117 (22)	522 (100)

[a]iEHR: interoperable electronic health record.

[b]HIE: health information exchange.

When looking at the results from a setting-specific perspective, where sufficient volumes existed, there were some notable differences from the overall results. Acute care settings were assessed by 59 out of 113 studies (52.2%) and represented 270 out of 522 (51.7%) distinct measurement outcomes. Of these, there was focus on each of the six dimensions, with considerable attention on service quality, system quality, and some aspects of productivity. Primary care settings were assessed by 44 out of 113 (38.9%) studies and represented 183 out of 522 (35.1%) distinct measurement outcomes that covered most of the six dimensions, with attention directed predominantly to productivity measures while service, system, and information quality received less focus. Emergency department settings were assessed by 34 out of 133 (30.1%) studies and represented 112 out of 522 (21.5%) distinct measurement outcomes. There was a lack of measurement outcomes for most of the six dimensions, with the exception of one type of productivity item and one type of quality item.

Summary of Findings From Reviews

We identified 17 reviews on iEHR use or impact (see Table 2). Of these reviews, 6 (35%) focused on barriers and facilitators to adoption or implementation, and 11 (65%) focused on benefits or impacts. A total of 10 of the 17 reviews (59%) were published between 2013 and 2016 and, with the exception of 3 reviews (18%) that were limited in scope to clinical research [13], chronic disease [14], or ambulatory primary care [15], most reviews (14/17, 82%) examined general benefits or impacts of iEHRs or HIEs.

The more recent reviews (ie, published since 2015) found little generalizable evidence of benefit or impact. The reviews highlight some less methodologically robust research that focused on resource use and perception of outcomes; these reviews found that iEHRs or HIEs increase productivity (eg, reduction in duplicate testing, emergency department costs, or hospital admissions [16,17]) and are valued by patient and physician stakeholders [17], all of which is consistent with the findings from our review. However, authors of one of the recent reviews [18] cautioned on overinterpreting the generalizability of this work given the methodological limitations of the primary studies and the developing state of iEHR or HIE evaluative work overall. The recent reviews on the benefit or impact of iEHRs or HIE identify three main areas for future research, including the following: (1) focus on how the setting in which iEHRs or HIEs are used affects specific health care outcomes [18]; (2) use of more rigorous, coordinated, and systematic approaches to evaluate the relationship between iEHRs or HIEs and health care outcomes [16]; and (3) need for better understanding of the organizational factors that affect iEHR or HIE contributions to improved clinical care [17].

Table 2. Summary of review articles included in this review.

Authors	Title	Source	Year	Primary focus
Anonymous [14]	Electronic tools for health information exchange: An evidence-based analysis	Ontario Health Technology Assessment Series. 13 (11).	2013	Benefits and impacts
Akhlaq et al [19]	Barriers and facilitators to health information exchange in low- and middle-income country settings: A systematic review	Health Policy & Planning. 31(9):1310-1325.	2016	Barriers and facilitators
Dobrev et al [20]	Report on methodology for evaluating the socio-economic impact of interoperable EHR and ePrescribing systems	EHR IMPACT. Prepared for the European Commission, Directorate General Information Society and Media, Brussels.	2008	Benefits and impacts
Eden et al [21]	Barriers and facilitators to exchanging health information: A systematic review	International Journal of Medical Informatics. 88:44-51.	2016	Barriers and facilitators
Edwards et al [22]	Barriers to cross-institutional health information exchange: A literature review	Journal of Healthcare Information Management. 24(3):22-34.	2010	Barriers and facilitators
Flott et al [23]	A patient-centered framework for evaluating digital maturity of health services: A systematic review	Journal of Medical Internet Research. 18(4):e75.	2016	Benefits and impacts
Fontaine et al [15]	Systematic review of health information exchange in primary care practices	Journal of the American Board of Family Medicine. 23(5):655-670.	2010	Benefits and impacts
Hersh et al [16]	Outcomes from health information exchange: Systematic review and future research needs	Journal of Medical Internet Research. 17(12):e39.	2015	Benefits and impacts
Hincapie and Warholak [24]	The impact of health information exchange on health outcomes	Applied Clinical Informatics. 2(4):499-507.	2011	Benefits and impacts
Johnson and Gadd [25]	Playing smallball: Approaches to evaluating pilot health information exchange systems	Journal of Biomedical Informatics. 40(6 Suppl):S21-S26.	2007	Barriers and facilitators
Joshi [26]	Clinical value-add for health information exchange (HIE)	Internet Journal of Medical Informatics. 6(1).	2010	Benefits and impacts
Kruse et al [27]	Barriers over time to full implementation of health information exchange in the United States	JMIR Medical Informatics. 2(2):e26.	2014	Barriers and facilitators
Mastebroek et al [28]	Health information exchange in general practice care for people with intellectual disabilities: A qualitative review of the literature	Research in Developmental Disabilities. 35(9):1978-1987.	2014	Barriers and facilitators
Parker et al [13]	Health information exchanges: Unfulfilled promise as a data source for clinical research	International Journal of Medical Informatics. 87:1-9.	2016	Benefits and impacts
Rahurkar et al [18]	Despite the spread of health information exchange, there is little evidence of its impact on cost, use, and quality of care	Health Affairs. 34(3):476-483.	2015	Benefits and impacts
Rudin et al [17]	Usage and effect of health information exchange: A systematic review	Annals of Internal Medicine. 161(11):803-811.	2014	Benefits and impacts
Vest and Jasperson [29]	What should we measure? Conceptualizing usage in health information exchange	Journal of the American Medical Informatics Association. 17(3):302-307.	2010	Benefits and impacts

Discussion

Principal Findings

Consideration of the review results against the benefits evaluation framework provides a lens to assess where evaluative work has been targeted and where there may be gaps where future evaluative efforts should focus. A total of 57.1% (298/522) of all measurement outcomes were positive. Quality of care (88/136, 64.7%) and productivity (109/170, 64.1%) were the dimensions with the highest percentage of positive measurement outcomes. Prominent themes in the quality of care category were around coordination of care and clinical decision support. For the productivity dimension, efficiency in clinical processes, time savings, and costs were the prominent themes. The left side of the benefits evaluation framework (see Figure 1), including system, service, and information quality, as well as user satisfaction, had relatively lower proportions of positive measurement outcomes ranging from 39%-53%. Many of the factors critical to achieving quality and productivity benefits require concentrated efforts on the left side of the framework. Change management efforts and other studies evaluating the benefits of iEHRs and other information systems suggest that user satisfaction increases when users have access to high-performing technology that is well integrated into their

workflow, interoperable with existing systems, and is able to provide them with the information they need when they need it [2,12,30,31]. In addition, appropriate levels of support and training are necessary to ensure use of information systems.

Overall, our findings suggest that positive results tend to attract more evaluation, which may be explained by efforts to use progressively more rigorous methodologies, but also may reflect inefficient allocation of limited evaluation resources that could be applied to less-studied aspects of iEHRs and HIEs. The findings also point to several broader evaluation dimensions and several specific measurement items that require more attention going forward, including use and access, for which we did not identify any measurement items, and the service quality, system quality, user satisfaction, and information quality dimensions, which had fewer measurement items than productivity and quality dimensions. It is important to note that we did not perform quality appraisals of the primary studies; therefore, these review results need to be interpreted cautiously, which is a general theme of the reviews we assessed. While promising work exists, there is a clear need for more rigorous and comprehensive evaluation, with priority to support methodologies that can produce high-quality evidence. Overall, the review findings highlight the need to support more robust and comprehensive evaluative work across Canada on the impact of connected health information, covering more disease domains, health care settings, and populations.

Limitations

This review identified a large number of studies that address the use and impact of connected health information through iEHRs and HIEs. The majority of the studies have been published within the last 5 years, which reflects a developing rather than mature evidence base. Given that the bulk of this evidence base is current, concerns regarding potential temporal biases that might not accurately reflect the quickly evolving developments of iEHRs and HIEs should be limited. However, our analysis did not assess whether systematic temporal differences in dimension-specific evaluations were present (eg, service and quality evaluated sooner after system launch vs productivity evaluated at more mature stages following launch). Consistent with the developing nature of the evidence base, it is notable that a sizable proportion of the articles identified in this review (12/130, 9.2%) came from grey literature sources, highlighting broader contributions to iEHR and HIE evaluation. Beyond the United States, which is the focus of the vast majority of the primary studies identified, we found few other primary studies globally, with Israel a distant second in terms of evaluative work on iEHRs and HIEs, followed by Canada, Finland, and the United Kingdom. Given fundamental differences in the organization of health systems and services and the dearth of evaluations in non-US settings, there are limits on how generalizable these assessments across jurisdictions will be.

Conclusions

In conclusion, this review captures evaluative work from an evolving and active field focused on the use and impact of iEHRs and HIEs. While the overall findings suggest many positive impacts of iEHRs and HIEs, the quality of the primary studies were not evaluated systematically. When broken down by specific measurement items, some measurement outcomes consistently presented positive outcomes, while others were mostly negative or equivocal, highlighting areas for more attention. Setting-specific findings provide further insight on where more evaluative attention is needed.

Acknowledgments

The authors acknowledge the contributions of Leena Saeed, Angela Du, and Peggy Tso-Ngui in conducting the systematic review. This project was funded by Canada Health Infoway.

Authors' Contributions

MJD and JPB conceived of and designed the study. JPB performed searches and collected data. All authors analyzed and interpreted the data. MJD wrote the first draft of the manuscript. All authors contributed to the writing and/or critical review of the manuscript and agree with the manuscript results and conclusions.

Conflicts of Interest

ST and SH are employees of Canada Health Infoway, an independent, not-for-profit organization funded by the federal government of Canada that helps to improve the health of Canadians by working with partners to accelerate the development, adoption, and effective use of digital health solutions across Canada.

Multimedia Appendix 1

Database-specific search strategies.

[PDF File (Adobe PDF File), 32KB - medinform_v7i2e12607_app1.pdf]

References

1. Gartner. Connected Health Information in Canada: A Benefits Evaluation Study. Toronto, ON: Canada Health Infoway; 2018 Apr. URL: https://tinyurl.com/y3gaoy9a [accessed 2019-05-24] [WebCite Cache ID 78c6uuN4Q]

2. Gheorghiu B, Hagens S. Measuring interoperable EHR adoption and maturity: A Canadian example. BMC Med Inform Decis Mak 2016 Jan 25;16:8 [FREE Full text] [doi: 10.1186/s12911-016-0247-x] [Medline: 26810606]
3. Leaver C. Use of Electronic Medical Records Among Canadian Physicians: 2017 Update. Toronto, ON: Canada Health Infoway; 2017. URL: https://tinyurl.com/y3phx5xn [accessed 2019-05-24] [WebCite Cache ID 78c7m40EV]
4. Environics Research. 2017 National Survey of Canadian Nurses: Use of Digital Health Technology in Practice. Final Executive Report. Toronto, ON: Canada Health Infoway; 2017 May. URL: https://tinyurl.com/y3unwa43 [accessed 2019-05-24] [WebCite Cache ID 78c9NnaLd]
5. Hagens S, Zelmer J, Frazer C, Gheorghiu B, Leaver C. Valuing national effects of digital health investments: An applied method. Stud Health Technol Inform 2015;208:165-169. [Medline: 25676967]
6. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol 2016 Dec;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
7. Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol 2006 Jul 26;6:35 [FREE Full text] [doi: 10.1186/1471-2288-6-35] [Medline: 16872487]
8. Pham B, Robson RC, Thomas SM, Hwee J, Page MJ, Tricco AC. Improving quality and efficiency in selecting, abstracting, and appraising studies for rapid reviews. In: Tricco AC, Langlois EV, Straus SE, editors. Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide. Geneva, Switzerland: World Health Organization; 2017:39-54 URL: https://tinyurl.com/y4l4pxwj
9. Sandelowski M, Docherty S, Emden C. Focus on qualitative methods. Qualitative metasynthesis: Issues and techniques. Res Nurs Health 1997 Aug;20(4):365-371. [Medline: 9256882]
10. Delone WH, McLean ER. The DeLone and McLean Model of Information Systems Success: A ten-year update. J Manag Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
11. Lau F, Hagens S, Muttitt S. A proposed benefits evaluation framework for health information systems in Canada. Healthc Q 2007;10(1):112-116, 118. [Medline: 17326376]
12. Tharmalingam S, Hagens S, Zelmer J. The value of connected health information: Perceptions of electronic health record users in Canada. BMC Med Inform Decis Mak 2016 Dec 16;16:93 [FREE Full text] [doi: 10.1186/s12911-016-0330-3] [Medline: 27422571]
13. Parker C, Weiner M, Reeves M. Health information exchanges: Unfulfilled promise as a data source for clinical research. Int J Med Inform 2016 Mar;87:1-9. [doi: 10.1016/j.ijmedinf.2015.12.005] [Medline: 26806706]
14. Health Quality Ontario. Electronic Tools for Health Information Exchange: An Evidence-Based Analysis. Toronto, ON: Health Quality Ontario; 2013 Sep. URL: https://tinyurl.com/y49jzjfl [accessed 2019-05-24] [WebCite Cache ID 78cDuEYUE]
15. Fontaine P, Ross SE, Zink T, Schilling LM. Systematic review of health information exchange in primary care practices. J Am Board Fam Med 2010;23(5):655-670 [FREE Full text] [doi: 10.3122/jabfm.2010.05.090192] [Medline: 20823361]
16. Hersh WR, Totten AM, Eden KB, Devine B, Gorman P, Kassakian SZ, et al. Outcomes From health information exchange: Systematic review and future research needs. JMIR Med Inform 2015 Dec 15;3(4):e39 [FREE Full text] [doi: 10.2196/medinform.5215] [Medline: 26678413]
17. Rudin RS, Motala A, Goldzweig CL, Shekelle PG. Usage and effect of health information exchange: A systematic review. Ann Intern Med 2014 Dec 02;161(11):803-811. [doi: 10.7326/M14-0877] [Medline: 25437408]
18. Rahurkar S, Vest JR, Menachemi N. Despite the spread of health information exchange, there is little evidence of its impact on cost, use, and quality of care. Health Aff (Millwood) 2015 Mar;34(3):477-483. [doi: 10.1377/hlthaff.2014.0729] [Medline: 25732499]
19. Akhlaq A, McKinstry B, Muhammad KB, Sheikh A. Barriers and facilitators to health information exchange in low- and middle-income country settings: A systematic review. Health Policy Plan 2016 Nov;31(9):1310-1325. [doi: 10.1093/heapol/czw056] [Medline: 27185528]
20. Dobrov A, Jones T, Kersting A, Artmann J, Streotmann KA, Stroetmann VN. Report on Methodology for Evaluating the Socioeconomic Impact of Interoperable EHR and ePrescribing Systems. Brussels, Belgium: European Commission; 2008 Jun. URL: http://www.ehr-impact.eu/downloads/documents/EHRI_D1_3_Evaluation_Methodology_v1_0.pdf [accessed 2019-05-28] [WebCite Cache ID 78i3v4mRh]
21. Eden KB, Totten AM, Kassakian SZ, Gorman PN, McDonagh MS, Devine B, et al. Barriers and facilitators to exchanging health information: A systematic review. Int J Med Inform 2016 Apr;88:44-51 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.01.004] [Medline: 26878761]
22. Edwards A, Hollin I, Barry J, Kachnowski S. Barriers to cross-institutional health information exchange: A literature review. J Healthc Inf Manag 2010;24(3):22-34. [Medline: 20677469]
23. Flott K, Callahan R, Darzi A, Mayer E. A patient-centered framework for evaluating digital maturity of health services: A systematic review. J Med Internet Res 2016 Apr 14;18(4):e75 [FREE Full text] [doi: 10.2196/jmir.5047] [Medline: 27080852]
24. Hincapie A, Warholak T. The impact of health information exchange on health outcomes. Appl Clin Inform 2011;2(4):499-507 [FREE Full text] [doi: 10.4338/ACI-2011-05-R-0027] [Medline: 23616891]

25. Johnson KB, Gadd C. Playing smallball: Approaches to evaluating pilot health information exchange systems. J Biomed Inform 2007 Dec;40(6 Suppl):S21-S26 [FREE Full text] [doi: 10.1016/j.jbi.2007.08.006] [Medline: 17931981]
26. Joshi J. Clinical value-add for health information exchange (HIE). Internet J Med Inform 2010;6(1):1-9 [FREE Full text]
27. Kruse CS, Regier V, Rheinboldt KT. Barriers over time to full implementation of health information exchange in the United States. JMIR Med Inform 2014 Sep 30;2(2):e26 [FREE Full text] [doi: 10.2196/medinform.3625] [Medline: 25600635]
28. Mastebroek M, Naaldenberg J, Lagro-Janssen AL, van Schrojenstein Lantman de Valk H. Health information exchange in general practice care for people with intellectual disabilities: A qualitative review of the literature. Res Dev Disabil 2014 Sep;35(9):1978-1987. [doi: 10.1016/j.ridd.2014.04.029] [Medline: 24864050]
29. Vest JR, Jasperson J. What should we measure? Conceptualizing usage in health information exchange. J Am Med Inform Assoc 2010;17(3):302-307 [FREE Full text] [doi: 10.1136/jamia.2009.000471] [Medline: 20442148]
30. Pan-Canadian Change Management Network. A Framework and Toolkit for Managing eHealth Change: People and Processes. Toronto, ON: Canada Health Infoway; 2013 Mar. URL: https://tinyurl.com/y2vhvm7r [accessed 2019-05-24] [WebCite Cache ID 78cEDCm5z]
31. Kuhn K, Lau F. Evaluation of a shared electronic health record. Healthc Q 2014;17(1):30-35. [Medline: 24844718]

Abbreviations

CINAHL: Cumulative Index to Nursing and Allied Health Literature
EHR: electronic health record
HIE: health information exchange
iEHR: interoperable electronic health record
PRESS: Peer Review of Electronic Search Strategies
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 30.10.18; peer-reviewed by S Rahurkar, B Gheorghiu; comments to author 21.11.18; revised version received 31.03.19; accepted 21.04.19; published 06.06.19.

Please cite as:
Dobrow MJ, Bytautas JP, Tharmalingam S, Hagens S
Interoperable Electronic Health Records and Health Information Exchanges: Systematic Review
JMIR Med Inform 2019;7(2):e12607
URL: http://medinform.jmir.org/2019/2/e12607/
doi:10.2196/12607
PMID:31172961

Original Paper

Establishment of a New Platform for the Management of Patients After Cardiac Surgery: Descriptive Study

Zhihui Zhu[1], BA; Haibo Zhang[1], MD; Yuehuan Li[1], MA; Xu Meng[1], MD; Yuqing Jiao[1], MD; Ying Qin[1], MA; Yan Li[1], MD; Jie Han[1], BA; Kun Liu[1], MA; Chongyang Liu[1], MD

Beijing Anzhen Hospital, Capital Medical University, Beijing, China

Corresponding Author:
Haibo Zhang, MD
Beijing Anzhen Hospital
Capital Medical University
No. 2, Anzhen Road
Chaoyang District
Beijing,
China
Phone: 86 13370103561
Email: zhanghb2318@163.com

Abstract

Background: Medical care for the Chinese population has been focused on first-line treatment, but with little follow-up on treated patients. As an important part of clinical work, follow-up evaluations are of great significance for the long-term survival of patients and for clinical and scientific research. However, the overall follow-up rate of discharged patients after surgery has been low for many years because of the limitations of certain follow-up methods and the presence of objective, practical problems.

Objective: This study aimed to construct a new two-way interactive telemedicine follow-up platform to improve the collection of clinical data after cardiac surgery and provide reliable and high-quality follow-up services.

Methods: Computer and network technologies were employed in the context of "Internet +" to develop follow-up databases and software compatible with a mobile network. Postoperative follow-up quality data including the follow-up rate and important postoperative indices were used as standards to evaluate the new follow-up management model after cardiac surgery.

Results: This system has been officially operated for more than 5 years. A total of 5347 patients undergoing cardiac surgery have been enrolled, and the total follow-up rate was 90.22%. In addition, 6349 echocardiographic images, 4717 electrocardiographic images, and 3504 chest radiographic images have been uploaded during follow-up assessments. The international standardized ratio was 20,696 person-times.

Conclusions: This new management follow-up platform can be used to effectively collect clinical data, provide technical support for academic research, extend medical services, and provide more help to patients. It is of great significance for managing patients after cardiac surgery.

(JMIR Med Inform 2019;7(2):e13123) doi:10.2196/13123

KEYWORDS

follow-up; cardiac surgery; telemedicine

Introduction

As an important part of clinical work, follow-up evaluations are of great significance for the long-term survival of patients and for clinical and scientific research. However, the overall follow-up rate of discharged patients after surgery has been low for many years because of the limitations of certain follow-up methods and the presence of objective, practical problems [1-2]. Collection of large-scale multiagency data in clinical practice is accepted and encouraged for cardiac surgery. From a research perspective, database research can help increase the understanding of cardiac surgery. Among other uses, this research enables the investigation of the effects of disease morbidity and mortality, high-risk groups, differences in health care services, and new equipment and technologies [3].

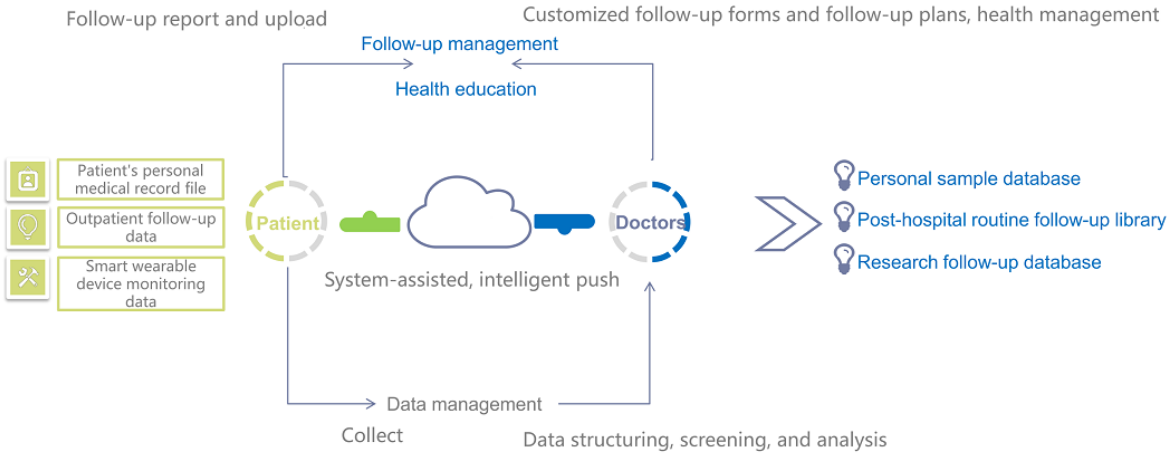
Medical activities in China tend to focus on "first-line treatment," and follow-up evaluations of patients are often overlooked. Thus, only the treatment of patients during

hospitalization is valued [4]. Traditional patient follow-up methods include outpatient review, text message follow-up, telephone follow-up, WeChat follow-up, and letter follow-up [5-9]. The main factors restricting patient participation in posthospital follow-up include the following: (1) Time effectiveness: Patient enthusiasm in participating in follow-up decreases as the time after discharge increases; (2) Age: Most major diseases affect aging patients, and their follow-up participation rate is low; (3) Communication and cost: Traditional means of contact restrict effective communication between patients and doctors, and most doctors and hospitals require re-examination of patients in a hospital (therefore, the additional costs of transportation, accommodations, meals, and registration are factors); (4) Supervision and guidance: Although patients have basic follow-up training after hospital discharge, they do not participate in the follow-up because they are not supervised; and (5) Geographical restrictions: The follow-up of nonlocal patients in key hospitals is a problem. Therefore, the compliance and follow-up rates of most patients in long-term postoperative follow-up care annually decline, and the patients often drop out.

In order to improve collection of clinical data after cardiac surgery and provide reliable and high-quality follow-up services, we tried to construct a new two-way interactive telemedicine follow-up platform. In 2000, Professor Meng of the Beijing Heart Transplantation and Valvular Surgery Center first applied the informatization management system in clinical care, teaching, and scientific research in China and established the cardiac surgery database. This database was primarily based on the permanent preservation of cardiac surgery information as well as scientific summary and research. Thus far, the database includes more than 20,000 cases of cardiac surgery, and after continuous improvement, the database has operated smoothly. In 2011, based on the original cardiac surgery database, using modern communication tools such as computers and internet technologies, the Beijing Heart Transplantation and Valvular Surgery Center began to work with professional software companies to establish a new two-way interactive telemedicine follow-up system for patients who have been discharged after surgery (see Figure 1 for the problems solved by the new follow-up management system).

Based on the computer network terminals, this telemedicine follow-up system considers the mobile client app as a popular way to coordinate using communication channels such as WeChat, short message service (SMS), and mailbox. The aim of this telemedicine follow-up system is to block abnormal disease processes and improve the quality of life of patients after surgery. This telemedicine follow-up system uses healthy concept inputs, reminders, and recordings as interventions. Through the series of core application functions and intelligent auxiliary methods of this telemedicine follow-up system, the problems encountered by doctors during the follow-up assessments of patients are solved. The follow-up system has been operating officially for more than 5 years, and it has gradually been applied to the cardiac surgery departments of seven hospitals in Beijing. The preparation, operation, maintenance, and follow-up effect of this new follow-up system are summarized below.

Figure 1. Problems solved by the new follow-up management system.



Methods

Establishment of the Follow-Up Platform

According to the characteristics of thoracic and cardiovascular surgery, the postoperative follow-up requirements of different surgical categories were collected and sorted and the postoperative follow-up index system and follow-up knowledge base were established. Professionals develop follow-up databases and related software that are compatible with mobile networks.

Development of the Follow-Up System and Standardization of Follow-Up Procedures

Patient Source

All patients were discharged from the Beijing Heart Transplantation and Valvular Surgery Center after surgery.

Patient Training and Registration

An electronic information registration file and a completed follow-up network registration were established for each patient discharged from the hospital after surgery. The information file

includes the patients' names, genders, dates of birth, contact telephone numbers, hospital numbers, times of admission and discharge, diagnoses, and types of surgery.

All patients and their family members participated in at least three follow-up training sessions before completing the network registration, including watching the mission video, installing the mobile app of the follow-up software, and using the app. Information registration and training for patients discharged from the hospital were completed through the network after confirmation from the follow-up management doctor.

Follow-Up Medical Staff Configuration and Workflow

Follow-Up Principle

The significance of the follow-up platform lies in the entire course of the follow-up assessment, with a review and warning of abnormal indices as the intervention method, which can be distinguished from a general remote follow-up evaluation. In principle, the patient's report was automatically returned by the follow-up platform. Figure 2 shows the follow-up workflow of the Beijing Heart Transplantation and Valvular Surgery Center.

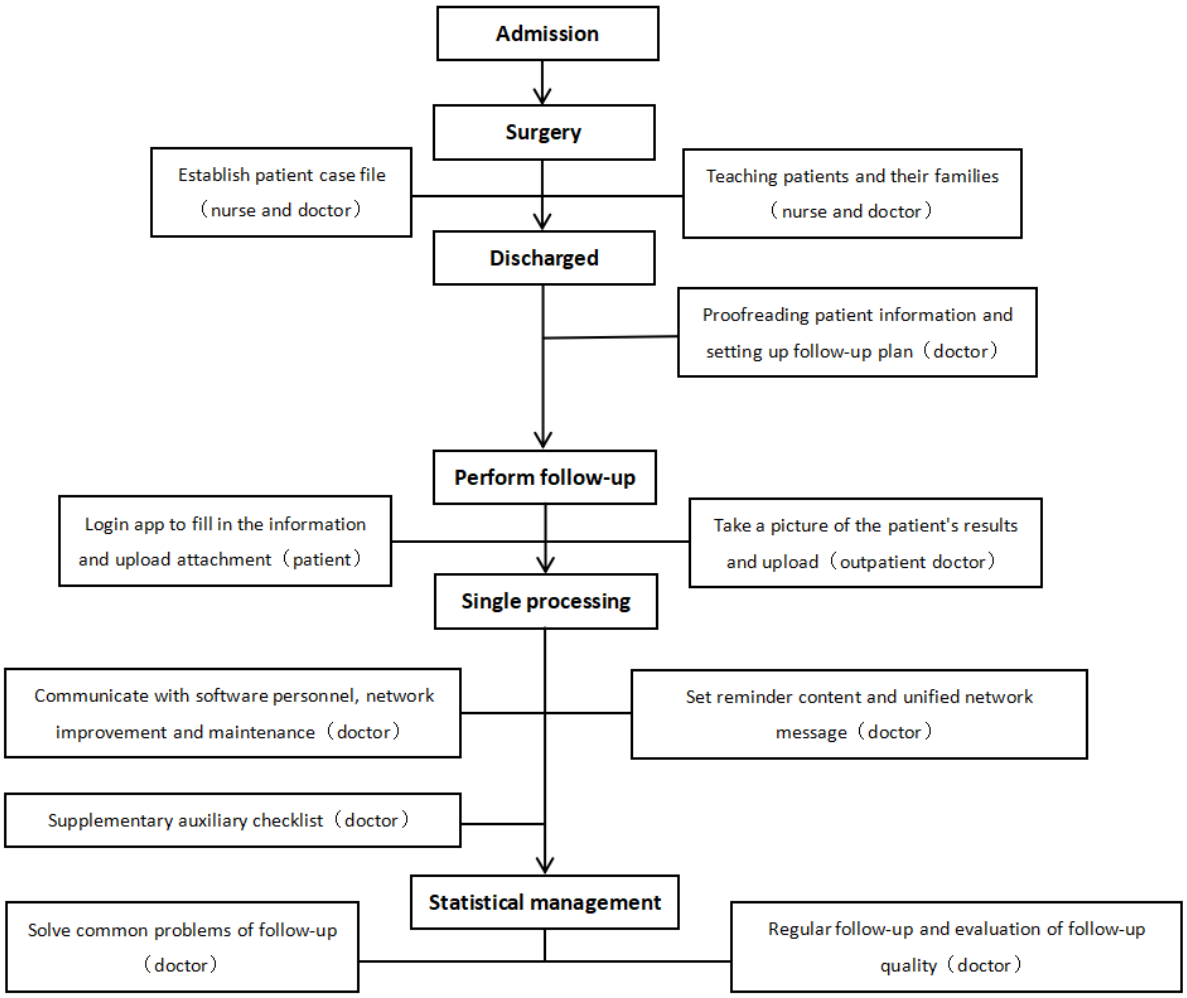
Medical Staff Requirements

Follow-up work was conducted under the leadership of the relevant person from the center, and participation of all medical staff was mandatory. The medical staff of the center were obligated to remind, guide, and strengthen the follow-up care of postoperative patients.

Statistical Analysis

The follow-up data were exported into an Excel spreadsheet. The patients had to complete at least one report (ie, the reference standard), and follow-up statistics were completed using SPSS 20.0 (IBM Corp, Armonk, NY).

Figure 2. Follow-up workflow of the Beijing Heart Transplantation and Valvular Surgery Center.



Results

Establishment of a New Two-Way Interactive Telemedicine Follow-Up Platform

Reminders

Personalized reminders for follow-up questionnaires and the review period were developed based on the diagnosis and surgical approach, which is conducive to enhancing patients' periodic review and completion of a health self-test.

Preliminary Judgment and Comparison Function

For the follow-up indicators and checklists, the system has built-in preliminary judgment and dynamic comparison functions that help patients improve their health awareness and detect problems early.

Health Records

The patients' follow-up questionnaires and the uploaded review data are recorded permanently.

Questions and Answers and Training Function

The system regularly summarizes patients' reports and answers common questions.

Clinical Research Assistance

The system can screen, extract, and provide preliminary statistics of the follow-up data.

Follow-Up List Settings

The follow-up list is determined by the diagnosis and surgical approach. Each patient generally completes two to three follow-up questionnaires, including the regular follow-up, the 36-item Short-Form (SF36) quality of life–assessment scale, and the international standardized ratio (INR) follow-up (for those taking warfarin). According to the actual situation of the center, the routine follow-up list is divided into five categories corresponding to valvular disease/atrial fibrillation (conventional item 1), coronary artery bypass grafting (conventional item 2), valvular disease/atrial fibrillation with coronary artery bypass grafting (conventional item 3), cardiac transplantation (conventional item 4), and congenital heart disease without valvular disease and pericarditis (conventional item 5). Before the patient is discharged from the hospital, the follow-up settings are confirmed via text message.

Registration of Surgical Patients

By the end of October 2018, 5347 patients receiving cardiac surgery were registered. Of these patients, 4522 were discharged after establishment of the follow-up platform and 825 were discharged before the establishment of the follow-up platform (including 46 deaths). The ages of the patients ranged from 3 to 90 years (average, 51.95 years). A total of 975 patients were aged over 65 years, accounting for 18.3% of the patient population. The male-to-female ratio was 52:48. Further, 3844 patients had heart valve surgery, accounting for 71.89%. There were 259 cases of cardiac transplantation, accounting for 4.84%. The earliest patients were registered was 21 years before the start of this study.

Follow-Up Details

Some patients do not use the new follow-up method for personal reasons, and we excluded data of these patients. Of the 5234 patients who were enrolled in the follow-up database, 4722 participated in the follow-up, accounting for 90.22%. Of these patients, 2408 were men and 2314 were women. The male-to-female ratio was 1:1, and their ages ranged between 3 and 90 years (average, 51.99 years). Of the 512 patients without follow-up compliance, 267 were men and 245 were women. The male-to-female ratio was 1:1, and their ages ranged between 3 and 85 years (average, 52.95 years). Of the 1021 patients who were discharged from the hospital ≤1 year ago, 893 participated during the follow-up period, and the follow-up rate was 87.46%. Of the 1703 patients who were discharged from the hospital 1-3 years ago, 1553 participated during the follow-up period, and the follow-up rate was 91.19%. Of the 1299 patients who were discharged from the hospital 3-5 years ago, 1193 participated during the follow-up period, and the follow-up rate was 91.84%. Of the 1008 patients who were discharged from the hospital 5-10 years ago, 905 participated during the follow-up period, and the follow-up rate was 89.78%. Of the 203 patients who were discharged from the hospital >10 years ago, 178 participated during the follow-up period, and the follow-up rate was 87.68%.

Of the important follow-up indicators, echocardiography, electrocardiography, chest radiography, blood test/blood biochemistry, INR, and the SF36 scale scores were associated with 6349, 4717, 3504, 1876, 20,696, and 1029 person-times, respectively.

Registration of patients who were discharged before the establishment of the follow-up platform was completed during the follow-up period. Therefore, to reflect the follow-up more objectively, patients who were discharged before the establishment of the follow-up platform were excluded and those who were discharged afterward were analyzed separately. This phase had 4522 registered patients; of these, 3896 participated during the follow-up period, accounting for 86.16%. In addition, 57 patients died after discharge, 19 of whom were involved in the follow-up assessment.

A total of 2497 patients underwent ultrasound examination 4999 times, 1983 patients uploaded electrocardiogram results 3928 times, 1856 patients uploaded chest x-ray results 3021 times, and 898 patients uploaded blood test results 1608 times. Echocardiography, an important indicator in this review, was to be performed during the 3-month follow-up assessment; therefore, patients who were discharged from the hospital less than 3 months ago were excluded, and the echocardiography, electrocardiogram, chest x-ray, and blood test follow-up rates were 75.21% (1732/2303), 59.9% (1074/1793), 54.86% (914/1666), and 25.68% (208/810), respectively.

After mechanical heart valve replacement, patients require long-term use of warfarin, and this group of patients was analyzed separately. Of the registered patients, 1513 underwent mechanical valve replacement, and 1390 completed 16,611 follow-up assessments. The follow-up rate was 91.87%. A total of 9983 patients participated in the INR test; the mean INR was 2.21, and the compliance rate was 57.47%.

Telephone Interviews With Patients Lost to Follow-Up

After the establishment of the follow-up platform, 1513 discharged patients underwent mechanical valve replacement. The telephone interview results of the 123 patients who did not participate in the follow-up assessment were as follows: 34 patients did not pay attention to postoperative follow-up; 18 patients had mobile phone problems (the primary contact number was a relative's phone number, not in service, no longer listed, or incorrect) and did not receive reminder text messages; 32 elderly patients did not know how to use the internet, the app, or text messaging; 21 patients directly contacted the doctor in charge; and 15 patients were hospitalized at a local hospital or during the follow-up of the local hospital. In addition, 3 patients had not been examined for INR after discharge.

Discussion

Changing the Follow-Up Concept to Improve Compliance

Innovations in information technology have fundamentally changed the way patients perceive time and distance; moreover, they have reshaped the way they interact and connect with others, including how they interact with medicine. As the public becomes more adept with using new technologies in all aspects of daily life, evolving applications in health care will change where and how patients and doctors interact [10]. Over the past four decades, telemedicine has become an increasingly cost-effective alternative to traditional "face-to-face" medicine and has evolved into an integrated technology used in hospitals, clinics, patients' homes, and many other environments. A growing body of literature has shown that telemedicine electronic media can connect doctors with other doctors, patients with doctors, and even patients with other patients, making clinical treatment and rehabilitation more effective and efficient [11].

Even the most rigorously designed randomized controlled trials are at risk for high rates of loss to follow-up [12]. For many years, improving patient follow-up concepts and compliance has been a major problem in clinical work [13-15]. The biggest innovation of this follow-up system is that patients actively, not passively, complete its report. This innovative concept brings sustainable, self-recognized benefits to patients, and the follow-up system helps patients create profiles. With this follow-up system, patients can describe their condition to the doctor at the next visit. This platform can help patients judge their physical condition and seek medical treatment in time. Past outpatient reviews and remote medical follow-up consultations have focused on individualized diagnosis and treatment; however, it is difficult to address patients' growing follow-up content in terms of medical risks and costs, if there is no guarantee from the corresponding medical or rehabilitation system. This system can help doctors manage group information of their patients. Doctors can track patients through this system, enriching patients' data sources. At the same time, this follow-up system can help doctors structure the checklist and build more data sources that can be extracted, analyzed, and applied.

The emergence of electronic health tools provides patients with more opportunities. By learning about specific diseases and receiving regular feedback and frequently enhanced choices, patient engagement can be increased [16]. Solid evidence shows that secondary prevention and cardiac rehabilitation programs reduce mortality among patients with cardiovascular disease after surgery. Internet applications and mobile platforms have broad prospects [13]. The emergence of the follow-up platform helps doctors establish a communication platform for doctors and patients and realize the transmission of health education. The follow-up platform frees some of the doctor's work during the follow-up. Under limited human and financial conditions, this follow-up platform changed the concept of the follow-up assessment and explored a new balance among patient benefits, costs, and risks.

Through the follow-up training, continuous updating of questions and answers between doctors and patients improved patient health awareness. Individualized follow-up questionnaires, built-in personalized follow-up questionnaires and review periods, system-automated judgments, reminders, and dynamic comparison functions might help patients' self-test and postoperative adjustment. The health file records basic patient information and postoperative dynamic changes and facilitates future visits and reviews. All system judgments and responses are derived from the built-in procedures of the network, which reduces human hours and time costs and avoids medical risks as much as possible.

Factors Affecting the Follow-Up Rate

Disease Development

Smith et al [17] believed that disease development is an important factor that affects the follow-up rate. Patients with poor prognoses and corresponding complications tend not to participate in follow-up evaluations. However, tracking patients who had been lost to follow-up for more than 10 years, Dexter et al [18] found that these patients were relatively young, were predominately women, and had fewer complications. During our follow-up period, the follow-up rate was 29.8% among the 57 patients who died, which was much lower than the overall follow-up rate. Of the 123 patients lost to follow-up after mechanical valve replacement, 12 (9.8%) had complications; this rate is also significantly higher than that of follow-up in the general population, which suggests that the short-term follow-up rate is associated with disease development. The statistical analysis of all registered patients did not show effects of age or gender on follow-up. In addition, it is difficult to assess the effect of complications on the follow-up period because of the brevity of the follow-up period.

Follow-Up Method

The patient follow-up period revealed that speed, the cumbersome nature of computer-based image imports, and internet access location are important factors that restrict the upload of patient checklists. The development of a mobile phone app can solve these problems [19-22]. The new management system provides a mobile phone app for patients, which makes it convenient for patients to upload their own re-examination reports and photos, thereby increasing the portability of the

report and making the report more convenient for storage. Furthermore, the system reminds the patient to check and re-examine themselves regularly in the fixed period during the follow-up, which is helpful for patients to receive more active formal postoperative rehabilitation guidance.

Some patients will be familiar with the doctors and will establish channels of communication with doctors, such as telephone, WeChat, and SMS; however, the doctor's contact information is publicly disclosed, which will affect the doctor's daily life. The follow-up platform can be used for communication between doctors and patients. Through the patient's follow-up report, the doctor can control the entire communication process. Through the system's automatic evaluation of indicators, doctors can choose to communicate only with patients in need. This greatly enhances the follow-up efficiency of doctors. Regarding the mobile aspect, doctors can manage their patients anytime, anywhere.

Outpatient tracking is an important follow-up method after patient discharge [9,23]. Although new technologies and treatment models continue to emerge, the basic ethical responsibilities of doctors have not changed. Medical practice is essentially an ethical activity built on the trust between patients and doctors [24]. Effective communication between these parties is the cornerstone of building trust and providing quality health care [25]. Using this follow-up system, the attending doctor can upload the early outpatient checklist for patients; thus, these patients can go directly to the doctors of the center after the review because the patients have had offline contact with the doctors, thereby establishing a real doctor-patient connection. Furthermore, the patients truly realize that a postoperative follow-up assessment is a key part of the entire rehabilitation based on the center's satisfactory follow-up training; therefore, they are able to participate in postoperative follow-up care more proactively.

Patient Rights and Network Security

Although telemedicine innovation has an enormous potential to benefit patients, it also presents new ethical and security challenges. Of particular concern is that the exchange of patients' health information and the provision of treatment and training through telemedicine might create new risks to the quality, safety, and continuity of care that could affect the doctor-patient relationship [26-29]. In addition to collection of patient hospitalization information, the establishment of this follow-up platform enables each patient to track the dynamics of the event as much as possible. The follow-up platform is equipped with a follow-up closure function, and the end-point event can be turned off to avoid disturbing the patient. From the patient's terminal, resetting the password after the first login should help protect patients' personal information. From the doctor's terminal, the multirole authority management setting can assign different rights to medical staff, which greatly protects patient privacy and data security.

Work Plan of the Follow-Up Platform

Because of the brief operation time of the follow-up platform, it is difficult to complete follow-up statistics at many levels. Currently, the activity of the patient is not high enough, and it is necessary to find and optimize derivative services that can be provided to patients to fully mobilize patient participation and compliance. Parker et al [30] analyzed the use of electronics and mobile internet apps among vulnerable groups across 18 telemedicine studies and suggested that encouraging goal setting, providing achievement rewards, strengthening patient responsibility for symptom monitoring, and providing educational guidance and additional support can improve patients' abilities to manage their diseases. Our future work will improve service design at the patient terminal, optimize the interactive experience, calculate the follow-up rate and follow-up quality of the different follow-up stages, complete the statistics of the follow-up methods and influencing factors, and screen and process the follow-up data for different projects. In addition, we will increase the popularization of science education, provide regular updates, and strengthen interactions with patients. In summary, the follow-up platform will be evaluated and continuously refined from a sustainability perspective.

This follow-up system completes the collection of clinical data after cardiac surgery and provides reliable and quality follow-up services for patients. Based on modern communication technologies, such as computer networks and data storage technology, a new type of follow-up management mode for thoracic and cardiovascular surgery based on wired and wireless network communication was established using the "Internet + medical" operation management mode, supported by the knowledge base of clinical medical diagnosis and treatment. The follow-up management mode provides an open, flexible, and efficient communication platform for doctors, researchers, and discharged patients. Through standardized follow-up behavior and follow-up methods, postoperative health education is strengthened and service is improved. This platform is oriented toward patients, serves clinical research, and provides technical support to comprehensively improve overall medical services and academic research.

Acknowledgments

This work was supported by the Beijing Science and Technology Commission of China (new warfarin anticoagulation management mode after heart valve operation; grant number: D171100002917003)

Conflicts of Interest

None declared.

References

1. Lovell ME, Morcuende JA. Patient location strategies for pediatric long-term follow-up studies. Iowa Orthop J 2006;26:91-95 [FREE Full text] [Medline: 16789455]
2. Rivara FP, Koepsell TD, Wang J, Durbin D, Jaffe KM, Vavilala M, et al. Comparison of telephone with World Wide Web-based responses by parents and teens to a follow-up survey after injury. Health Serv Res 2011 Jun;46(3):964-981 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01236.x] [Medline: 21275989]
3. Saxena A, Newcomb AE, Dhurandhar V, Bannon PG. Application of clinical databases to contemporary cardiac surgery practice: Where are we now? Heart, Lung and Circulation 2016 Mar;25(3):237-242. [doi: 10.1016/j.hlc.2015.01.006]
4. Fang J, Huang B, Xu D, Li J, Au WW. Innovative Application of a Home-Based and Remote Sensing Cardiac Rehabilitation Protocol in Chinese Patients After Percutaneous Coronary Intervention. Telemedicine and e-Health 2018 Sep 07. [doi: 10.1089/tmj.2018.0064]
5. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013;12:CD007458. [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
6. Lyu K, Zhao J, Wang B, Xiong G, Yang W, Liu Q, et al. Smartphone Application WeChat for Clinical Follow-up of Discharged Patients with Head and Neck Tumors: A Randomized Controlled Trial. Chin Med J (Engl) 2016 Dec 05;129(23):2816-2823 [FREE Full text] [doi: 10.4103/0366-6999.194635] [Medline: 27900995]
7. Palmer MJ, Barnard S, Perel P, Free C. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. Cochrane Database Syst Rev 2018 Dec 22;6:CD012675. [doi: 10.1002/14651858.CD012675.pub2] [Medline: 29932455]
8. Atherton H, Sawmynaden P, Sheikh A, Majeed A, Car J. Email for clinical communication between patients/caregivers and healthcare professionals. Cochrane Database Syst Rev 2012 Nov 14;11:CD007978. [doi: 10.1002/14651858.CD007978.pub2] [Medline: 23152249]
9. Schougaard LMV, Mejdahl CT, Petersen KH, Jessen A, de Thurah A, Sidenius P, et al. Effect of patient-initiated versus fixed-interval telePRO-based outpatient follow-up: study protocol for a pragmatic randomised controlled study. BMC Health Serv Res 2017 Jan 26;17(1). [doi: 10.1186/s12913-017-2015-8]
10. Chaet D, Clearfield R, Sabin JE, Skimming K, Council on Ethical Judicial Affairs American Medical Association. Ethical Practice in Telehealth and Telemedicine. J Gen Intern Med 2017 Oct 26;32(10):1136-1140 [FREE Full text] [doi: 10.1007/s11606-017-4082-2] [Medline: 28653233]
11. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
12. Madden K, Scott T, McKay P, Petrisor BA, Jeray KJ, Tanner SL, et al. Predicting and Preventing Loss to Follow-up of Adult Trauma Patients in Randomized Controlled Trials. The Journal of Bone and Joint Surgery 2017;99(13):1086-1092. [doi: 10.2106/JBJS.16.00900]
13. Antypas K, Wangberg SC. An Internet- and mobile-based tailored intervention to enhance maintenance of physical activity after cardiac rehabilitation: short-term results of a randomized controlled trial. J Med Internet Res 2014 Mar 11;16(3):e77 [FREE Full text] [doi: 10.2196/jmir.3132] [Medline: 24618349]
14. Hamilton SJ, Mills B, Birch EM, Thompson SC. Smartphones in the secondary prevention of cardiovascular disease: a systematic review. BMC Cardiovasc Disord 2018 Feb 7;18(1). [doi: 10.1186/s12872-018-0764-x]
15. Moradi A, Moeini M, Sanei H. The effect of interactive text message follow-up on health promoting lifestyle of patients with acute coronary syndrome. Iranian J Nursing Midwifery Res 2017;22(4):287. [doi: 10.4103/ijnmr.IJNMR_89_16]
16. Ciere Y, Cartwright M, Newman SP. A systematic review of the mediating role of knowledge, self-efficacy and self-care behaviour in telehealth patients with heart failure. J Telemed Telecare 2012 Sep 27;18(7):384-391. [doi: 10.1258/jtt.2012.111009]
17. Smith JS, Watts HG. Methods for locating missing patients for the purpose of long-term clini-cal studiesJ. J Bone Joint Surg Am 1998;80(3):431-438 [FREE Full text] [Medline: 9531213]
18. Louie DL, Earp BE, Blazar PE. Finding Orthopedic Patients Lost to Follow-up for Long-term Outcomes Research Using the Internet: An Update for 2012. Orthopedics 2012 Jul 01;35(7):595-599. [doi: 10.3928/01477447-20120621-06]
19. Vezzoni G, Guazzelli M, Barachini P. The. G Ital Dermatol Venereol 2011 Apr;146(2):95-101. [Medline: 21505395]
20. Ye J, Zuo Y, Xie T, Wu M, Ni P, Kang Y, et al. A telemedicine wound care model using 4G with smart phones or smart glasses: A pilot study. Medicine (Baltimore) 2016 Aug;95(31):e4198 [FREE Full text] [doi: 10.1097/MD.0000000000004198] [Medline: 27495023]
21. Elsbernd A, Hjerming M, Visler C, Hjalgrim LL, Niemann CU, Boisen K, et al. Cocreated Smartphone App to Improve the Quality of Life of Adolescents and Young Adults with Cancer (Kræftværket): Protocol for a Quantitative and Qualitative Evaluation. JMIR Res Protoc 2018 May 10;7(5):e10098. [doi: 10.2196/10098]
22. Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27(7):MR000042. [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]

23. Emberton M. Outpatient follow up. BMJ 1995 Nov 18;311(7016):1315-1316 [FREE Full text] [doi: 10.1136/bmj.311.7016.1315]
24. Pellegrino ED. Professionalism, profession and the virtues of the good physician. Mt Sinai J Med 2002 Nov;69(6):378-384. [Medline: 12429956]
25. Aung K, Wu WK, Tokumi A, Kuo P, Day CS. Does a Directive to an Internet Site Enhance the Doctor-Patient Interaction? The Journal of Bone and Joint Surgery 2015;97(13):1112-1118. [doi: 10.2106/JBJS.N.00741]
26. Miller TE, Derse AR. Between strangers: the practice of medicine online. Health Aff (Millwood) 2002;21(4):168-179. [Medline: 12117127]
27. Hall JL, McGraw D. For Telehealth To Succeed, Privacy And Security Risks Must Be Identified And Addressed. Health Affairs 2014 Feb;33(2):216-221. [doi: 10.1377/hlthaff.2013.0997]
28. Agha Z, Schapira RM, Laud PW, McNutt G, Roter DL. Patient Satisfaction with Physician–Patient Communication During Telemedicine. Telemedicine and e-Health 2009 Nov;15(9):830-839. [doi: 10.1089/tmj.2009.0030]
29. Huesch MD. Privacy Threats When Seeking Online Health Information. JAMA Intern Med 2013 Oct 28;173(19):1838. [doi: 10.1001/jamainternmed.2013.7795]
30. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris MF. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192. [doi: 10.1136/bmjopen-2017-019192]

Abbreviations

SMS: short message service
SF36: 36-item Short Form
INR: international standardized ratio

Edited by G Eysenbach; submitted 13.12.18; peer-reviewed by H Yang, ZH Sun; comments to author 23.01.19; revised version received 25.01.19; accepted 27.01.19; published 01.04.19.

Please cite as:
Zhu Z, Zhang H, Li Y, Meng X, Jiao Y, Qin Y, Li Y, Han J, Liu K, Liu C
Establishment of a New Platform for the Management of Patients After Cardiac Surgery: Descriptive Study
JMIR Med Inform 2019;7(2):e13123
URL: https://medinform.jmir.org/2019/2/e13123/
doi:10.2196/13123
PMID:30932865

Original Paper

Identifying Clinical Terms in Medical Text Using Ontology-Guided Machine Learning

Aryan Arbabi[1,2], MSc; David R Adams[3], MD, PhD; Sanja Fidler[1], PhD; Michael Brudno[1,2], PhD

[1]Department of Computer Science, University of Toronto, Toronto, ON, Canada

[2]Centre for Computational Medicine, Hospital for Sick Children, Toronto, ON, Canada

[3]Section on Human Biochemical Genetics, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, United States

Corresponding Author:
Michael Brudno, PhD
Department of Computer Science
University of Toronto
10 King's College Road, SF 3303
Toronto, ON, M5S 3G4
Canada
Phone: 1 4169782589
Email: brudno@cs.toronto.edu

Abstract

Background: Automatic recognition of medical concepts in unstructured text is an important component of many clinical and research applications, and its accuracy has a large impact on electronic health record analysis. The mining of medical concepts is complicated by the broad use of synonyms and nonstandard terms in medical documents.

Objective: We present a machine learning model for concept recognition in large unstructured text, which optimizes the use of ontological structures and can identify previously unobserved synonyms for concepts in the ontology.

Methods: We present a neural dictionary model that can be used to predict if a phrase is synonymous to a concept in a reference ontology. Our model, called the Neural Concept Recognizer (NCR), uses a convolutional neural network to encode input phrases and then rank medical concepts based on the similarity in that space. It uses the hierarchical structure provided by the biomedical ontology as an implicit prior embedding to better learn embedding of various terms. We trained our model on two biomedical ontologies—the Human Phenotype Ontology (HPO) and Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT).

Results: We tested our model trained on HPO by using two different data sets: 288 annotated PubMed abstracts and 39 clinical reports. We achieved 1.7%-3% higher F1-scores than those for our strongest manually engineered rule-based baselines (P=.003). We also tested our model trained on the SNOMED-CT by using 2000 Intensive Care Unit discharge summaries from MIMIC (Multiparameter Intelligent Monitoring in Intensive Care) and achieved 0.9%-1.3% higher F1-scores than those of our baseline. The results of our experiments show high accuracy of our model as well as the value of using the taxonomy structure of the ontology in concept recognition.

Conclusion: Most popular medical concept recognizers rely on rule-based models, which cannot generalize well to unseen synonyms. In addition, most machine learning methods typically require large corpora of annotated text that cover all classes of concepts, which can be extremely difficult to obtain for biomedical ontologies. Without relying on large-scale labeled training data or requiring any custom training, our model can be efficiently generalized to new synonyms and performs as well or better than state-of-the-art methods custom built for specific ontologies.

(JMIR Med Inform 2019;7(2):e12596) doi:10.2196/12596

KEYWORDS

concept recognition; medical text mining; biomedical ontologies; machine learning; phenotyping; human phenotype ontology

Introduction

Background

Automatic recognition of medical concepts in unstructured text is a key component of biomedical information retrieval systems. Its applications include analysis of unstructured text in electronic health records (EHR) [1-3] and knowledge discovery from biomedical literature [4,5]. Many medical terminologies are structured as ontologies, adding relations between concepts and often including several synonyms for each term. One of the most widely used ontologies in the medical space is SNOMED-CT (Systematized Nomenclature of Medicine - Clinical Terms) [6], which provides structured relationships for over 300,000 medical concepts. SNOMED-CT is commonly used in EHR Systems to help summarize patient encounters and is fully integrated with the International Classification of Diseases - Ninth Revision (ICD-9) billing codes used in the United States and many other jurisdictions. The Human Phenotype Ontology (HPO) [7] is an arrangement of terms used to describe the visible manifestations, or phenotypes, of human genetic diseases. With ~12,000 terms, the HPO has become the standard ontology used in rare disease research and clinical genetics and has been adopted by the International Rare Diseases Research Consortium [8], ClinGen [9], and many other projects. Although both SNOMED-CT and the HPO provide a number of synonyms for each term, they usually miss many valid synonymous terms, as manually curating every term that refers to a concept is extremely difficult, if not impossible. For example, HPO provides four additional synonyms for the term “Renal neoplasm,” including “Kidney cancer” and “Renal tumors,” but it does not include synonyms such as “Renal cancer.” There are also many concepts in HPO, such as “Retinal neoplasm,” which are not given any synonyms in the ontology.

Many concept recognition and text annotation tools have been developed for biomedical text. Examples of popular tools for general purpose are the NCBO (National Center for Biomedical Ontology) annotator [10], OBO (Open Biological and Biomedical Ontologies) annotator [11], MetaMap [12], and Apache cTAKES (Clinical Text Analysis and Knowledge Extraction System) [13]. Other tools focusing on more specific domains have also been developed, such as BioLark [14] for automatic recognition of terms from the HPO and a tool by Lobo et al [15], which combines a machine learning approach with manual validation rules to detect HPO terms. Another example is the phenotype search tool provided by PhenoTips [16], which uses Apache Solr indexed on the HPO and has an extensive set of rule-based techniques to rank matching phenotypes for a query. Many of these systems consist of a pipeline of natural language processing components including a tokenizer, part-of-speech tagger, sentence boundary detector, and named entity recognizer (NER)/annotator. Generally, the NER/annotator component of these tools are based on text matching, dictionary look-ups, and rule-based methods, which usually require significant engineering effort and are often unable to handle novel synonyms that are not annotated in the ontology.

On the other hand, in the more general domain of natural language processing, many machine learning–based text classification and NER tools have been recently introduced [17-19]. Typically, these methods do not require manual rule-based engineering; however, they are dependent on large annotated text data for training. Popular among them is a model known as LSTM-CRF, in which long short-term memory (LSTM) [20], a variation of recurrent neural networks (RNNs) widely used for processing sequences such as text, is used to extract rich representations of the tokens in a sentence and is then followed by a conditional random field (CRF) [21] on top of these representations to recognize named entities.

Although these methods address a similar problem, they cannot be used directly for concept recognition, as the number of named entity classes is typically much lower than that of the concepts in medical ontologies. For instance, CoNLL-2003 [22], a data set widely used for evaluations of such methods, contains only four classes: locations, persons, organizations, and miscellaneous. As a result, these methods typically have a large number of training and test examples for each class, while in our setting, we are trying to recognize tens or hundreds of thousands of terms and may have only a few or even no examples of a specific term. Automatic creation of training data by exact match searching of the synonyms in a large corpus will not fully utilize synonyms that have no or low coverage in the data set, can bring bias by mislabeling valid out-of-ontology synonyms in the extracted snippets as negatives, and overfit to the context of the more frequent senses. Hence, in a setting where the training data does not fully cover all the classes, methods based on dictionary look-up might have some advantage, as they can identify a concept in a given text by simply matching it to a synonym available in their dictionary without requiring training data annotated with that concept.

In this paper, we develop a hybrid approach, called Neural Concept Recognizer (NCR), by introducing a neural dictionary model that learns to generalize to novel synonyms for concepts. Our model is trained on the information provided by the ontology, including the concept names, synonyms, and taxonomic relations between the concepts, and can be used to rank the concepts that a given phrase can match as a synonym. Our model consists of two main components: an encoder, which maps an input phrase to a vector representation, and an embedding table, which consists of the vector representations learned for the ontology concepts. The classification is performed based on the similarity between the phrase vector and the concept vectors. To allow for the use of our model to also detect concepts from longer texts, we scan the input text with fixed-size windows and report a phrase as matching a concept if it is above a threshold that is chosen from an appropriate validation data set.

Our work introduces a novel machine learning–based method for automatic concept recognition of medical terms in clinical text, and we have provided empirical results to demonstrate the accuracy of our methods in several settings. We trained our neural dictionary model on the HPO and used it to recognize concepts from 228 PubMed abstracts and 39 clinical reports of patients with rare genetic diseases. Additionally, we used a subset of concepts from SNOMED-CT that have matching terms

in ICD-9 and experimented on 2000 Intensive Care Unit (ICU) discharge summaries from a Multiparameter Intelligent Monitoring in Intensive Care (MIMIC-III) data set [23]. In both settings, we trained our model solely on the ontology data and did not use the text corpora except to set the recognition sensitivity threshold and choose model hyperparameters from a small validation set. Although the main focus of this work is recognizing HPO and SNOMED-CT concepts, our method can be easily trained on other biomedical ontologies. The results of our experiments show the high accuracy of our model, which is on par with or better than hand-trained concept recognition methods. Our tool has already been used in two applications. It has been integrated with the PhenoTips tool to suggest concepts for clinical reports [16] and to automatically recognize occurrences of phenotypes in a clinical report for subsequent data visualization [24].

Related Works

Recently, several machine learning methods have been used in biomedical NER or concept recognition. Habibi et al [25] trained the LSTM-CRF NER model, introduced by Lample et al [17], to recognize five entity classes of genes/proteins, chemicals, species, cell lines and diseases. They tested their model on several biomedical corpora and achieved better results than previous rule-based methods. In another work, Vani et al [26] introduced a novel RNN–based model and showed its efficiency on predicting ICD-9 codes in clinical notes. Both of these methods require a training corpus annotated with the concepts (loosely annotated in the case of Vani et al [26]).

Curating such an annotated corpus is more difficult for typical biomedical ontologies, as the corpus has to cover thousands of classes. For example, the HPO contains 11,442 concepts (classes), while, to the best of our knowledge, the only publicly available corpus hand annotated with HPO concepts [14] contains 228 PubMed abstracts with only 607 unique annotations that are not an exact match of a concept name or a synonym. Thus, training a method to recognize the presence of concepts in biomedical text requires a different approach when there is a large number of concepts.

The concepts in an ontology often have a hierarchical structure (ie, a taxonomy), which can be utilized in representation learning. Hierarchies have been utilized in several recent machine learning approaches. Deng et al [27] proposed a CRF-based method for image classification that takes into account inheritance and exclusion relations between the labels. Their CRF model transfers knowledge between classes by summing the weights along the hierarchy, leading to improved performance. Vendrov et al [28] introduced the order-embedding penalty to learn representations of hierarchical entities and used it for image caption retrieval tasks. Gaussian embeddings were introduced by Neelakantan et al [29] and learn a high-dimensional Gaussian distribution that can model entailment instead of single point vectors. Most recently, Nickel et al [30] showed that learning representations in a hyperbolic space can improve performance for hierarchical representations.

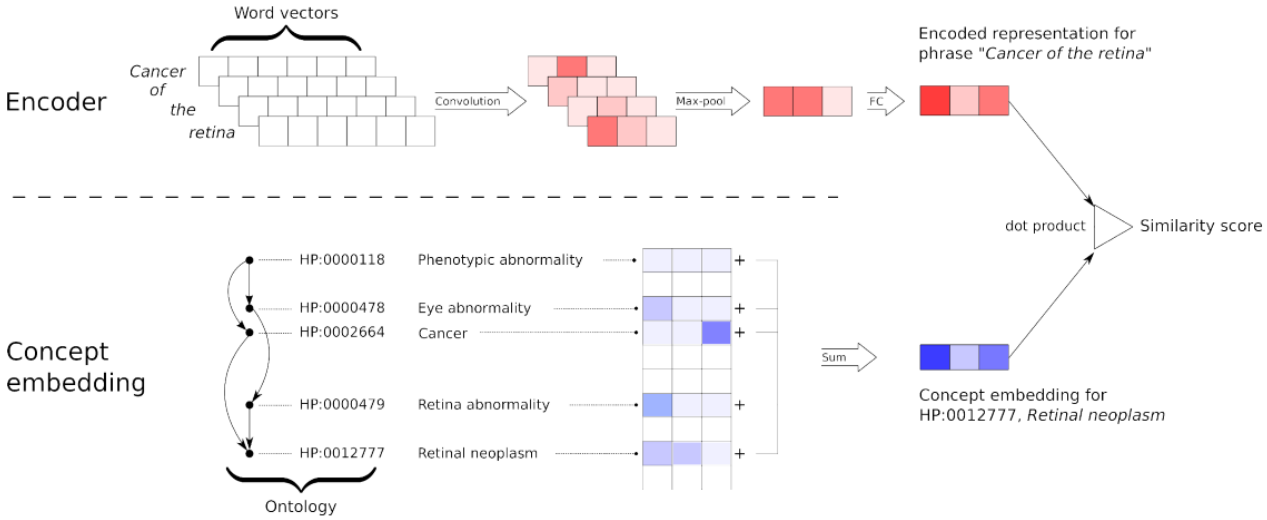
Methods

In this section, we first describe the neural dictionary model that computes the likelihood that a given phrase matches each concept from an ontology, and then demonstrate how to apply the model to larger text fragments such as a full sentence, which may have multiple (or no) terms.

Overview of the Neural Dictionary Model

The neural dictionary model receives a word or a phrase as input and finds the probability of the concepts in the ontology matching it. The model consists of a text encoder, which is a neural network that maps the query phrase into vector representation, and an embedding matrix with rows corresponding to the ontology concepts (Figure 1). We use the dot product of the query vector and a concept vector as the measure of similarity.

Figure 1. Architecture of the neural dictionary model. The encoder is shown at the top, and the procedure for computing the embedding for a concept is illustrated at the bottom. Encoder: a query phrase is first represented by its word vectors, which are then projected by a convolution layer into a new space. Then, a max-over-time pooling layer is used to aggregate the set of vectors into a single one. Thereafter, a fully connected layer maps this vector into the final representation of the phrase. Concept embedding: a matrix of raw embeddings is learned, where each row represents one concept. The final embedding of a concept is retrieved by summing the raw embeddings for that concept and all of its ancestors in the ontology. FC: fully connected.



Encoder

We use word embeddings to represent the input words learned in a pre-processing step by running fastText [31] on publicly available MEDLINE/PubMed abstracts. The goal of this unsupervised step is to map semantically similar words (eg, synonyms) to close vectors. We selected fastText for this task primarily because it takes into account the subword information, which is important in the medical domain where there are many semantically close words with slight morphologic variations.

Inspired by the work of Kim et al [32], our encoder projects these word vectors into another space using a convolution neural network. We have used a much simpler network, consisting of a single convolution layer, with a filter size of one word. Although our choice of filter size has the disadvantage of losing the word order information, in our settings, this was outweighed by the benefit of having fewer network parameters to learn. We also tried other types of encoders such as different variations of LSTMs and small variants of attention-based encoders [33]. However, given the small amount of training data available, simpler encoders were more effective.

After the first layer of projection, the output vectors were aggregated into a single vector (v) using a max-over-time pooling operation, as shown in the following equation $v=\max_t\{\mathrm{ELU}(Wx^{(t)}+b)\}$, where $x^{(t)}$ is the word vector for the tth word in the phrase; W and b are the weight matrix and the bias vector of the convolution filter, respectively; and ELU [34] is the activation function we used in the convolution layer. It should also be noted that the max operation used in the equation above is an element-wise operation that takes the maximum value of each feature across projected word vectors. Finally, a fully connected layer with the weights U was applied on v, followed by a ReLU (rectified linear unit) activation and $l2$ normalization. The result e was used as the encoded vector representation of the phrase:

Concept Representations

Our model includes a component that learns representations for concepts and measures the similarity between an input phrase and the concepts by computing the dot product between these representations and the encoded phrase e.

We denote these representations by the matrix H, where each row corresponds to one concept. Our model does not learn H directly, but instead learns a matrix  where each row  represents the features of concept c that are "novel" compared to its ancestors. Then, H can be derived by multiplying  by the taxonomy's ancestry matrix A: 

Each element of the ancestry matrix $A_{i,j}$ is nonzero only if concept j is an ancestor of i (including $i=j$) and is calculated as:

The final embedding of a concept would be the final embedding of its parent (or the average of its parents, in cases of multi-inheritance) plus its own raw embedding (ie, ). In other words, the parent concept provides the global location in the embedding space, whereas the child concepts learn their local locations with respect to that space.

This has two major advantages. First, it incorporates the taxonomic structure as implicit prior information on the geometry of the concept embeddings. Second, by binding the embeddings of the concepts, training becomes more efficient, as for each concept, it is sufficient to learn only the local location with respect to its parent, rather than learning the absolute location from scratch. Furthermore, when the location of a concept gets updated, both its descendants and ancestors will also get updated, even if they do not have samples present in

the mini-batch. More specifically, as a concept gets updated, the global locations provided to all its descendants are automatically updated as well, while the actual raw embedding of its ancestors will get updated through the backpropagation process. The results of our experiments quantitatively and qualitatively show the advantage of this approach in our task.

Finally, the classification is done by computing the dot product (plus a bias term) followed by a softmax layer as follows:

The taxonomy information can be ignored by setting A to the identity matrix I. In this scenario, the model would behave like an ordinary softmax classifier with the weight matrix

Training Procedure

Training is performed on the names and synonyms provided by the ontology. If a concept has multiple synonyms, each synonym-concept pair is considered as a separate training example. The parameters learned during the training are the encoder parameters W and U, and the concept representations through  The fastText word vectors used in our experiments had a dimensionality of 100, while we set the dimensionality of the concept embeddings to be 1024. We used a filter size of 1024 for the convolution layer in the encoder, and the output of the dense layer used after the max-pooling layer was 1024. We trained our model by minimizing the cross-entropy loss between the softmax output and the class labels using Adam optimizer [35], with a learning rate of 0.002 and a batch size of 256. We trained our model for 100 epochs.

Concept Recognition in a Sentence

To use our neural dictionary model to recognize concepts in a sentence or larger text, we extract all n-grams of one to seven words in the text and used the neural dictionary model to match each n-gram to a concept. We filter irrelevant n-grams by removing the candidates whose matching score (the softmax probability provided by the neural dictionary model) is lower than a threshold. This threshold is chosen based on the performance of the method (f-measure) on a validation set.

We also use random n-grams from an unrelated corpus (in our case Wikipedia) as negative examples labeled with a dummy *none* concept when training the neural dictionary model. This is done to reduce false positives that do not match to any concept (as opposed to false positives that are due to misclassification between two different concepts). To reduce the compute time, we made the assumption that phenotypic phrases have a length of at most 10 tokens, which we chose based on the empirical evidence that less than 0.8% of the names/synonyms in the HPO are longer than 10 tokens. As a result, the lengths of these n-grams were uniformly selected to be between 1 and 10.

After all the n-grams satisfying the conditions are captured, a postprocessing step is performed to ensure that the results are consistent. For every pair of overlapping captured n-grams, if both n-grams match the same concept, we retain the smaller n-gram. Otherwise, if they are matched to different concepts, we choose the longer n-gram, as this reduces the chances of choosing shorter general concepts in the presence of a more specific, longer, concept. For example, when annotating the sentence "The patient was diagnosed with conotruncal heart defect," our method will favor choosing the longer, more specific concept "conotruncal heart defect" rather than the more general concept "heart defect."

Results

Overview

To evaluate our model, we trained the model on the HPO and SNOMED-CT and applied it to a number of medical texts. We evaluated the model on two different tasks. In the first task, the model ranks concepts matching an input isolated phrase (synonym classification) and in the second task, concepts are recognized and classified from a document (concept recognition).

To assess the effectiveness of the techniques used in our model, we trained four variations of the model as follows:

- NCR: The full model, with the same architecture as described in the section Overview of the Neural Dictionary Model. The training data for this model includes negative examples.
- NCR-H: In this version, the model ignores the taxonomic relations by setting the ancestry matrix A to the identity matrix I.
- NCR-N: Similar to the original NCR, this version utilizes the taxonomic relations. However, this model has not been trained on negative samples.
- NCR-HN: A variation that ignores the taxonomy and has not been trained on negative examples.

To improve stability, we trained 10 different versions of our model, varying the random initialization of the model parameters and randomly reshuffling the training data across minibatches at the beginning of each training epoch. We created an ensemble of these 10 models by averaging their prediction probabilities for any given query and used this ensemble in all experiments.

Data Sets

In most of our experiments, we used the HPO to train the neural dictionary model. To maintain consistency with previous work, we used the 2016 release of the HPO, which contains a total of 11,442 clinical phenotypic abnormalities seen in human disease and provides a total of 19,202 names and synonyms for them, yielding an average of 1.67 names per concept.

We evaluated the accuracy of our model trained on the HPO on two different data sets:

- PubMed: This data set contains 228 PubMed article abstracts, gathered and manually annotated with HPO concepts by Groza et al [14].
- Undiagnosed Diseases Program (UDP): This data set includes 39 clinical reports provided by National Health Institutes UDP [36]. Each case contains the medical history of a patient in unstructured text format and a list of phenotypic findings, recorded as a set of HPO concepts, gathered by the examining clinician from the patient encounter.

In order to examine the effectiveness of our model on different ontologies, we also trained the model on a subset of SNOMED-CT, which is a comprehensive collection of medical concepts that includes their synonyms and taxonomy. We evaluated the trained model for concept recognition using a subset of 2000 ICU discharge summaries from MIMIC-III. The discharge summaries are composed of unstructured text and are accompanied by a list of disease diagnosis terms in the form of ICD-9 codes.

Since SNOMED-CT provides a more sophisticated hierarchy than ICD-9 and a mapping between the two exists, we used a subset of SNOMED-CT concepts that include the ICD-9 concepts. We considered the 1292 most frequent ICD-9 concepts that have a minimum of 50 occurrences in MIMIC-III. These were filtered to 1134 concepts that also have at least one mapping SNOMED-CT concept, which were mapped to a total of 8405 SNOMED-CT concepts (more SNOMED-CT concepts because of one-to-many mappings). To have a single connected hierarchy of concepts, we also added all missing ancestors of these SNOMED-CT terms, resulting in a total of 11,551 SNOMED-CT concepts. To these additional 3146 SNOMED-CT concepts, we assigned the ICD-9 code mapped to the original SNOMED-CT term that had induced them (ie, their descendent). We trained NCR using these 11,551 SNOMED-CT concepts and the 21,550 names and synonyms associated with them.

Synonym Classification Results

In this experiment, we evaluated our method's performance in matching isolated phrases with ontology concepts. For this purpose, we extracted 607 unique phenotypic phrases that did not have an exact match among the names and synonyms in the HPO from the 228 annotated PubMed abstracts. We used our model to classify HPO concepts for these phrases and ranked them by their score.

In addition to the four variations of our model, we compared our method with one based on Apache Solr, customized to suggest HPO terms for phenotypic queries. This tool is currently in use as a component of the phenotyping software PhenoTips [16]. The results of this experiment are provided in Table 1. Since all the phrases in this data set are true phenotypic terms and PhenoTips reports at most 10 concepts for each phrase, we measured the fraction of the predictions where the correct label was among the top 1 (R@1) and top 5 (R@5) recalled concepts, instead of precision/recall. NCR outperformed PhenoTips by 20%-30% in this experiment. While NCR-N slightly outperformed regular NCR based on R@1, the experiments here contained no queries without phenotypic terms, which is the task that NCR-N was built to model.

An example phrase from this data set is "reduced retinal pigment," labeled as HP:0007894. In our version of the HPO, there are four names/synonyms for this phrase: "hypopigmentation of the fundus," "decreased retinal pigmentation," "retinal depigmentation," and "retinal hypopigmentation." NCR correctly identified this concept as its top match. In contrast, the correct concept was not in the top 10 concepts reported by PhenoTips; the top reported concept was "retinal pigment epithelial mottling."

Table 1. Synonym classification experiments on 607 phenotypic phrases extracted from 228 PubMed abstracts. Largest values for each category are italicized.

Method	Accuracy (%)	
	R@1[a]	R@5[b]
PhenoTips	28.9	49.3
NCR[c]	51.6	*80.6*
NCR-H[d]	45.5	69.8
NCR-N[e]	*55.8*	78.2
NCR-HN[f]	50.2	71.8

[a]R@1: recall using top 1 result from each method.

[b]R@5: recall using top 5 results from each method.

[c]NCR: Neural Concept Recognizer.

[d]NCR-H: variation of the NCR model that ignores taxonomic relations.

[e]NCR-N: variation of the NCR model that has not been trained on negative samples.

[f]NCR-HN: variation of the NCR model that ignores the taxonomy and has not been trained on negative examples.

Concept Recognition Results

We evaluated the four versions of NCR for concept recognition and compared them with four rule-based methods: NCBO annotator [10], cTAKES [13], BioLarK [14], and OBO annotator [11]. The NCBO annotator is a general concept recognition tool with access to hundreds of biomedical ontologies, including the HPO. cTAKES is a more general medical knowledge extraction system primarily designed for SNOMED-CT, while BioLarK and the OBO annotator are concept recognizers primarily tailored for the HPO. Another method, called IHP (Identifying Human Phenotypes) [15], was recently introduced for identifying HPO terms in unstructured text using machine learning for named entity recognition and a rule-based approach for further extending them. However, this method is not directly comparable, as it only reports the text spans that are a phenotype and does not classify or rank matching HPO terms.

In order to choose a score threshold for filtering irrelevant concepts, we used 40 random PubMed abstracts as a validation set and compared the micro F1-score with different threshold values. The selected thresholds were 0.85, 0.8, 0.8, and 0.75 for NCR, NCR-H, NCR-N, and NCR-HN, respectively. Since the UDP data set contained fewer reports (39 in total), we did not choose a separate UDP validation set and used the same threshold determined for the PubMed abstracts. We tested our methods on the remaining 188 PubMed abstracts and the 39 UDP reports and calculated micro and macro versions of precision, recall, and F1-score, as shown in the following equations:

In these equations, D is the set of all documents and R_d and L_d notate the set of reported concepts and label concepts for the document d, respectively. In cases where $|L_d|$ or $|R_d|$ were zero, we assigned a macro recall and macro precision of 1.0, respectively.

We also calculated a less strict version of accuracy measurements that takes the taxonomic relations of the concepts into consideration. For this, we extended the reported set and the label set for each document to include all their ancestor concepts, which we notate by $E(L_d)$ and $E(R_d)$, respectively, and calculated an extended version of the precision and recall, as well as the Jaccard Index of the extended sets. The following equations show how these accuracies are derived:

The measured micro and macro accuracies are provided in Tables 2 and 3 for the PubMed abstract and UDP data sets, respectively. The taxonomy-based extended accuracies and the Jaccard index results are available in Tables 4 and 5 for the abstracts and UDP data sets, respectively. In both experiments, based on the measurements of the Jaccard index and all three versions of micro, macro, and extended F1-scores, NCR had higher accuracy than all other baselines. Furthermore, by comparing the NCR and NCR-H, we observed that using the hierarchy information considerably improved the F1-score of the model in the abstract data set, although the F1-score of the UDP set was slightly lower. Finally, comparison of NCR and NCR-N showed that using negative examples during the training improved the overall accuracy for the abstract data set, while not using the negatives led to a narrow advantage with the UDP data set.

To verify the statistical significance of NCR's superiority to the baselines, we aggregated both the abstract and UDP data sets for a total of 227 documents and calculated the F1-score for each document separately. This method is different from that used to calculate the F1-score presented in Tables 2-5, which only show a single measurement of F1-score per category. We compared the main version of NCR against BioLarK, which was our strongest baseline. NCR performed statistically significantly better (P=.003, Wilcoxon test).

To evaluate the effectiveness of the techniques employed in NCR on a different ontology, we trained the four variations of our model on the SNOMED-CT subset, using 200 MIMIC reports as the validation set and the remaining 1800 reports as a test set. We mapped each reported SNOMED-CT concept to the corresponding ICD-9 code and calculated the accuracy measurements (Table 6).

The results show that using the hierarchy information improved both micro and macro F1-scores. Since the labels were only available as ICD-9 codes, which do not hold a sufficiently rich hierarchical structure as opposed to HPO and SNOMED-CT, the Jaccard index and the extended accuracy measurements were less meaningful and were not calculated. We also ran the original cTAKES, which is optimized for SNOMED-CT concepts, on the 1800 test documents and filtered its reported SNOMED-CT results to ones that have a corresponding ICD-9. Although cTAKES had a high recall, the overall F1-scores were lower than those for NCR. Furthermore, using a method similar to the one used to calculate the statistical significance for the improvement relative to BioLark in the section above, we compared NCR with cTAKES and found that NCR performed statistically significantly better (P<.001, Wilcoxon test).

Table 2. Micro and macro measurements for concept recognition experiments on 188 PubMed abstracts. Neural Concept Recognizer models were trained on Human Phenotype Ontology. Largest values for each category are italicized.

Method	Micro (%)			Macro (%)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
BioLarK	78.5	60.5	68.3	76.6	66.0	70.9
cTAKES[a]	72.2	55.6	62.8	74.0	61.4	67.1
OBO[b]	78.3	53.7	63.7	79.5	58.6	67.5
NCBO[c]	*81.6*	44.0	57.2	79.5	48.7	60.4
NCR[d]	80.3	62.4	*70.2*	*80.5*	68.2	*73.9*
NCR-H[e]	74.4	61.5	67.3	72.2	67.1	69.6
NCR-N[f]	78.1	*62.5*	69.4	76.6	*68.3*	72.2
NCR-HN[g]	77.1	57.2	65.7	76.5	63.4	69.3

[a]cTAKES: Clinical Text Analysis and Knowledge Extraction System.

[b]OBO: Open Biological and Biomedical Ontologies

[c]NCBO: National Center for Biomedical Ontology.

[d]NCR: Neural Concept Recognizer.

[e]NCR-H: variation of the NCR model that ignores taxonomic relations.

[f]NCR-N: variation of the NCR model that has not been trained on negative samples.

[g]NCR-HN: variation of the NCR model that ignores the taxonomy and has not been trained on negative examples.

Table 3. Micro and macro measurements for concept recognition experiments on 39 Undiagnosed Diseases Program clinical notes. Neural Concept Recognizer models were trained on Human Phenotype Ontology. Largest values for each category are italicized.

Method	Micro (%)			Macro (%)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
BioLarK	27.6	21.0	23.9	28.7	21.6	24.6
cTAKES[a]	31.5	18.9	23.6	*37.5*	20.2	26.2
OBO[b]	26.8	20.5	23.2	28.8	20.1	23.7
NCBO[c]	*33.4*	16.9	22.5	37.1	19.9	25.9
NCR[d]	24.5	27.2	25.8	26.5	27.6	27.0
NCR-H[e]	25.1	26.8	25.9	26.2	27.0	26.6
NCR-N[f]	24.3	*28.5*	26.2	27.0	*28.9*	*27.9*
NCR-HN[g]	25.5	27.2	*26.4*	27.4	27.7	27.6

[a]cTAKES: Clinical Text Analysis and Knowledge Extraction System.

[b]OBO: Open Biological and Biomedical Ontologies

[c]NCBO: National Center for Biomedical Ontology.

[d]NCR: Neural Concept Recognizer.

[e]NCR-H: variation of the NCR model that ignores taxonomic relations.

[f]NCR-N: variation of the NCR model that has not been trained on negative samples.

[g]NCR-HN: variation of the NCR model that ignores the taxonomy and has not been trained on negative examples.

Table 4. Extended measurements for concept recognition experiments on 188 PubMed abstracts. Neural Concept Recognizer models were trained on Human Phenotype Ontology. Largest values for each category are italicized.

Method	Extended value (%)			Jaccard value (%)
	Precision	Recall	F1-score	
BioLarK	91.5	80.8	85.8	76.9
cTAKES[a]	95.6	73.9	83.3	72.1
OBO[b]	92.4	77.9	84.5	74.4
NCBO[c]	*95.8*	65.4	77.7	64.3
NCR[d]	93.3	82.1	*87.3*	*79.1*
NCR-H[e]	86.5	*83.8*	85.1	76.7
NCR-N[f]	90.6	83.1	86.7	78.2
NCR-HN[g]	89.7	78.9	83.9	73.2

[a]cTAKES: Clinical Text Analysis and Knowledge Extraction System.

[b]OBO: Open Biological and Biomedical Ontologies

[c]NCBO: National Center for Biomedical Ontology.

[d]NCR: Neural Concept Recognizer.

[e]NCR-H: variation of the NCR model that ignores taxonomic relations.

[f]NCR-N: variation of the NCR model that has not been trained on negative samples.

[g]NCR-HN: variation of the NCR model that ignores the taxonomy and has not been trained on negative examples.

Table 5. Extended measurements for concept recognition experiments on 39 Undiagnosed Diseases Program clinical notes. Neural Concept Recognizer models were trained on Human Phenotype Ontology. Largest values for each category are italicized.

Method	Extended value (%)			Jaccard index (%)
	Precision	Recall	F1-score	
BioLarK	58.9	42.6	49.5	29.5
cTAKES[a]	68.5	36.7	47.8	27.3
OBO[b]	59.2	46.4	52.0	31.3
NCBO[c]	*69.8*	37.2	48.5	27.2
NCR[d]	57.1	49.4	*53.0*	*31.5*
NCR-H[e]	54.0	49.4	51.6	30.5
NCR-N[f]	54.7	*50.5*	52.5	31.4
NCR-HN[g]	56.5	49.0	52.5	31.3

[a]cTAKES: Clinical Text Analysis and Knowledge Extraction System.

[b]OBO: Open Biological and Biomedical Ontologies

[c]NCBO: National Center for Biomedical Ontology.

[d]NCR: Neural Concept Recognizer.

[e]NCR-H: variation of the NCR model that ignores taxonomic relations.

[f]NCR-N: variation of the NCR model that has not been trained on negative samples.

[g]NCR-HN: variation of the NCR model that ignores the taxonomy and has not been trained on negative examples.

Table 6. Results for concept recognition experiments on 1800 Multiparameter Intelligent Monitoring in Intensive Care documents. The Neural Concept Recognizer models were trained on a subset of the Systematized Nomenclature of Medicine - Clinical Terms ontology. Largest values for each category are italicized.

Method	Micro (%)			Macro (%)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
cTAKES[a]	9.1	*37.0*	14.6	8.7	*36.5*	14.1
NCR[b]	10.9	26.7	*15.5*	10.6	26.9	15.2
NCR-H[c]	10.0	30.6	15.1	9.6	30.4	14.6
NCR-N[d]	*11.2*	24.8	15.4	*11.1*	25.3	*15.4*
NCR-HN[e]	9.6	28.6	14.4	9.2	28.9	13.9

[a]cTAKES: Clinical Text Analysis and Knowledge Extraction System.

[b]NCR: Neural Concept Recognizer.

[c]NCR-H: variation of the NCR model that ignores taxonomic relations.

[d]NCR-N: variation of the NCR model that has not been trained on negative samples.

[e]NCR-HN: variation of the NCR model that ignores the taxonomy and has not been trained on negative examples.

Qualitative Results

To better understand how utilizing the hierarchy information affects our model, we used t-SNE (t-distributed stochastic neighbor embedding) to embed and visualize the learned concept representations for the rows of matrix *H* for NCR-N (using hierarchy) and NCR-NH (not using the hierarchy), trained on the HPO. These representations are illustrated in Figure 2, where colors are assigned to concepts based on their high-level ancestor (the 23 children of the root). If a concept had multiple high-level ancestors, we chose one randomly. As is evident in the plots, the representations learned for NCR-N were better clustered than those for NCR-NH.

Interestingly, in the representations learned for NCR-N, concepts in categories that share children with many other categories, such as “Neoplasm” (dark grey), are located in the center of the plot, close to various other categories, while a category like “Abnormality of ear” (orange) forms its own cluster far from center and is separated from other categories.

To further investigate the false positives reported by NCR, we manually investigated the false positives reported by our method in three clinical reports randomly chosen from the UDP data set. We looked at false positives from the extended version of evaluations, which included concepts reported by our method, where neither the concepts nor any of their descendants were in the label set. This yielded a total number of 73 unique false positives for the three documents. Based on a manual analysis of these terms conducted by a medical expert on rare genetic diseases (coauthor DA), 47.9% of the reported false positives were actually correctly adding more information to the closest phenotype reported in the label set. One such example is “Congenital hypothyroidism on newborn screening.” Although our method correctly recognized “Congenital hypothyroidism,” the closest concept in the extended label set was “Abnormality of the endocrine system.” In an additional 8.2% of cases, our model correctly reported a more specific concept than that presented in the patient record, but the concept was sufficiently close to a specified phenotype for it not to be considered a novel finding. Furthermore, 16.4% of the reported false positives were, in fact, mentioned in the text, albeit as negations, such as “Group fiber atrophy was not seen.” In 6.8% of these cases, the reported phenotype was mentioned but not confidently diagnosed, such as “possible esophagitis and gastric outlet delay.”

Figure 2. Visualization of the representations learned for Human Phenotype Ontology concepts. The representations are embedded into two dimensions using t-SNE. The colors denote the high-level ancestors of the concepts. The plot on the left shows the representations learned in NCR-N, where the taxonomy information was used in training, and the plot on the right shows representations learned for NCR-HN, where the taxonomy was ignored. NCR-HN: variation of the NCR model that ignores the taxonomy and has not been trained on negative examples; NCR-N: variation of the NCR model that has not been trained on negative samples; t-SNE: t-distributed stochastic neighbor embedding.

Discussion

Principal Findings

Our experiments showed the high accuracy of NCR compared to the baselines in both synonym classification and concept recognition, where NCR consistently achieved higher F1-scores across different data sets. Furthermore, we showed that NCR's use of the hierarchical information contributes to its higher performance.

In the synonym classification task, as evident in Table 1, all variations of NCR had a much better performance than the tool provided by PhenoTips. Furthermore, comparison of NCR and NCR-H showed that use of the hierarchy information considerably improved accuracy.

In concept recognition experiments, NCR had a better F1-score and Jaccard index than BioLarK and cTAKES on PubMed abstracts (Tables 2 and 4) and UDP reports (Tables 3 and 5). On both data sets, NCR had a higher recall, showing its ability to better generalize to synonymous terms that occurred in the text. In some experiments, NCBO achieved the highest precision; however, we should note that in the same experiments, NCR achieved a much better recall rate, and when taking both precision and recall into account, NCR had the highest F1-score.

Among different variations of NCR, use of the hierarchy information always led to a higher F1-score and Jaccard index. Having negative samples during training also generally improved accuracy; however, in some cases, this difference was small, and in some cases, NCR-N showed slightly better results.

Although the PubMed abstracts were manually annotated with HPO concepts by Groza et al [14], the text provided for UDP is not annotated and there is no explicit association between the provided HPO terms and phenotypic phrases in the text. However, since both the text and the terms referred to the same patients, a correspondence exists between them. This can explain the overall higher accuracy of all methods on PubMed data compared to UDP data. As a result, these performance measurements would be more meaningful when observed in a relative manner, which shows the better performance of NCR than the baselines.

The experiments on MIMIC data, where the model was trained on SNOMED-CT, resulted in a much lower accuracy than the two experiments performed using the HPO. In addition to the problem of implicit correspondence between labels and actual mentions in the text, in this experiment, we used a mapping between ICD-9 and SNOMED-CT terms, which can introduce further inconsistencies. On the other hand, for the sake of evaluating the techniques employed in our model on another ontology, use of the SNOMED-CT hierarchy, similar to the case with the HPO, improves the F1-scores (Table 3).

In addition to the quantitative results showing the advantage of using the hierarchy information, our visualization of the concept representations in Figure 2 shows that the representations learned for NCR-N are more cohesive compared to those for NCR-HN. Although in theory, NCR-N has the flexibility to learn representations identical to those of NCR-HN, the way our model utilizes the taxonomy connects the embedding of related concepts during training, which leads to better separated clusters.

NCR has already been used in several applications in practice. Currently, a version of NCR trained on the HPO is deployed as a component of PhenoTips software [16] and is being used in both annotation of clinical notes and term suggestion for manually entered phenotypes. Another example is PhenoLines [24], a software for visualizing disease subtypes, that relies on a mapping between HPO and Unified Medical Language System (UMLS) [37] terms. NCR was effectively used to help improve the coverage of their mapping. The code for NCR is available under the MIT license [38].

Conclusions

In this paper, we presented a neural dictionary model that ranks matching concepts for a query phrase and can be used for concept recognition in larger text. Unlike other machine learning–based concept recognition tools, our training is solely performed on the ontology data (except the unsupervised learning of the word vectors) and does not require any annotated corpus. Another novelty of our model is our approach to using the taxonomic relations between concepts that, based on our experiments, improve synonym classification. Use of these taxonomic relations makes the training of our model easier by sharing knowledge between different concepts and providing implicit prior information on the similarity between concepts for the model. Furthermore, using multiple sources of information can improve the robustness of the model to potential errors in the input ontologies (eg, due to a mislabeled synonym).

NCR uses convolutional neural networks to encode query phrases into vector representations and computes their similarity to embeddings learned for ontology concepts. The model benefits from knowledge transfer between child and parent concepts by summing the raw embeddings of a concept's ancestors to compute its final embedding. We tested our neural dictionary model by classifying 607 phenotypic phrases, and our model achieved a considerably higher accuracy than another method designed for this task and baseline versions of our model that do not use the taxonomy information. We also tested our method for concept recognition on full text using four data sets. In one setting, we trained our model on the HPO and tested it on two data sets, including 188 PubMed paper abstracts and 39 UDP clinical records, while in another setting, we trained the model on a subset of SNOMED-CT medical concepts and tested it on 1800 MIMIC ICU discharge notes. Our results showed the efficiency of our methods in both settings.

One major challenge for the concept recognition task is to filter candidates that do not match any class in the ontology. In our experiments, we approached this challenge by adding negative samples from Wikipedia in the training. Although this improved the results, it did not fully solve the problem, as there can be many relevant medical terms in a clinical text that are neither in an ontology nor available in any negative examples.

Although our experiments have shown the high accuracy of our model in classifying synonyms, we believe there is much more room for improvement in the overall concept recognition method, especially the way that n-grams are selected and filtered. Limitations of NCR include its relatively slower speed than several dictionary-based and rule-based methods and its limited ability to utilize contextual information for concept recognition. An interesting direction for future work is to investigate the possibility of using unsupervised methods for encoding phrases, such as skip-thought vectors [39] or the recently introduced language representation model BERT (Bidirectional Encoder Representations from Transformers) [40], to use the massive amount of available unannotated biomedical corpora for better generalization of classifying synonymous phrases and concept recognition.

Acknowledgments

We thank Michael Glueck for his valuable comments and discussions, and Mia Husic for help in improving the manuscript. We also thank Tudor Groza for his helpful comments and for providing us the BioLarK API used for the experiments. This work was partially supported by an NSERC Discovery grant to MB.

Conflicts of Interest

None declared.

References

1. Simmons M, Singhal A, Lu Z. Text Mining for Precision Medicine: Bringing Structure to EHRs and Biomedical Literature to Understand Genes and Health. Adv Exp Med Biol 2016;939:139-166 [FREE Full text] [doi: 10.1007/978-981-10-1503-8_7] [Medline: 27807747]
2. Jonnagaddala H. Mining Electronic Health Records to Guide and Support Clinical Decision Support Systems. In: Healthcare Ethics and Training: Concepts, Methodologies, Tools, and Applications. PA, USA: IGI Global; 2017:184-201.
3. Luo Y, Thompson WK, Herr TM, Zeng Z, Berendsen MA, Jonnalagadda SR, et al. Natural Language Processing for EHR-Based Pharmacovigilance: A Structured Review. Drug Saf 2017 Dec;40(11):1075-1089. [doi: 10.1007/s40264-017-0558-6] [Medline: 28643174]
4. Gonzalez GH, Tahsin T, Goodale BC, Greene AC, Greene CS. Recent Advances and Emerging Applications in Text and Data Mining for Biomedical Discovery. Brief Bioinform 2016 Jan;17(1):33-42 [FREE Full text] [doi: 10.1093/bib/bbv087] [Medline: 26420781]
5. Piñero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M, et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford) 2015;2015:bav028 [FREE Full text] [doi: 10.1093/database/bav028] [Medline: 25877637]
6. SNOMED International. URL: http://www.snomed.org/snomed-ct/five-step-briefing [accessed 2019-04-18] [WebCite Cache ID 77jn1Od9U]
7. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res 2017 Dec 04;45(D1):D865-D876 [FREE Full text] [doi: 10.1093/nar/gkw1039] [Medline: 27899602]

8. Lochmüller H, Le Cam Y, Jonker AH, Lau LPI, Baynam G, Kaufmann P, et al. 'IRDiRC Recognized Resources': a new mechanism to support scientists to conduct efficient, high-quality research for rare diseases. Eur J Hum Genet 2017 Dec;25(2):162-165 [FREE Full text] [doi: 10.1038/ejhg.2016.137] [Medline: 27782107]
9. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, ClinGen. ClinGen--the Clinical Genome Resource. N Engl J Med 2015 Dec 04;372(23):2235-2242 [FREE Full text] [doi: 10.1056/NEJMsr1406261] [Medline: 26014595]
10. Jonquet C, Shah NH, Musen MA. The open biomedical annotator. Summit Transl Bioinform 2009 Mar 01;2009:56-60 [FREE Full text] [Medline: 21347171]
11. Taboada M, Rodríguez H, Martínez D, Pardo M, Sobrido MJ. Automated semantic annotation of rare disease cases: a case study. Database (Oxford) 2014;2014 [FREE Full text] [doi: 10.1093/database/bau045] [Medline: 24903515]
12. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21 [FREE Full text] [Medline: 11825149]
13. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
14. Groza T, Köhler S, Doelken S, Collier N, Oellrich A, Smedley D, et al. Automatic concept recognition using the human phenotype ontology reference and test suite corpora. Database (Oxford) 2015;2015 [FREE Full text] [doi: 10.1093/database/bav005] [Medline: 25725061]
15. Lobo M, Lamurias A, Couto FM. Identifying Human Phenotype Terms by Combining Machine Learning and Validation Rules. Biomed Res Int 2017;2017:8565739 [FREE Full text] [doi: 10.1155/2017/8565739] [Medline: 29250549]
16. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al. PhenoTips: patient phenotyping software for clinical and research use. Hum Mutat 2013 Aug;34(8):1057-1065. [doi: 10.1002/humu.22347] [Medline: 23636887]
17. Lample G. arXiv. 2016. Neural Architectures for Named Entity Recognition URL: https://arxiv.org/abs/1603.01360 [accessed 2019-04-24] [WebCite Cache ID 77sNQlmqS]
18. Huang Z, Xu W, Yu K. arXiv. 2015. Bidirectional LSTM-CRF Models for Sequence Tagging URL: https://arxiv.org/abs/1508.01991 [accessed 2019-04-24] [WebCite Cache ID 77sNfWP89]
19. Ma X, Hovy E. arXiv. 2016. End-to-End Sequence Labeling via Bi-Directional Lstm-Cnns-Crf URL: https://arxiv.org/abs/1603.01354 [accessed 2019-04-24] [WebCite Cache ID 77sNvEEZi]
20. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation 1997 Nov;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735]
21. Lafferty J, McCallum A, Pereira FCN. University of Pennsylvania ScholarlyCommons. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data URL: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis_papers [accessed 2019-04-24]
22. Tjong Kim Sang EF, De Meulder F. Introduction to the CoNLL-Language-Independent Named Entity Recognition. 2003 Presented at: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL; 2003; Edmonton, Canada p. 142-147 URL: https://www.aclweb.org/anthology/W03-0419
23. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
24. Glueck M, Naeini MP, Doshi-Velez F, Chevalier F, Khan A, Wigdor D, et al. PhenoLines: Phenotype Comparison Visualizations for Disease Subtyping via Topic Models. IEEE Trans Vis Comput Graph 2018 Dec;24(1):371-381. [doi: 10.1109/TVCG.2017.2745118] [Medline: 28866570]
25. Habibi M, Weber L, Neves M, Wiegandt DL, Leser U. Deep learning with word embeddings improves biomedical named entity recognition. Bioinformatics 2017 Jul 15;33(14):i37-i48 [FREE Full text] [doi: 10.1093/bioinformatics/btx228] [Medline: 28881963]
26. Vani A, Jernite Y, Jernite D. arXiv. 2017. Grounded Recurrent Neural Networks URL: https://arxiv.org/abs/1705.08557 [accessed 2019-04-24] [WebCite Cache ID 77sOcriI7]
27. Deng J, Ding N, Jia Y, Frome A, Murphy K, Bengio S, et al. Large-Scale Object Classification Using Label Relation Graphs. In: Computer Vision – ECCV 2014. NY, USA: Springer; 2014.
28. Vendrov I, Kiros R, Fidler S, Urtasun R. arXiv. 2015. Order-Embeddings of Images and Language URL: https://arxiv.org/abs/1511.06361 [accessed 2019-04-24] [WebCite Cache ID 77sPDXI71]
29. Neelakantan A, Roth B, McCallum A. arXiv. 2015. Compositional Vector Space Models for Knowledge Base Inference URL: https://arxiv.org/abs/1504.06662 [accessed 2019-04-24] [WebCite Cache ID 77sPQJhCH]
30. Nickel M, Kiela D. arXiv. 2017. Poincaré Embeddings for Learning Hierarchical Representations URL: https://arxiv.org/abs/1705.08039 [accessed 2019-04-24] [WebCite Cache ID 77sPb3fYf]
31. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics 2017 Dec;5:135-146. [doi: 10.1162/tacl_a_00051]
32. Kim Y. arXiv. 2014. Convolutional Neural Networks for Sentence Classification URL: https://arxiv.org/abs/1408.5882 [accessed 2019-04-24] [WebCite Cache ID 77sPyeki6]
33. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. arXiv. 2017. Attention Is All You Need URL: https://arxiv.org/abs/1706.03762 [accessed 2019-04-24] [WebCite Cache ID 77sQBej4D]

34. Clevert AD, Unterthiner T, Hochreiter S. arXiv. 2015. Fast and Accurate Deep Network Learning by Exponential Linear Units (Elus) URL: https://arxiv.org/abs/1511.07289 [accessed 2019-04-24] [WebCite Cache ID 77sQPixhj]
35. Kingma DP, Ba J. arXiv. 2014. A Method for Stochastic Optimization URL: https://arxiv.org/abs/1412.6980 [accessed 2019-04-24] [WebCite Cache ID 77sQZGHQF]
36. Tifft CJ, Adams DR. The National Institutes of Health undiagnosed diseases program. Curr Opin Pediatr 2014 Dec;26(6):626-633 [FREE Full text] [doi: 10.1097/MOP.0000000000000155] [Medline: 25313974]
37. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
38. GitHub. Neural CR URL: https://github.com/ccmbioinfo/NeuralCR [accessed 2019-04-24] [WebCite Cache ID 77sLcntjU]
39. Kiros R, Zhu Y, Salakhutdinov R, Zemel RS, Torralba A, Urtasun R, et al. Skip-thought vectors. In: Advances in neural information processing systems. 2015 Presented at: Proceedings of the 28th International Conference on Neural Information Processing Systems Skip-Thought Vectors; December 07-12, 2015; Montreal, Canada p. 3302.
40. Devlin J, Chang MW, Lee K, Toutanova K. arXiv. 2018. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding URL: https://arxiv.org/abs/1810.04805 [accessed 2019-04-24] [WebCite Cache ID 77sRFfJCW]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
CRF: conditional random field
cTAKES: Clinical Text Analysis and Knowledge Extraction System
EHR: electronic health records
HPO: Human Phenotype Ontology
ICD-9: International Classification of Diseases - Ninth Revision
ICU: Intensive Care Unit
IHP: Identifying Human Phenotypes
LSTM: long short-term memory
MIMIC: Multiparameter Intelligent Monitoring in Intensive Care
NER: named entity recognizer
NCBO: National Center for Biomedical Ontology
NCR: Neural Concept Recognizer
NCR-H: variation of the NCR model that ignores taxonomic relations
NCR-HN: variation of the NCR model that ignores the taxonomy and has not been trained on negative examples
NCR-N: variation of the NCR model that has not been trained on negative samples
OBO: Open Biological and Biomedical Ontologies
R@1: recall using top 1 results from each method
R@5: recall using top 5 results from each method
ReLU: rectified linear unit
SNOMED-CT: Systematized Nomenclature of Medicine - Clinical Terms
t-SNE: t-distributed stochastic neighbor embedding
UDP: Undiagnosed Diseases Program

Edited by G Eysenbach; submitted 23.10.18; peer-reviewed by M Jiang, N Miyoshi, F Li, V Rajan; comments to author 26.12.18; revised version received 20.02.19; accepted 29.03.19; published 10.05.19.

Please cite as:
Arbabi A, Adams DR, Fidler S, Brudno M
Identifying Clinical Terms in Medical Text Using Ontology-Guided Machine Learning
JMIR Med Inform 2019;7(2):e12596
URL: http://medinform.jmir.org/2019/2/e12596/
doi:10.2196/12596
PMID:31094361

Original Paper

Fast Healthcare Interoperability Resources, Clinical Quality Language, and Systematized Nomenclature of Medicine—Clinical Terms in Representing Clinical Evidence Logic Statements for the Use of Imaging Procedures: Descriptive Study

Eseosa Odigie[1*], BSc; Ronilda Lacson[1,2*], MD, PhD; Ali Raja[1,2], MBA, MD; David Osterbur[1,3], PhD; Ivan Ip[1,2], MPH, MD; Louise Schneider[1,2], MD; Ramin Khorasani[1,2], MPH, MD

[1]Center for Evidence-Based Imaging, Brigham and Women's Hospital, Brookline, MA, United States

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

[3]Countway Medical Library, Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Ronilda Lacson, MD, PhD
Center for Evidence-Based Imaging
Brigham and Women's Hospital
20 Kent Street
Brookline, MA, 02445
United States
Phone: 1 6175259712
Fax: 1 6175257575
Email: rlacson@rics.bwh.harvard.edu

Abstract

Background: Evidence-based guidelines and recommendations can be transformed into “If-Then” Clinical Evidence Logic Statements (CELS). Imaging-related CELS were represented in standardized formats in the Harvard Medical School Library of Evidence (HLE).

Objective: We aimed to (1) describe the representation of CELS using established Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT), Clinical Quality Language (CQL), and Fast Healthcare Interoperability Resources (FHIR) standards and (2) assess the limitations of using these standards to represent imaging-related CELS.

Methods: This study was exempt from review by the Institutional Review Board as it involved no human subjects. Imaging-related clinical recommendations were extracted from evidence sources and translated into CELS. The clinical terminologies of CELS were represented using SNOMED CT and the condition-action logic was represented in CQL and FHIR. Numbers of fully and partially represented CELS were tallied.

Results: A total of 765 CELS were represented in the HLE as of December 2018. We were able to fully represent 137 of 765 (17.9%) CELS using SNOMED CT, CQL, and FHIR. We were able to represent terms using SNOMED CT in the temporal component for action (“Then”) statements in CQL and FHIR in 755 of 765 (98.7%) CELS.

Conclusions: CELS were represented as shareable clinical decision support (CDS) knowledge artifacts using existing standards—SNOMED CT, FHIR, and CQL—to promote and accelerate adoption of evidence-based practice. Limitations to standardization persist, which could be minimized with an add-on set of standard terms and value sets and by adding time frames to the CQL framework.

(JMIR Med Inform 2019;7(2):e13590) doi:10.2196/13590

KEYWORDS

knowledge representation; guidelines; evidence-based medicine; clinical decision support

Introduction

Background

Imaging clinical decision support (CDS) applies health information technology (IT) to inform clinical decision making at the point of care regarding the need for imaging or the optimal study based on the best available evidence [1]. Legislation has called for the use of health IT, including CDS, for health promotion and health quality improvement [2,3]. Subsequently, regulations promulgated in response to the Protecting Access to Medicare Act (PAMA) state that health care providers should reference appropriate use criteria or evidence-based clinical knowledge while ordering certain advanced imaging exams [4]. Such an evidence-based approach to appropriate medical imaging by way of CDS systems can help mitigate health care costs and imaging utilization, while providing appropriate and safe health care to those who require these procedures [5-7].

Many guidelines, recommendations, systematic reviews, and clinical decision rules have been published or endorsed by national societies in the peer-reviewed literature and as best practices by other provider groups related to appropriate use of advanced imaging procedures for certain indications. The knowledge contained in these recommendations and guidelines can be transformed into Clinical Evidence Logic Statements (CELS) that can be implemented into CDS systems. However, to be widely shared and usable in such systems, CELS must be translated into established standardized syntax and formats such as Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) [8], Clinical Quality Language (CQL) [9], and Fast Healthcare Interoperability Resources (FHIR) [10]. SNOMED International does not charge for the use of SNOMED CT in SNOMED International Member countries or territories; CQL and FHIR are Health Level Seven (HL7) standards and are available at no cost under a licensing agreement by which HL7 will retain its copyright. Key components of each standard are summarized in the subsections that follow.

Systematized Nomenclature of Medicine—Clinical Terms Compositional Grammar

SNOMED CT compositional grammar is a standard ontology for representing clinical concepts and establishes relationships between them [8]. Clinical terms such as “X-ray knee” can be modeled in SNOMED CT, where each concept is linked to an identifying number. Concepts in SNOMED CT are organized into expressions. Precoordinated expressions are represented by a single concept identifier. Postcoordinated expressions are those that are represented by combining two or more concept identifiers. SNOMED CT establishes rules and hierarchies that define attributes, qualifiers, and relationships between concepts [11]. SNOMED CT also enables reference sets, which can be used to group SNOMED CT components (ie, concepts).

Health Level Seven Clinical Quality Language Standard

The HL7 CQL Specification was developed to standardize the representation of clinical logic for clinical quality improvement [12]. More specifically, CQL was developed with the target of harmonizing expression logic. An additional component of the CQL Standard is the Expression Logical Model (ELM) [12]. Each CQL logic file is also represented as an ELM Extensible Markup Language (XML) document, which allows for an action to be represented for CDS. CQL files can reference clinical terms represented using SNOMED CT [8]. CQL files can also reference data models, such as the Quality Information and Clinical Knowledge (QUICK) logical model [13]. The QUICK data model defines the format and structure of the “retrieve” expressions in a CQL library. The retrieve declaration gathers a list of clinical data that is specific to the context of the patient or the population and to the retrieve itself.

Fast Healthcare Interoperability Resources

FHIR is a standard for sharing health care information with multiple functional areas known as resources [10]. These modules or individual components can be combined into a framework that can be implemented in a health care system. The modules are generated in a format that can be recognized and utilized by most health care systems, while also allowing for flexibility and customization of these resources through extensions. Data representation in FHIR can be in the XML, JavaScript Object Notation (JSON), or Turtle formats and it uses both CQL and SNOMED CT standards in its representations. The FHIR “decisionsupportrule” resource [14], expressed through the ELM, represents shareable knowledge artifacts for CDS.

The Harvard Medical School Library of Evidence (HLE) provides a repository of medical evidence, publicly available from the HLE website, from a range of recommendation sources that can be utilized in CDS systems [15,16]. Each unit of medical evidence is represented as a CELS of “If-Then” logic statement form (eg, If [age>X] And [symptom] Then Not [procedure]). We aimed to (1) describe the representation of CELS using the established standards of SNOMED CT, CQL, and FHIR and (2) assess the limitations of using these standards to represent the CELS in the HLE.

Methods

Study Design and Setting

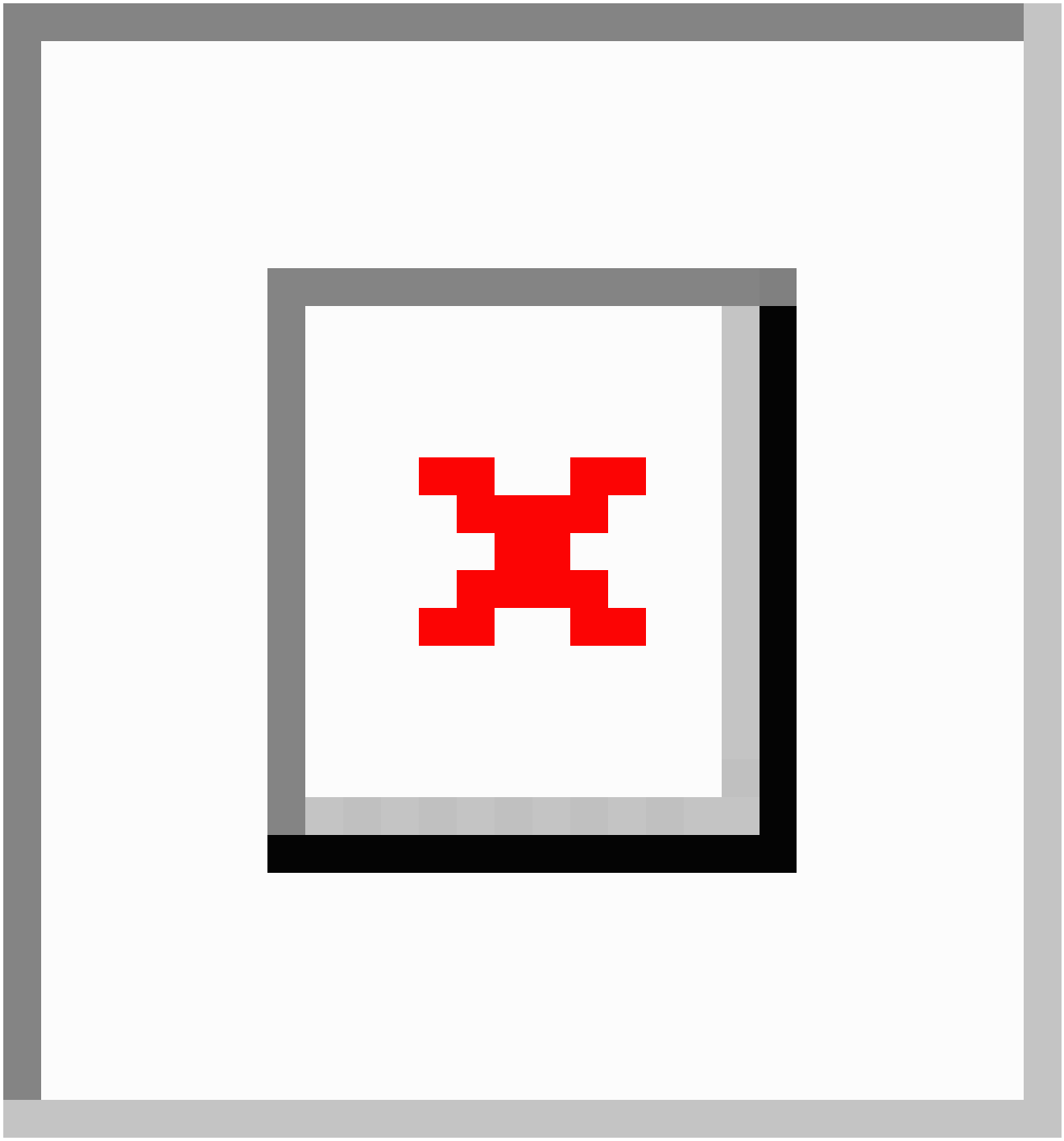
This descriptive study was exempt from the requirement of review from the Institutional Review Board as it did not include human subjects. The HLE currently contains imaging-related recommendations from clinical decision rules, professional society guidelines, and locally developed best practice guidelines [17]. As of December 20, 2018, there were a total of 765 completely graded CELS from 134 evidence sources in the HLE. A total of 235 of the CELS are Choosing Wisely content [18], pertaining to Priority Clinical Areas (eg, cervical or neck pain and suspected pulmonary embolism) specified by the Centers for Medicare and Medicaid Services [4,19].

Representing Clinical Evidence Logic Statements in Established Standards: An Overview

Steps in the process of translating a unit of evidence into FHIR so that it can be used in CDS are summarized in Figure 1. Recommendations are extracted from evidence sources, including published guidelines, recommendations, systematic reviews, clinical decision rules, and local best practices; each

extracted recommendation is known as a unit of evidence. Each unit of evidence is then organized into an "If [condition] Then [action]" format which is known as a CELS. Therefore, each CELS consists of clinical terms and logic operators and has associated metadata (eg, source and author).

Figure 1. Relationship between the standards. HL7: Health Level Seven.



Transforming a Unit of Evidence to Fast Healthcare Interoperability Resources: An Example

Overview

An in-depth transformation of a unit of evidence to FHIR is described below using a peer-reviewed article with recommendations for using ventilation-perfusion single-photon emission computed tomography (VQ SPECT) imaging for diagnosing pulmonary embolism [20]. The article recommends using VQ SPECT in patients with suspected pulmonary embolism (PE), and can be written as the following CELS: "If [Suspected PE] Then [VQ SPECT]."

Previous studies related to the HLE have identified three main types of variations in logic: single-decision statements, branching statements, and score-based statements [16]. The "Suspected PE" recommendation is an example of a single-decision statement.

Representing the Terms Using Systematized Nomenclature of Medicine: Clinical Terms

We modeled “Suspected PE” in SNOMED CT as follows: code “suspected PE”: '417113001'. This is an example of a precoordinated match.

Representing Clinical Logic in Clinical Quality Language

The CQL file is structured into a series of categories including the following:

1. Library: this is the name of the reference file, which is referenced by the secondary ELM file needed for each clinical decision.
2. Using: this term defines the data model that will be used (eg, QUICK).
3. Code System: this identifies the standardized code system, such as SNOMED CT.
4. Value Set: this identifies the specific codes within the code system that will be referenced in the clinical logic; either extensional or intensional value sets can be used.
5. Context: this can either be patient or population. For clinical logic in the HLE, the context is patient, as most data references the patient.
6. Define: this is a statement that creates a local name for conditions (eg, in an “If” statement).

Figure 2. Suspected pulmonary embolism Clinical Quality Language (CQL) file. QUICK: Quality Information and Clinical Knowledge; SNOMED: Systematized Nomenclature of Medicine; PE: pulmonary embolism.

```
USING QUICK

codesystem "SNOMED": 'https://www.snomed.org/'

code "suspected PE": '417113001' from "SNOMED" display 'suspected PE'

context Patient

define "C1":
        exists ["Condition": "suspected PE"]
```

Although one can define an infinite number of subsets, these *define* statements should be organized and succinct. Naming the *define* statements creates a local name for all the conditions and rules that either exist or do not exist to make up a defined statement subset; this also allows one to reference the list of conditions in a future *define* statement, so that the *define* statements can be stacked. In most cases, the last *define* statement in the CQL file will be a *define* statement that contains all the conditions in the “If” statement that must be true to initiate the “Then” portion of the recommendation. A complete CQL file is shown in Figure 2.

Representation in Fast Healthcare Interoperability Resources

The ELM file (see Figure 3) is the second file necessary to share clinical logic written in CQL. As mentioned previously, the ELM is a machine-readable, canonical representation of the CQL logic, which is the intermediate step in implementing the logic written in CQL. This is where the Event, Context, and Actions are defined. The organization of the ELM file is dictated by the FHIR standard “decisionsupportrule” resource. It can be formatted in the XML, JSON, or Turtle formats; HLE uses XML.

The setup of the ELM file is shown in Textbox 1. Each individual decision rule, which thereby contains an individual action, has its own XML file.

Figure 3. Suspected pulmonary embolism Extensible Markup Language (XML) file. HL7: Health Level Seven; FHIR: Fast Healthcare Interoperability Resources; CELS: Clinical Evidence Logic Statement; VQ SPECT: ventilation-perfusion single-photon emission computed tomography; PE: pulmonary embolism; SNOMED: Systematized Nomenclature of Medicine.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<DecisionSupportRule xmlns="http://hl7.org/fhir">
    <action>
        <participantType/>
        <textEquivalent value="CELS can be applied"/>
        <title value="CELS can be applied"/>
        <type value="create"/>
    </action>
    <condition value="C1"/>
    <id value="395-1"/>
    <library>
        <reference value="395-1.cql"/>
    </library>
    <moduleMetadata>
        <contributor>
            <name value="[redacted]"/>
            <type value="editor"/>
        </contributor>
        <contributor>
            <name value="[redacted]"/>
            <type value="editor"/>
        </contributor>
        <contributor>
            <name value="[redacted]"/>
            <type value="editor"/>
        </contributor>
        <experimental value="true"/>
        <identifier/>
        <lastReviewDate value="08/24/2018"/>
        <publicationDate value="11/03/2017"/>
        <publisher value="[redacted]"/>
        <status value="draft"/>
        <title value="CELS 395-1 V/Q SPECT Chest  for suspected PE"/>
        <topic>
            <text value="Three-year clinical experience with VQ SPECT for diagnosing
pulmonary embolism: diagnostic performance."/>
        </topic>
        <type value="decision-support-rule"/>
        <version value="1.0.0"/>
    </moduleMetadata>
    <trigger>
        <eventData>
            <codeFilter>
                <code/>
                <path value="code SNOMED"/>
            </codeFilter>
            <type value="VQ SPECT"/>
        </eventData>
        <eventName>V/Q SPECT_Chest_order</eventName>
        <type>data-added</type>
    </trigger>
</DecisionSupportRule>
```

Textbox 1. Setup of the Expression Logical Model (ELM) file.

<action></action>: This portion of the Extensible Markup Language (XML) file references whether the clinical logic deems the test inappropriate or appropriate. The action can either be “Rule can be applied” if the test is appropriate or “Rule cannot be applied” if the test is not appropriate.

<condition value = />: In this part of the file, the name of the final *define* statement that contains all the conditions that renders the action true is referenced. Thus, a subset of clinical logic is referenced.

<moduleMetadata></moduleMetadata>: General information about the library file is referenced (eg, author names and title).

<library></library>: This portion of the XML file references the name of the Clinical Quality Language (CQL) file that contains the logic relevant to the action and condition.

<trigger></trigger>: This portion of the XML file references the clinical order that is related to the clinical logic and contains the event that triggers a decision rule. This trigger is defined in the implementation environment, and not defined in CQL.

The FHIR framework allows for the combination of multiple CQL files with their corresponding XML files. FHIR supports single-decision statements, branching-decision statements, and score-based statements.

Representation of Branching Statements

Branching statements are recommendations that are applicable to patients with similar indications but fulfil various criteria; for example, recommendations for managing pulmonary embolism in pregnant patients versus nonpregnant patients. Thus, there are more than two CELS generated for the evidence source. The first step of translating these units of evidence is creating a decision tree. Each end point of the decision tree corresponds to a CELS, to be represented in CQL.

Representation of Score-Based Statements

Score-based units of evidence also produce more than two CELS, corresponding to evidence-based scores; for example, recommendation for managing acute appendicitis for an Acute Inflammatory Response (AIR) score of 5. The ELM files created for each CQL file in the branched and score-based statement follow the same format as the ELM files in the single-decision statement.

Assessing Representation of Clinical Evidence Logic Statements in Established Standards

The HLE contained a total of 2616 CELS at the time of data analysis for this publication. Among these, we counted the number of CELS that we were able to represent in SNOMED CT, CQL, and FHIR and reported this as a percentage of the total number of cells. For each of these, we characterized those CELS that could not be represented in SNOMED CT and those that could not be represented in CQL. CELS were defined as represented in SNOMED CT when all terms in the CELS could be represented using SNOMED CT. CELS are defined as represented in CQL when the action of the CELS, after the “Then” portion, could be represented in the ELM in the FHIR format.

Results

We were able to represent terms using SNOMED CT in the temporal component for action ("Then") statements in CQL and FHIR in 755 of 765 (98.7%) of CELS. Of the completely graded 765 CELS in the evidence library, 17.9% (n=137) were fully represented using SNOMED CT, CQL, and FHIR (see Table 1).

Reasons why CELS were not adequately represented are included in Textboxes 2-4 and are summarized as follows:

1. Clinical terms are unrepresented using SNOMED CT. Some clinical terms within logic statements contained one or more clinical terms not represented in SNOMED CT (eg, AIR score for acute appendicitis) [21-31].
2. Standard English phrases were unrepresented using SNOMED CT. Some common phrases that were not represented using SNOMED CT include "new feature" or "vehicle rollover."
3. Temporal phrases were unrepresented in CQL. An additional number of CELS were not adequately represented as the "Then" portion of the logic statement because a temporal component could not be represented in CQL (eg, computed tomography [CT] chest in 12 months) and, subsequently, with the FHIR "decisionsupportrule" resource (see Textbox 4).

Table 1. Partially represented CELS[a] in the Harvard Medical School Library of Evidence.

Type of CELS	Number of CELS (N=765), n (%)
CELS fully represented using SNOMED CT[b], CQL[c], and FHIR[d]	137 (17.9)
CELS partially represented using SNOMED CT	628 (82.1)
CELS partially represented due to CQL	10 (1.3)

[a]CELS: Clinical Evidence Logic Statement.

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

[c]CQL: Clinical Quality Language.

[d]FHIR: Fast Healthcare Interoperability Resources.

Textbox 2. Clinical terms unrepresented using Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT).

- Acute Inflammatory Response (AIR) score [21]
- Alvarado score [22]
- Canadian Computed Tomography (CT) Head Rule [23]
- Canadian Cervical Spine Rule (CCSR) [24]
- New Orleans/Charity head trauma rule [25]
- National Emergency X-Radiography Utilization Study (NEXUS) head trauma rule [26]
- Magnetic resonance imaging (MRI) shoulder with dedicated metal suppression protocol
- O_2 saturation on room air
- Optimizing imaging in suspected appendicitis (OPTIMAP) score [27]
- Revised Geneva (rGeneva) score [28]
- Simple calculated osteoporosis risk estimation (SCORE) score [29]
- Simplified Motor Score (SMS) [30]
- Sex, timing, origin, nausea, erythrocytes (STONE) score [31]

Textbox 3. Standard English phrases unrepresented using Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT).

- New feature
- Suitable candidate
- Time-of-flight (TOF) magnetic resonance angiography (MRA)
- Vehicle rollover

Textbox 4. Temporal phrases unrepresented in Clinical Quality Language (CQL).

- Computed tomography (CT) chest in 12 months
- CT chest in 18-24 months
- CT chest in 3 months, 9 months, And 24 months
- CT chest in 3-6 months
- CT chest in 6-12 months
- CT chest in 9-12 months And 24 months
- Low-dose CT annually for 3 years

Examples of partially unrepresented CELS include:

1. Alvarado score for suspected appendicitis; this is an example of an evidence source with three partially represented CELS, since the term "Alvarado score" does not exist in the SNOMED CT standard ontology, as indicated by the asterisk:
   a. If [Alvarado score* >=4] And [Alvarado score* <=6] Then [CT abdomen]
   b. If [Alvarado score* <4] Then Not [CT Abdomen]
   c. If [Alvarado score* >6] Then Not [CT Abdomen]
2. Guidelines for management of small pulmonary nodules detected on CT scans—a statement from the Fleischner Society [32]:
   a. If [pulmonary nodule on chest CT] And [nodule size <=4mm] And [high risk] Then [CT chest in 12 months]
   b. If [pulmonary nodule on chest CT] And [nodule size >4mm] And [nodule size <=6] And [low risk] Then [CT Chest in 12 months]
   c. If [pulmonary nodule on chest CT] And [nodule size >4mm] And [nodule size <=6] And [high risk] Then [CT chest in 6-12 months]

These CELS are examples of partially represented CELS due to actions such as "CT chest in 6-12 months" and "CT chest in 12 months." These actions contain a future temporal component that cannot be represented in CQL. CT chest can be represented in a define statement. However, a define statement in CQL for scheduling a procedure at a future time cannot be created. Furthermore, value sets for terms such as "high risk" are not available.

Discussion

Principal Findings

Overall, 17.9% (137/765) of CELS were represented as shareable CDS knowledge artifacts using existing standards, SNOMED CT, FHIR, and CQL to promote and accelerate adoption of evidence-based practice. More work to represent imaging-related CELS need to be undertaken to standardize clinical knowledge included in the HLE. A few limitations to utilizing these standards for CDS implementation in the evidence library were identified. While SNOMED CT is robust, some terms do not exist in its ontology. For example, names for known rules or scores such as "Revised Geneva score" [28] cannot be represented. The HLE is currently in the process of creating an add-on set of terms in SNOMED CT so that these terms will have an ID and mapping.

In addition, English words, which contribute to the meaning of a clinical recommendation (eg, "high risk" and "suitable candidate"), may not be represented using SNOMED CT. In those situations, one can substitute terms that are in SNOMED CT that are a synonym or close in meaning to the original term. Use of value sets to enumerate concepts that may map to a criterion in a recommendation may be useful. However, these mappings are not always exact and may change the interpretation of the clinical recommendation. This limitation can also possibly be amended through an add-on set of terms to SNOMED CT. SNOMED CT is updated twice yearly and updates can include newly added concepts. In addition, developing an add-on set of terms can be expedited by creating more value sets, specifically sets of code from hierarchy-based definitions that are algorithmically defined (ie, intensional value sets) or enumerated (ie, extensional value sets). These can be disseminated publicly (eg, via the Value Set Authority Center) to accelerate cross-organizational efforts for terminology standardization. More importantly, guidelines and recommendations should be limited to using standardized terminology prior to getting published.

The limitations of the FHIR framework include determining ways to represent temporal actions and phrases. CQL gives rise to a temporal framework as the time frame of a condition can be defined; for example, "If CT chest in the past 12 months" can be represented. However, this allowance is limited to conditions within the "If" statement, and currently there is no temporal framework in the CQL file or future temporal component in the ELM file that runs an action (eg, "CT chest in 12 months" cannot be written in the XML file). Further, recurring actions such as "9-12 months And 24 months" cannot be represented. The current actions determine whether imaging at that exact time is appropriate or not. Clinical logic could be restructured in the "If-Then" statement to incorporate the time frame into the "If," but this does not work in every situation. A systemized and structured approach to these statements with time frames should be added onto the CQL and FHIR framework by developers or CDS implementation services that expands upon the representation of time within the actions of CDS.

To summarize, a unit of evidence in the HLE is structured as a CELS. We represented terms using a standard terminology, SNOMED CT. The conditions or "If" statements are represented in CQL. Using the FHIR resource "decisionsupportrule," we

combine the action and the condition to represent CDS knowledge artifacts. CELS are publicly available and represented using existing standards to promote and accelerate adoption of evidence into daily practice to improve the quality of care and reduce waste.

Conclusions

CELS were represented as shareable CDS knowledge artifacts using existing standards—SNOMED CT, FHIR, and CQL—to promote and accelerate adoption of evidence-based practice. However, more work needs to be done to represent terminology and value sets and to model future temporal action in CDS recommendations.

Acknowledgments

The authors would like to thank Ms Laura Peterson for reviewing the manuscript.

Authors' Contributions

All authors contributed to the study design and data acquisition; EO, RL, and RK are responsible for data analysis and interpretation. All authors contributed significant intellectual content during the manuscript preparation and revisions, approved the final version, and accept accountability for the overall integrity of the research process and the manuscript.

Conflicts of Interest

None declared.

References

1. Khorasani R, Hentel K, Darer J, Langlotz C, Ip I, Manaker S, et al. Ten commandments for effective clinical decision support for imaging: Enabling evidence-based practice to improve quality and reduce waste. AJR Am J Roentgenol 2014 Nov;203(5):945-951. [doi: 10.2214/AJR.14.13134] [Medline: 25341131]
2. Department of Health and Human Services. Federal Register (volume 72, issue 209). 2009 Oct 30. HIPAA administrative simplification: Enforcement URL: https://www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf [accessed 2019-05-03] [WebCite Cache ID 787zG6aj2]
3. Gold M, McLaughlin C. Assessing HITECH implementation and lessons: 5 years later. Milbank Q 2016 Sep;94(3):654-687 [FREE Full text] [doi: 10.1111/1468-0009.12214] [Medline: 27620687]
4. Medicare: Considerations for Expansion of the Appropriate Use Criteria Program. Washington, DC: US Government Accountability Office; 2015 Sep 30. URL: https://www.gao.gov/assets/680/672856.pdf [accessed 2019-05-04] [WebCite Cache ID 7897g4drI]
5. Bright T, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux R, et al. Effect of clinical decision-support systems: A systematic review. Ann Intern Med 2012 Jul 03;157(1):29-43. [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
6. Dunne R, Ip I, Abbett S, Gershanik E, Raja A, Hunsaker A, et al. Effect of evidence-based clinical decision support on the use and yield of CT pulmonary angiographic imaging in hospitalized patients. Radiology 2015 Jul;276(1):167-174 [FREE Full text] [doi: 10.1148/radiol.15141208] [Medline: 25686367]
7. Semelka R, Armao D, Elias JJ, Huda W. Imaging strategies to reduce the risk of radiation in CT studies, including selective substitution with MRI. J Magn Reson Imaging 2007 May;25(5):900-909. [doi: 10.1002/jmri.20895] [Medline: 17457809]
8. US National Library of Medicine. Bethesda, MD: US National Library of Medicine, National Institutes of Health, Health & Human Services; 2009. SNOMED CT URL: http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html [accessed 2019-04-21] [WebCite Cache ID 77pK7mQhp]
9. Lin Y, Staes C, Shields D, Kandula V, Welch B, Kawamoto K. Design, development, and initial evaluation of a terminology for clinical decision support and electronic clinical quality measurement. AMIA Annu Symp Proc 2015;2015:843-851 [FREE Full text] [Medline: 26958220]
10. Benson T, Grieve G. Principles of Health Interoperability: SNOMED CT, HL7 and FHIR. 3rd edition. London, UK: Springer-Verlag; 2016.
11. SNOMED CT: Compositional Grammar v2.3.1 - Specification and Guide. London, UK: International Health Terminology Standards Development Organisation (IHTSDO); 2016 Nov 11. URL: https://tinyurl.com/y238f5sk [accessed 2019-05-04] [WebCite Cache ID 789B9FqZx]
12. HL7.org. 2018. HL7 Standard: Clinical Quality Language Specification, Release 1 STU3 (CQL 1.3) URL: https://cql.hl7.org/ [accessed 2019-04-21] [WebCite Cache ID 77pLYH8Db]
13. HL7. HL7.org. 2019. Quick data model URL: http://hl7.org/fhir/us/qicore/2018Jan/quick/index.html [accessed 2019-04-21] [WebCite Cache ID 77pM9tCpe]
14. HL7.org. FHIR STU 3 Candidate URL: http://hl7.org/fhir/2016May/resourcelist.html [accessed 2019-04-21] [WebCite Cache ID 77pMLBD3Z]

15. Lacson R, Raja A, Osterbur D, Ip I, Schneider L, Bain P, et al. Assessing strength of evidence of appropriate use criteria for diagnostic imaging examinations. J Am Med Inform Assoc 2016 Jan;23(3):649-653. [doi: 10.1093/jamia/ocv194] [Medline: 26911819]
16. Varada S, Lacson R, Raja A, Ip I, Schneider L, Osterbur D, et al. Characteristics of knowledge content in a curated online evidence library. J Am Med Inform Assoc 2018 May 01;25(5):507-514. [doi: 10.1093/jamia/ocx092] [Medline: 29092054]
17. Harvard Medical School Library of Evidence. URL: http://libraryofevidence.med.harvard.edu/ [accessed 2019-04-21] [WebCite Cache ID 77pMU1Cv7]
18. Choosing Wisely. URL: http://www.choosingwisely.org/ [accessed 2019-04-21] [WebCite Cache ID 77pMei8Hj]
19. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare program: Medicare clinical diagnostic laboratory tests payment system. Final rule. Fed Regist 2016 Jun 23;81(121):41035-41101 [FREE Full text] [Medline: 27373013]
20. Grüning T, Drake B, Farrell S, Nokes T. Three-year clinical experience with VQ SPECT for diagnosing pulmonary embolism: Diagnostic performance. Clin Imaging 2014;38(6):831-835. [doi: 10.1016/j.clinimag.2014.04.003] [Medline: 24852679]
21. Scott A, Mason S, Arunakirinathan M, Reissis Y, Kinross J, Smith J. Risk stratification by the Appendicitis Inflammatory Response score to guide decision-making in patients with suspected appendicitis. Br J Surg 2015 Apr;102(5):563-572. [doi: 10.1002/bjs.9773] [Medline: 25727811]
22. McKay R, Shepherd J. The use of the clinical scoring system by Alvarado in the decision to perform computed tomography for acute appendicitis in the ED. Am J Emerg Med 2007 Jun;25(5):489-493. [doi: 10.1016/j.ajem.2006.08.020] [Medline: 17543650]
23. Stiell I, Wells G, Vandemheen K, Clement C, Lesiuk H, Laupacis A, et al. The Canadian CT Head Rule for patients with minor head injury. Lancet 2001 May 05;357(9266):1391-1396. [Medline: 11356436]
24. Coffey F, Hewitt S, Stiell I, Howarth N, Miller P, Clement C, et al. Validation of the Canadian c-spine rule in the UK emergency department setting. Emerg Med J 2011 Oct;28(10):873-876. [doi: 10.1136/emj.2009.089508] [Medline: 21068171]
25. Haydel M, Preston C, Mills T, Luber S, Blaudeau E, DeBlieux P. Indications for computed tomography in patients with minor head injury. N Engl J Med 2000 Jul 13;343(2):100-105. [doi: 10.1056/NEJM200007133430204] [Medline: 10891517]
26. Mower W, Hoffman J, Herbert M, Wolfson A, Pollack CJ, Zucker M, NEXUS II Investigators. Developing a decision instrument to guide computed tomographic imaging of blunt head injury patients. J Trauma 2005 Oct;59(4):954-959. [Medline: 16374287]
27. Leeuwenburgh M, Stockmann H, Bouma W, Houdijk A, Verhagen M, Vrouenraets B, OPTIMAP Study Group. A simple clinical decision rule to rule out appendicitis in patients with nondiagnostic ultrasound results. Acad Emerg Med 2014 May;21(5):488-496 [FREE Full text] [doi: 10.1111/acem.12374] [Medline: 24842498]
28. Le Gal G, Righini M, Roy P, Sanchez O, Aujesky D, Bounameaux H, et al. Prediction of pulmonary embolism in the emergency department: The revised Geneva score. Ann Intern Med 2006 Feb 07;144(3):165-171. [Medline: 16461960]
29. Cadarette S, Jaglal S, Murray T. Validation of the simple calculated osteoporosis risk estimation (SCORE) for patient selection for bone densitometry. Osteoporos Int 1999;10(1):85-90. [doi: 10.1007/s001980050199] [Medline: 10501785]
30. Haukoos J, Gill M, Rabon R, Gravitz C, Green S. Validation of the Simplified Motor Score for the prediction of brain injury outcomes after trauma. Ann Emerg Med 2007 Jul;50(1):18-24. [doi: 10.1016/j.annemergmed.2006.10.004] [Medline: 17113193]
31. Moore C, Bomann S, Daniels B, Luty S, Molinaro A, Singh D, et al. Derivation and validation of a clinical prediction rule for uncomplicated ureteral stone--the STONE score: Retrospective and prospective observational cohort studies. BMJ 2014 Mar 26;348:g2191 [FREE Full text] [doi: 10.1136/bmj.g2191] [Medline: 24671981]
32. MacMahon H, Austin J, Gamsu G, Herold C, Jett J, Naidich D, Fleischner Society. Guidelines for management of small pulmonary nodules detected on CT scans: A statement from the Fleischner Society. Radiology 2005 Nov;237(2):395-400. [doi: 10.1148/radiol.2372041887] [Medline: 16244247]

Abbreviations

AIR: Acute Inflammatory Response
CCSR: Canadian Cervical Spine Rule
CDS: clinical decision support
CELS: Clinical Evidence Logic Statement
CQL: Clinical Quality Language
CT: computed tomography
ELM: Expression Logical Model
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level Seven
HLE: Harvard Medical School Library of Evidence
IT: information technology
JSON: JavaScript Object Notation
MRA: magnetic resonance angiography

MRI: magnetic resonance imaging
NEXUS: National Emergency X-Radiography Utilization Study
OPTIMAP: optimizing imaging in suspected appendicitis
PAMA: Protecting Access to Medicare Act
PE: pulmonary embolism
QUICK: Quality Information and Clinical Knowledge
rGeneva: Revised Geneva
SCORE: simple calculated osteoporosis risk estimation
SMS: Simplified Motor Score
SNOMED CT: Systematized Nomenclature of Medicine—Clinical Terms
STONE: sex, timing, origin, nausea, erythrocytes
TOF: time-of-flight
VQ SPECT: ventilation-perfusion single-photon emission computed tomography
XML: Extensible Markup Language

Edited by G Eysenbach; submitted 01.02.19; peer-reviewed by J Durack, D Willett; comments to author 13.03.19; revised version received 21.03.19; accepted 24.03.19; published 13.05.19.

Please cite as:
Odigie E, Lacson R, Raja A, Osterbur D, Ip I, Schneider L, Khorasani R
Fast Healthcare Interoperability Resources, Clinical Quality Language, and Systematized Nomenclature of Medicine—Clinical Terms in Representing Clinical Evidence Logic Statements for the Use of Imaging Procedures: Descriptive Study
JMIR Med Inform 2019;7(2):e13590
URL: http://medinform.jmir.org/2019/2/e13590/
doi:10.2196/13590
PMID:31094359

Original Paper

Development of a Consumer Health Vocabulary by Mining Health Forum Texts Based on Word Embedding: Semiautomatic Approach

Gen Gu[1], MS; Xingting Zhang[2], MS; Xingeng Zhu[1], BS; Zhe Jian[3], BS; Ken Chen[1], BS; Dong Wen[2], MS; Li Gao[4], MD; Shaodian Zhang[1,5], PhD; Fei Wang[1,6], PhD; Handong Ma[1], MS; Jianbo Lei[2,7], PhD, MD

[1]Synyi Research, Shanghai, China

[2]Center for Medical Informatics, Peking University, Beijing, China

[3]Harbin Medical University, Harbin, China

[4]School of Stomatology, Peking University, Beijing, China

[5]APEX Data & Knowledge Management Lab, Shanghai Jiao Tong University, Shanghai, China

[6]Department of Healthcare Policy and Research, Weill Cornell Medicine, New York, NY, United States

[7]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou city, Sichuan Province, China

Corresponding Author:
Jianbo Lei, PhD, MD
Center for Medical Informatics
Peking University
No. 1 Administration Building, Peking University Health Science Center, Haidian Dist
Beijing,
China
Phone: 86 (010)82805906
Email: jblei@hsc.pku.edu.cn

Abstract

Background: The vocabulary gap between consumers and professionals in the medical domain hinders information seeking and communication. Consumer health vocabularies have been developed to aid such informatics applications. This purpose is best served if the vocabulary evolves with consumers' language.

Objective: Our objective is to develop a method for identifying and adding new terms to consumer health vocabularies, so that it can keep up with the constantly evolving medical knowledge and language use.

Methods: In this paper, we propose a consumer health term–finding framework based on a distributed word vector space model. We first learned word vectors from a large-scale text corpus and then adopted a supervised method with existing consumer health vocabularies for learning vector representation of words, which can provide additional supervised fine tuning after unsupervised word embedding learning. With a fine-tuned word vector space, we identified pairs of professional terms and their consumer variants by their semantic distance in the vector space. A subsequent manual review of the extracted and labeled pairs of entities was conducted to validate the results generated by the proposed approach. The results were evaluated using mean reciprocal rank (MRR).

Results: Manual evaluation showed that it is feasible to identify alternative medical concepts by using professional or consumer concepts as queries in the word vector space without fine tuning, but the results are more promising in the final fine-tuned word vector space. The MRR values indicated that on an average, a professional or consumer concept is about 14th closest to its counterpart in the word vector space without fine tuning, and the MRR in the final fine-tuned word vector space is 8. Furthermore, the results demonstrate that our method can collect abbreviations and common typos frequently used by consumers.

Conclusions: By integrating a large amount of text information and existing consumer health vocabularies, our method outperformed several baseline ranking methods and is effective for generating a list of candidate terms for human review during consumer health vocabulary development.

(JMIR Med Inform 2019;7(2):e12704) doi:10.2196/12704

KEYWORDS

consumer health vocabulary; word embedding; representation learning; natural language processing; consumer health information; ontology enrichment

Introduction

Background

In 2015, a survey of Chinese internet users showed that medicine and health care are the two most popular searched topics and accounted for 55.15% of all searches [1]. However, it is difficult for most users to express medical concepts using professional terms such as bronchus, brain, and extracellular space [2-4], and online forums and news media explain such professional medical terms with very little detail. The gap between consumer language and medical terminology makes searching and retrieving information difficult and biases the understanding of health information [5-7].

Development of the consumer health vocabularies, which map languages of consumers and medical experts, is a potential way to bridge the gap. Several commercial and noncommercial groups, such as Apelon and Public Health Terminology by Intelligent Medical Objects, and Open Access and Collaborative Consumer Health Vocabulary [8], tried to bind consumer health vocabularies with the Unified Medical Language System (UMLS) or the International Classification of Diseases (ICD). Several factors dominate the expansion of a quality consumer health vocabulary: a comprehensive search to identify related nonstandard expressions, abbreviations and common typos, a consensus between consumers' point of view and professional classification, and periodic updating for new terms. These factors make the expansion process complicated, costly, and time-consuming.

To accelerate the expansion process, researchers developed many approaches to extract and map consumer terms automatically or semiautomatically, including the n–gram-based approach [9], pattern-based approach [7], co-occurrence analysis [10], and machine learning methods [9,11]. Although the consumer health vocabularies mined through these hand-crafted heuristic approaches are more accurate, many relevant pairs could be missing. Recent theoretical and experimental results from Wang et al [12] showed that matching professional-consumer concept pairs through text embedding approaches can capture the semantic similarities between professional concepts and consumer concepts, thus yielding a high recall. However, with only unsupervised algorithms, many irrelevant pairs could be generated. In order to retain the advantage of text embedding and improve precision as much as possible, we propose a semisupervised representation learning method to make the concept embedding specific in the consumer vocabulary mining process. With the knowledge introduced by the reviewer, concept embeddings can continuously improve themselves. Our approach provides a related consumer term list sorted by their semantic distance to a particular medical term and helps reviewers identify synonym pairs efficiently. We extracted consumer health terms from one of the most popular health forums in China and manually evaluated the performance of our approach. The experimental results are promising, showing performance improvement of up to 16% with a small amount of seed pairs.

Synonym Identification

Two mainstream approaches for identifying synonyms are rule-based algorithm and word similarity measurement. A rule-based algorithm identifies synonyms by semantic patterns. For example, Vydiswaran et al took advantage of common linking phrases such as "also called," "also known as," and "also referred to as" to extract synonyms from Wikipedia [7]. There are many ways to calculate word similarities, including n-gram, edit-distance (Levenshtein distance), WordNet-distance, and cosine-distance between word vectors [13-16]. Among them, training distributed word vectors and extracting synonyms from top closest words are the most popular ways. Henriksson et al created word vectors using latent semantic analysis with random indexing and permutation to identify medical synonyms and abbreviation-expansion pairs [17,18]. He et al created word vectors including linguistic, contextual, and statistical features and used K-means to gather new consumer health terms on social media [19]. Elhadad et al created word vectors combining both contextual and semantic features and cluster terms from breast cancer forums into predefined semantic categories [20]. Wang et al created word vectors using word2vec, an open-source natural language processing tool released by Google, to extract symptoms from UMLS [12].

Development of Consumer Health Vocabularies

As early as 1998, Marshall [21] mapped the consumer health terminology from WellMed (a health care website) to SNOMED (Systematized Nomenclature of Medicine) and UMLS, which helped patients search information using nonprofessional expressions. In 2001, Patrick expanded UMLS, the Eurodicautom of the European Commission's Translation Service, and the European Commission Glossary of popular and technical medical terms, by adding words from the Dictionary of American Regional English, but only focused on diabetes-related terms [22]. Both Marshall and Patrick constructed their consumer health vocabularies manually, which is inefficient and unscalable. In 2005, Zeng developed a two-step approach, which combined corpus-based text analysis and manual review, to build an open-source consumer health vocabulary [23]. To reduce the labor in term mapping, Zeng improved the two-step approach by adding n-gram, logistic regression, and even natural language processing and machine learning algorithm (parts of speech, noun phrases, and named entities recognition) [9-11]. Since then, the two-step semiautomatic approach—term identification algorithms followed by manual review—evolved into a common practice in many consumer health vocabulary researches [7,10,24].

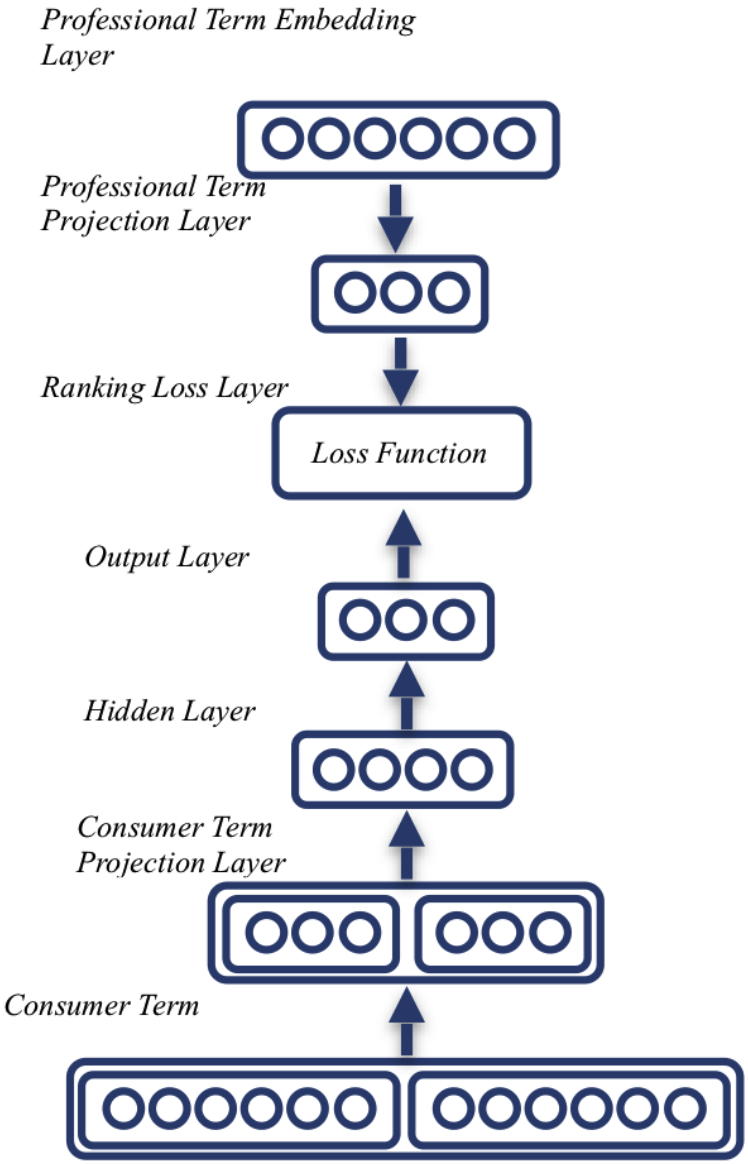
Methods

Overview

To alleviate the problems mentioned above, we propose a consumer health term–finding framework based on a distributed word vector space model. The overview of the framework is shown in Figure 1. The workflow can be interpreted as a feed-forward neural network. The first step of our approach requires a corpus of raw text for the unsupervised pretraining of the embedding matrix *E* as the embedding layer (Figure 1). Text embedding approaches have proven to be very effective in capturing the similarities between words and phrases, which can yield professional-consumer concept pairs that cannot be found by feature- and pattern-based methods.

We used THULAC (Tsinghua University - Lexical Analyzer for Chinese) [25], a Chinese segmentation tool, to change Chinese text into words. Thereafter, word embedding tools were used to compute the vector space, as described above. We collected professional concepts from the official Chinese version of ICD-10, calculated the frequency of words in the corpus, and extracted those with a count of over 1000 and their corresponding consumer concepts with context as seed pairs. All the weights were initialized uniformly at random. In order to obtain a model that takes professional concepts as input and consumer concept list as output after training, we introduced an embedding space–adapting process consisting of an embedding projection and a supervised ranking method. The embedding projection contains a projection layer, a hidden layer, and a target projection layer (Figure 1) to achieve a "smaller" embedding space that preserves more supervisory signal.

Figure 1. The overall architecture of our model. The consumer and professional terms start from both ends of the model, going through some embedding and projection layers, so that they are projected into a unified semantic space, where the ranking loss will be measured.



The output concept list is not necessarily the most similar word of the input concept in the word vector space, but with the supervised signal of seed terms, their similarity is more of a measure of professional-consumer concept pair similarity. A special ranking loss function is used in the ranking loss layer to calculate the similarity of professional-consumer concept pairs. After manually selecting the output professional-consumer concept pairs, we input the selection results into the training data. New professional-consumer concept pairs were discovered through iterations.

Word Embedding

Word embeddings are generally trained to reconstruct linguistic contexts of words by optimizing an objective function that can be measured without annotations. One popular approach is to estimate the embeddings by maximizing the probability that the words within a given window size are predicted correctly. Word embedding takes a large corpus of text as its input and produces a vector space, with each unique word in the corpus assigned a corresponding vector in the space. In word embedding training, one of the key issues is the formulation of the training objective function, minimization or maximization of which may produce meaningful word vector representations. Ideally, the training objective function should reflect the fact that the semantic word similarities measured on learned word vectors are consistent with human cognition. Recently Wang et al [26] performed a comprehensive comparative study on the different word embedding techniques in biomedical texts. Of those, we chose three popular word embedding methods: Word2Vec [27], Global Vectors (GloVe) [28], and FastText [29].

Word2Vec

Word2Vec is a widely used method in natural language processing for generating word embeddings. It has two different training strategies: (1) Continuous Bag-of-Words, in which the model is given a sequence of words without the middle one and attempts to predict this omitted word, and (2) Skip-Gram, in which the model is given a word and attempts to predict its neighboring words. In both cases, the model consists of only a single weight matrix (apart from the word embedding), which results in a fast log-linear training process that can capture semantic information [26].

Global Vectors

The GloVe method was proposed by Pennington et al [28] and obtained state-of-the-art results for syntactic and semantic analogies tasks. This method has a co-occurrence matrix M that is constructed by looking at the context words. Each element M_{ij} in the matrix represents the probability of the word *i* being similar to the word *j*. In the matrix M, the vectors are randomly generated and trained with the equation $P(w_i, w_j)=log(M_{ij})=w_iw_j+b_i+b_j$, where w_i and w_j are word vectors and b_i and b_j are biases.

FastText

FastText is a recently developed method [29] proposed by the same group who developed word2vec, in which the embeddings are associated with character n-grams and the words are represented as the summation of these representations. Specifically, a word representation is induced by summing character n-gram vectors with vectors of surrounding words. Therefore, this method attempts to capture morphological information to induce word embedding.

Adapting Embedding With Supervised Training

As mentioned in the Introduction, word embedding is a useful unsupervised technique to capture the similarity between words and phrases, which can yield high recall of professional-consumer concept pairs. It can also be used as a pretraining phase prior to supervised training. However, even if the embeddings provide compact, real, valued representations of each word in a vocabulary, it only indicates that word embeddings produce a semantic space that models synonymy to a certain degree. Current methods use pretrained embedding to initialize model parameters and then use the labeled data to guide them for the intended task (eg, we use professional-consumer concept pairs that already exist as the supervision to produce a semantic space dedicated to finding such pairs). If, as in our case, only a small amount of supervised data are available, this can lead to severe overfitting. Furthermore, rare words will receive very few updates and their embedding will be poorly adapted for our task. We propose two solutions to avoid these problems.

Embedding Projection

Let denote the original embedding matrix obtained. We define the adapted embedding matrix as the multiplication $S \cdot E$, where the projection matrix and $s< e$. We estimate the parameters of the matrix *S* using the labeled dataset, while *E* is kept fixed. In other words, we determine the optimal projection of the embedding matrix *E* into a subspace. The ideal embedding subspace relies on two fundamental principles:

1. With dimensionality reduction of the embedding, the model can better fit the complexity of our consumer health vocabularies task or the amount of available data. As the number of professional-consumer concept pairs increase, the size of the embedding can be adjusted.
2. Using a projection, all embeddings are indirectly updated, not only for the words present in the labeled dataset.

Let $M=[w_1 \ldots w_n]$ denote a message of n words. Each column $w \in \{0,1\}^{v \times 1}$ of m represents a word in one-hot form. is the projection vector for each word, given by $P= S \cdot E \cdot M$. A simple adapting rule is to keep the original *S* fixed and append a new random initial matrix to *S* to obtain the new *S'* for retraining.

Compared to a conventional feed-forward network employing embedding for natural language, two main differences arise. First, the input layer is factorized into two components—the embedding attained in unsupervised form *E* and the projection matrix *S*. Second, the size of the subspace in which the embeddings are projected is much smaller than that of the original embedding with typical reductions above one order of magnitude. As is usual in this kind of model, all the parameters can be trained with gradient methods, using the back-propagation update rule.

Supervised Ranking Method

One of the challenges for supervised word embedding training is the difficulty of defining the exact similarity values between two words. Especially in our case, the professional concept and the consumer concept are different. The similarity measure is affected by many factors such as the dimensionality of the embedding, the employed learning algorithms, and the corpus size. Although the similarity values are quite different, the ranking of similarity values is more robust than the values itself.

Inspired by this finding, we employed ranking information as the supervised training targets. The ranking loss function  is obtained as,

where V is the vocabulary, ω_v is a specific word, and

is the set of synonym words of ω_v in the labeled set.

is the rank of ω_r in the labeled set, and

is the rank of ω_r according to its cosine similarity with ω_v measured in the embedding space.

Because the ranking loss is not differentiable, we choose to minimize the semantic similarity loss between the desired ranking position and the real ranking position in the embedding space as a surrogate. Given the desired ranking position, the similarity value corresponding to the desired ranking position is employed as the real training target. Minimizing the difference of similarity values between the desired position and the real position may also reduce the ranking loss. The similarity value lies in function  , given below, where

denotes the sorted similarity values for word ω_v:

Experiment

To evaluate the effectiveness of the proposed model, three groups of experiments were designed. The three kinds of word embeddings with different vector size were further trained by the proposed model and evaluated. The baselines were the original word embeddings described above. The effect of the projection layer was studied in the second group experiments. Two comparison groups were involved, one that used the standard structure without the projection layer and another that used the proposed projection layer.

Data Sets

We tested our methods with the corpus obtained from two different Chinese communities to cover different perspectives. The Tianya community is one of the most popular online forums in China, and the data are open and easy to retrieve. The health care sector of the Tianya community—Tianya Hospital—has a large number of disease consulting posts initiated by consumers, and about 180 Mb of data are used in our experiment. Haodf is the largest Chinese medical question-and-answer website where all questions are created by patients and answered by doctors, and about 2 Gb of data are used in our experiment. We considered these to be ideal sources of consumer health corpus, used Scrapy [30] for a full-text crawling from those corpora, and removed user information before further processing. The messages from the consumer forums were preprocessed as follows: URLs were replaced with a token URL and words occurring less than 30 times in the corpus were replaced by a special UNKNOW symbol. We collected professional concepts from the Chinese official version of ICD-10, calculated the frequency of words in these two corpora, extracted them with a count of over 1000, and manually collected their corresponding consumer concepts with context as seed pairs. The annotation process was performed by one medical professional and reviewed by three medical professionals. We finally obtained 224 seed pairs that all three reviewers consistently agreed upon.

Evaluation Method

We use the mean reciprocal rank to evaluate the quality of word embeddings. Mean reciprocal rank is a statistic measure for evaluating any process that produces a list of possible response to a sample of queries and orders them by probability of correctness. The mean reciprocal rank is the average of the reciprocal ranks of results for a sample of queries Q:

where $rank_i$ refers to the rank position of the first relevant synonym for the i-th query.

We use a large collection of candidate medical concepts and build a small set of ground truth professional-consumer concept pairs. We randomly select 100 pairs from our seed pairs for evaluation.

Results

Principal Findings

In general, the performance of the proposed method is detailed in Table 1. All word embeddings are significantly enhanced after fine tuning. The performance of the best word embedding is FastText, with a 400-dimensional vector with Haodf and projection matrix size set to 40, and it is also significantly improved in all datasets. The rich n-gram features used in FastText are important in Chinese synonym finding and have much higher performance than others. These remarkable improvements demonstrate that our method may transfer the complementary knowledge from the weak embeddings into the strong embeddings.

Effect of the Projection Layer

Table 2 shows the system performance with no projection matrix and different projection matrix size. As baselines, we considered a simple log-linear approach, which uses the unsupervised embeddings directly as features in a log-linear classifier. We

tested the model performance. Furthermore, we observed that updating the embeddings always led to inferior results. This suggests that pretrained embeddings should be kept fixed, when little labeled data are available to retrain them.

Manual Review of the Recommended Consumer Health Terms

In order to ensure the accuracy of professional-consumer concept pairs, manual review is inevitable. Table 3 showed the top 10 candidates for word “diarrhea” (“腹泻”) provided by our method for reviewers. Most words illustrated here are symptoms or clinical findings related to “diarrhea,” such as “vomiting” (“呕吐”), “abdominal pain” (“腹痛”), and “dyspepsia” (“消化不良”). We see two synonyms in the table: “having loose bowels” (“拉肚子”, ranked third) and “diarrhea” (“腹泄”, ranked seventh). The former is a consumer health term that is rarely used by professionals, and the latter is a typo of “diarrhea” (“腹泻”). In the manual review, researchers reviewed a sample of the candidate terms suggested by the system to assess whether these terms should be added into the consumer health vocabularies.

Table 1. Performance of three word embedding methods with different embedding sizes. Italicized values indicate the best performance of the date set.

Corpus and tuning state	GloVe[a]			Word2Vec			FastText		
	100[b]	200	400	100	200	400	100	200	400
Tianya									
Before	0.266	0.263	0.282	0.289	0.296	0.272	0.341	0.340	0.319
After	0.313	0.308	0.320	0.325	0.338	0.341	0.355	0.362	*0.371*
Haodf									
Before	0.270	0.273	0.289	0.288	0.290	0.295	0.320	0.322	0.331
After	0.321	0.326	0.332	0.313	0.344	0.346	0.361	0.365	*0.385*

[a]GloVe: Global Vectors.

[b]Values in this row indicate embedding size.

Table 2. Performance of FastText-200 with different sizes of the projection matrix.

Corpus	Projection matrix size				
	0[a]	20	40	80	160
Tianya	0.350	0.352	0.362	0.357	0.345
Haodf	0.338	0.342	0.365	0.360	0.331

[a]Projection matrix size 0 is used to denote the baseline (log-linear model).

Table 3. Top 10 candidates for the seed word “diarrhea.”

Rank	Medical words in Chinese	Medical words in English
1	呕吐	Vomiting
2	腹胀	Ventosity
3	拉肚子	Having loose bowels
4	腹痛	Abdominal pain
5	便秘	Constipation
6	消化不良	Dyspepsia
7	腹泄	Diarrhea
8	厌食	Anorexia
9	肠鸣	Borborygmus
10	返酸	Acid reflux

Discussion

Bridging the language gap between consumers and medical professionals is a fundamental problem in medical internet research. There has been some research on building the consumer health vocabulary for English medical terms, but the research on other languages is scarce. The model developed in this paper was evaluated using Chinese terms and could help

professionals collect consumer health vocabularies related to certain clinical topics and discover synonyms in a more effective and efficient way.

From the methodology perspective, we adopted unsupervised word embedding as the backbone of our approach. This mechanism encodes words into vectors based on the context they are likely to be put into and projects them into a common semantic space. We further fine-tuned the word embeddings to make them align with the limited supervision information provided. A previous study used word vectors trained on a large-scale corpus to explore semantic relationships such as analogy, subordination, and comparison [26,27]. In our corpus, the context of a diagnostic term could always be related diseases, symptoms, and drugs. Therefore, the embeddings of similar terms or synonyms with similar context will be close to each other in the space after the training process.

Our algorithm can correctly identify over 80% of the synonyms by just searching from the top 10 candidates of a certain medical term. We further summarize these synonyms into three classes: (1) Colloquial expressions; for example, consumers say “having loose bowels” (“拉肚子”) rather than “diarrhea” (“腹泻”) and “zits” (“青春痘”) rather than “acne” (“粉刺”). (2) Typos; for example, consumers always misspell “腹泻” (“diarrhea”) as “腹泄” and “黄疸” (“jaundice”) as “黄胆.” (3) The symptoms or findings from traditional Chinese medicine; for example, Chinese medicine refers to “stomachache” (“胃痛”) as “epigastric pain” (“胃脘痛”). Besides the “typo” synonyms, other two classes of synonyms do not necessarily share common characters with each other or source terms. Therefore, simple character-based matching approaches such as n-gram and edit-distance do not help in these cases. Semantic pattern-based algorithms depend less on exact common characters; however, consumers in online communities and social media express themselves in a more casual way, and we may not be able to create and maintain a comprehensive semantic pattern list to capture all the variations and diversities. Our method can fill in such a language gap and effectively expand the synonyms and consumer health vocabularies. We validated the effectiveness of our approach with the 180-Mb Tianya corpus and the 2-Gb Haodf corpus. The results indicate that the larger the corpus, the better the learning.

One limitation of our approach is that we cannot handle a case when a new professional term is needed in the vocabulary, especially the newly formulated professional term, for example, the extracellular space and the interstitial system [31]. This is because we adopted a matching-based framework. However, it is not difficult to extend the current algorithm to gain such capability. For example, we can normalize the similarity between a specific consumer term to all professional terms in the dictionary and thus make these similarities a probability distribution. Thereafter, we can use entropy for this distribution to determine whether we need a new professional term. A high entropy indicates that the consumer term is not really similar to any of the existing professional terms, and thus, a new term may be needed.

For the first time in the Chinese medical terminology field, this study verified the effectiveness of word semantic representations and their potential for linking narrative consumer terms to clinical terms. This approach can discover consumer expressions such as spelling errors and nonstandard abbreviations, which are usually missed in the traditional consumer health vocabularies, and enrich the consumer health vocabularies to meet consumer requirements for information retrieval. The candidate consumer term list automatically generated by our model can be employed as an important reference for professionals to discover synonyms in a more efficient way.

Acknowledgments

All coauthors contributed equally. This study is partly sponsored by the National Natural Science Foundation of China (Grants: #81471756 #81771937). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

References

1. China Institute of Popular Science. An Analysis Report on the Search Demand Behavior of Chinese Netizens in Popular Science in 2015. Science and Technology Review 2016;34(12):29-33 [FREE Full text]
2. 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]
3. Holtz B, Vasold K, Cotten S, Mackert M, Zhang M. Health Care Provider Perceptions of Consumer-Grade Devices and Apps for Tracking Health: A Pilot Study. JMIR Mhealth Uhealth 2019 Jan 22;7(1):e9929 [FREE Full text] [doi: 10.2196/mhealth.9929] [Medline: 30668515]
4. Lei Y, Han H, Yuan F, Javeed A, Zhao Y. The brain interstitial system: Anatomy, modeling, in vivo measurement, and applications. Prog Neurobiol 2017 Oct;157:230-246 [FREE Full text] [doi: 10.1016/j.pneurobio.2015.12.007] [Medline: 26837044]

5. Yangarber R, Lin W, Grishman R. Unsupervised learning of generalized names. In: COLING.: Association for Computational Linguistics; 2002 Presented at: Proceedings of the 19th international conference on Computational linguistics; 2002; Taipei, Taiwan. [doi: 10.3115/1072228.1072382]
6. Slaughter L, Ruland C, Rotegård AK. Mapping cancer patients' symptoms to UMLS concepts. AMIA Annu Symp Proc 2005:699-703 [FREE Full text] [Medline: 16779130]
7. Vydiswaran VGV, Mei Q, Hanauer DA, Zheng K. Mining consumer health vocabulary from community-generated text. AMIA Annu Symp Proc 2014;2014:1150-1159 [FREE Full text] [Medline: 25954426]
8. Keselman A, Smith CA, Divita G, Kim H, Browne AC, Leroy G, et al. Consumer health concepts that do not map to the UMLS: where do they fit? J Am Med Inform Assoc 2008;15(4):496-505 [FREE Full text] [doi: 10.1197/jamia.M2599] [Medline: 18436906]
9. Doing-Harris KM, Zeng-Treitler Q. Computer-assisted update of a consumer health vocabulary through mining of social network data. J Med Internet Res 2011 May 17;13(2):e37 [FREE Full text] [doi: 10.2196/jmir.1636] [Medline: 21586386]
10. Jiang L, Yang C. Using Co-occurrence Analysis to Expand Consumer Health Vocabularies from Social Media Data. 2013 Sep 11 Presented at: 2013 IEEE International Conference on Healthcare Informatics; 2013; Philadelphia, PA p. 74-81. [doi: 10.1109/ICHI.2013.16]
11. Zeng QT, Tse T, Divita G, Keselman A, Crowell J, Browne AC, et al. Term identification methods for consumer health vocabulary development. J Med Internet Res 2007 Feb 28;9(1):e4 [FREE Full text] [doi: 10.2196/jmir.9.1.e4] [Medline: 17478413]
12. Wang Y, Tang J, Vydiswaran V, Zheng K, Xu H, Mei Q. Matching Consumer Health Vocabulary with Professional Medical Terms Through Concept Embedding. : American Medical Informatics Association; 2017 Presented at: AMIA 2017 Annual Symposium; 2017; Washington, DC.
13. Jelinek F, Mercer R, Roukos S. Classifying words for improved statistical language models. : IEEE; 1990 Presented at: International Conference on Acoustics, Speech, and Signal Processing; 1990; Albuquerque, NM p. 621-624. [doi: 10.1109/ICASSP.1990.115789]
14. Blondel V, Gajardo A, Heymans M, Senellart P, Van Dooren PM. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. SIAM review 2004;46(4):A. [doi: 10.2307/20453570]
15. Martino B. Semantic web services discovery based on structural ontology matching. IJWGS 2009;5(1):46-65. [doi: 10.1504/IJWGS.2009.023868]
16. Hill F, Cho K, Jean S, Devin C, Bengio Y. arXiv. 2014. Embedding word similarity with neural machine translation URL: https://arxiv.org/abs/1412.6448 [accessed 2019-05-13] [WebCite Cache ID 78LAJpW1Q]
17. Henriksson A, Conway M, Duneld M, Chapman W. Identifying synonymy between SNOMED clinical terms of varying length using distributional analysis of electronic health record. 2013 Presented at: AMIA Annual Symposium Proceedings; 2013; Washington, DC.
18. Henriksson A, Moen H, Skeppstedt M, Daudaravičius V, Duneld M. Synonym extraction and abbreviation expansion with ensembles of semantic spaces. J Biomed Semantics 2014 Feb 05;5(1):6 [FREE Full text] [doi: 10.1186/2041-1480-5-6] [Medline: 24499679]
19. He Z, Chen Z, Oh S, Hou J, Bian J. Enriching consumer health vocabulary through mining a social Q&A site: A similarity-based approach. J Biomed Inform 2017 Dec;69:75-85 [FREE Full text] [doi: 10.1016/j.jbi.2017.03.016] [Medline: 28359728]
20. Elhadad N, Zhang S, Driscoll P, Brody S. Characterizing the sublanguage of online breast cancer forums for medications, symptoms, and emotions. AMIA Annu Symp Proc 2014;2014:516-525 [FREE Full text] [Medline: 25954356]
21. Marshall P. Bridging the terminology gap between health care professionals and patients with the Consumer Health Terminology (CHT). 2000 Presented at: Proceedings of the AMIA Symposium; 2000; Los Angeles, CA.
22. Patrick T, Monga H, Sievert M, Houston Hall J, Longo D. Evaluation of controlled vocabulary resources for development of a consumer entry vocabulary for diabetes. J Med Internet Res 2001;3(3):E24 [FREE Full text] [doi: 10.2196/jmir.3.3.e24] [Medline: 11720966]
23. Zeng Q, Tse T, Crowell J, Divita G, Roth L, Browne A. Identifying consumer-friendly display (CFD) names for health concepts. 2005 Presented at: AMIA annual symposium proceedings; 2005; Washington, DC.
24. Jiang L, Yang C, Li J. Discovering Consumer Health Expressions from Consumer-Contributed Content. In: Social Computing, Behavioral-Cultural Modeling and Prediction. Berlin, Heidelberg: Springer; 2013:164-174.
25. Maosong S, Xinxiong C, Kaixu Z, Zhipeng G, Zhiyuan L. THULAC: An efficient lexical analyzer for Chinese [in Chinese]. 2016. URL: http://thulac.thunlp.org/ [accessed 2019-05-11] [WebCite Cache ID 78HtNhzB9]
26. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform 2018 Nov;87:12-20. [doi: 10.1016/j.jbi.2018.09.008] [Medline: 30217670]
27. Mikolov T, Chen K, Corrado G, Dean J. arXiv. 2013. Efficient estimation of word representations in vector space URL: https://arxiv.org/abs/1301.3781 [accessed 2019-05-13] [WebCite Cache ID 78LAlyreV]
28. Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation. : Association for Computational Linguistics; 2014 Presented at: EMNLP; 2014; Doha, Qatar. [doi: 10.3115/v1/D14-1162]

29. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics 2017;5:135-146. [doi: 10.1162/tacl_a_00051]
30. Myers D, McGuffee J. Choosing Scrapy. J Comput Sci Coll 2015;31(1):83-89.
31. Aibo W, Rui W, Dehua C, Xinrui H, Lan Y, Huipo L, et al. The Drainage of Interstitial Fluid in the Deep Brain is Controlled by the Integrity of Myelination. Aging and disease 2018. [doi: 10.14336/AD.2018.1206]

Abbreviations

GloVe: Global Vectors
ICD: International Classification of Diseases
MRR: mean reciprocal rank
SNOMED: Systematized Nomenclature of Medicine
THULAC: Tsinghua University - Lexical Analyzer for Chinese
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 06.11.18; peer-reviewed by J Bian, Y Wang, V Vydiswaran, PH Lin; comments to author 12.12.18; revised version received 19.03.19; accepted 05.04.19; published 23.05.19.

Please cite as:
Gu G, Zhang X, Zhu X, Jian Z, Chen K, Wen D, Gao L, Zhang S, Wang F, Ma H, Lei J
Development of a Consumer Health Vocabulary by Mining Health Forum Texts Based on Word Embedding: Semiautomatic Approach
JMIR Med Inform 2019;7(2):e12704
URL: http://medinform.jmir.org/2019/2/e12704/
doi:10.2196/12704
PMID:31124461

Original Paper

Use of Telemedicine to Screen Patients in the Emergency Department: Matched Cohort Study Evaluating Efficiency and Patient Safety of Telemedicine

Nicholas James Rademacher[1], MD; Gai Cole[2], MHA, MBA, DrPH; Kevin J Psoter[3], PhD; Gabor Kelen[4], MD; Jamie Wei Zhi Fan[5], MHA; Dennis Gordon[5]; Junaid Razzak[4], MD, PhD, FACEP

[1]Department of Emergency Medicine, University of Michigan, Ann Arbor, MI, United States

[2]The Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, MD, United States

[3]Department of Pediatrics, The Johns Hopkins School of Medicine, Baltimore, MD, United States

[4]Department of Emergency Medicine, The Johns Hopkins School of Medicine, Baltimore, MD, United States

[5]Center for Population Health IT, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Junaid Razzak, MD, PhD, FACEP
Department of Emergency Medicine
The Johns Hopkins School of Medicine
5801 Smith Ave
Ste 3220
Baltimore, MD, 21209
United States
Phone: 1 4107356400
Fax: 1 4107356425
Email: junaid.razzak@jhu.edu

Abstract

Background: Early efforts to incorporate telemedicine into Emergency Medicine focused on connecting remote treatment clinics to larger emergency departments (EDs) and providing remote consultation services to EDs with limited resources. Owing to continued ED overcrowding, some EDs have used telemedicine to increase the number of providers during surges of patient visits and offer scheduled “home” face-to-face, on-screen encounters. In this study, we used remote on-screen telemedicine providers in the “screening-in-triage” role.

Objective: This study aimed to compare the efficiency and patient safety of in-person screening and telescreening.

Methods: This cohort study, matched for days and proximate hours, compared the performance of real-time remote telescreening and in-person screening at a single urban academic ED over 22 weeks in the spring and summer of 2016. The study involved 337 standard screening hours and 315 telescreening hours. The primary outcome measure was patients screened per hour. Additional outcomes were rates of patients who left without being seen, rates of analgesia ordered by the screener, and proportion of patients with chest pain receiving or prescribed a standard set of tests and medications.

Results: In-person screeners evaluated 1933 patients over 337 hours (5.7 patients per hour), whereas telescreeners evaluated 1497 patients over 315 hours (4.9 patients per hour; difference=0.8; 95% CI 0.5-1.2). Split analysis revealed that for the final 3 weeks of the evaluation, the patient-per-hour rate differential was neither clinically relevant nor statistically discernable (difference=0.2; 95% CI –0.7 to 1.2). There were fewer patients who left without being seen during in-person screening than during telescreening (2.6% vs 3.8%; difference=–1.2; 95% CI –2.4 to 0.0). However, compared to prior year-, date-, and time-matched data on weekdays from 1 am to 3 am, a period previously void of provider screening, telescreening decreased the rate of patients LWBS from 25.1% to 4.5% (difference=20.7%; 95% CI 10.1-31.2). Analgesia was ordered more frequently by telescreeners than by in-person screeners (51.2% vs 31.6%; difference=19.6%; 95% CI 12.1-27.1). There was no difference in standard care received by patients with chest pain between telescreening and in-person screening (29.4% vs 22.4%; difference=7.0%; 95% CI –3.4 to 17.4).

Conclusions: Although the efficiency of telescreening, as measured by the rate of patients seen per hour, was lower early in the study period, telescreening achieved the same level of efficiency as in-person screening by the end of the pilot study. Adding

telescreening during 1-3 am on weekdays dramatically decreased the number of patients who left without being seen compared to historic data. Telescreening was an effective and safe way for this ED to expand the hours in which patients were screened by a health care provider in triage.

(JMIR Med Inform 2019;7(2):e11233) doi:10.2196/11233

KEYWORDS

telemedicine; telehealth; screening; triage; emergency medicine; emergency health services; emergency medical service; left without being seen; emergency room; emergency department; tele-medicine

Introduction

Over nearly three decades, the volume of emergency department (ED) visits has steadily grown [1-3]. The inability to slow down utilization has resulted in continued ED crowding and considerable delays prior to ED evaluation and treatment with the associated adverse effects on patient outcomes [4-9].

One solution to expedite emergency care in the face of growing demand is to place a provider proximate to triage evaluation. Apart from fulfilling requirements of the Emergency Medicine Treatment and Labor Act, early provider evaluation assists with (1) identification of patients who may be critically ill but not yet classified as such by the triage nurse, (2) identification of patients who can be quickly discharged, (3) early initiation of treatment, and (4) reduction in the number of patients who left without being seen (LWBS) by a qualified medical provider (typically a physician, nurse practitioner, or physician assistant). ED screening is particularly important for patient safety during times of surge and during hours with reduced staffing, when patient volume and crowding outpace an ED's ability to provide prompt evaluation [10,11].

The application of telemedicine to screening ("telescreening") is one additional solution to address the increased ED demands. Through a real-time audio-visual interface between patients and remote care providers, telescreening optimizes providers' time, potentially minimizes expensive staffing requirement, and may increase the pool of providers available during undesirable times due to the ability to provide care from home or other remote settings.

Telemedicine in the ED has traditionally been used to connect minor treatment clinics to larger EDs and to facilitate specialty consultation [12-19]. Additional applications, such as adding remote providers during times of patient volume surge [20], and direct-to-consumer home visits [21] have recently shown to be effective for and popular among patients.

In April 2016, our ED initiated a pilot telescreening program to expand the hours in which screening by a provider in triage took place. The objective of this evaluation was to compare the efficiency and patient safety metrics between ED remote real-time telescreening and in-person screening encounters.

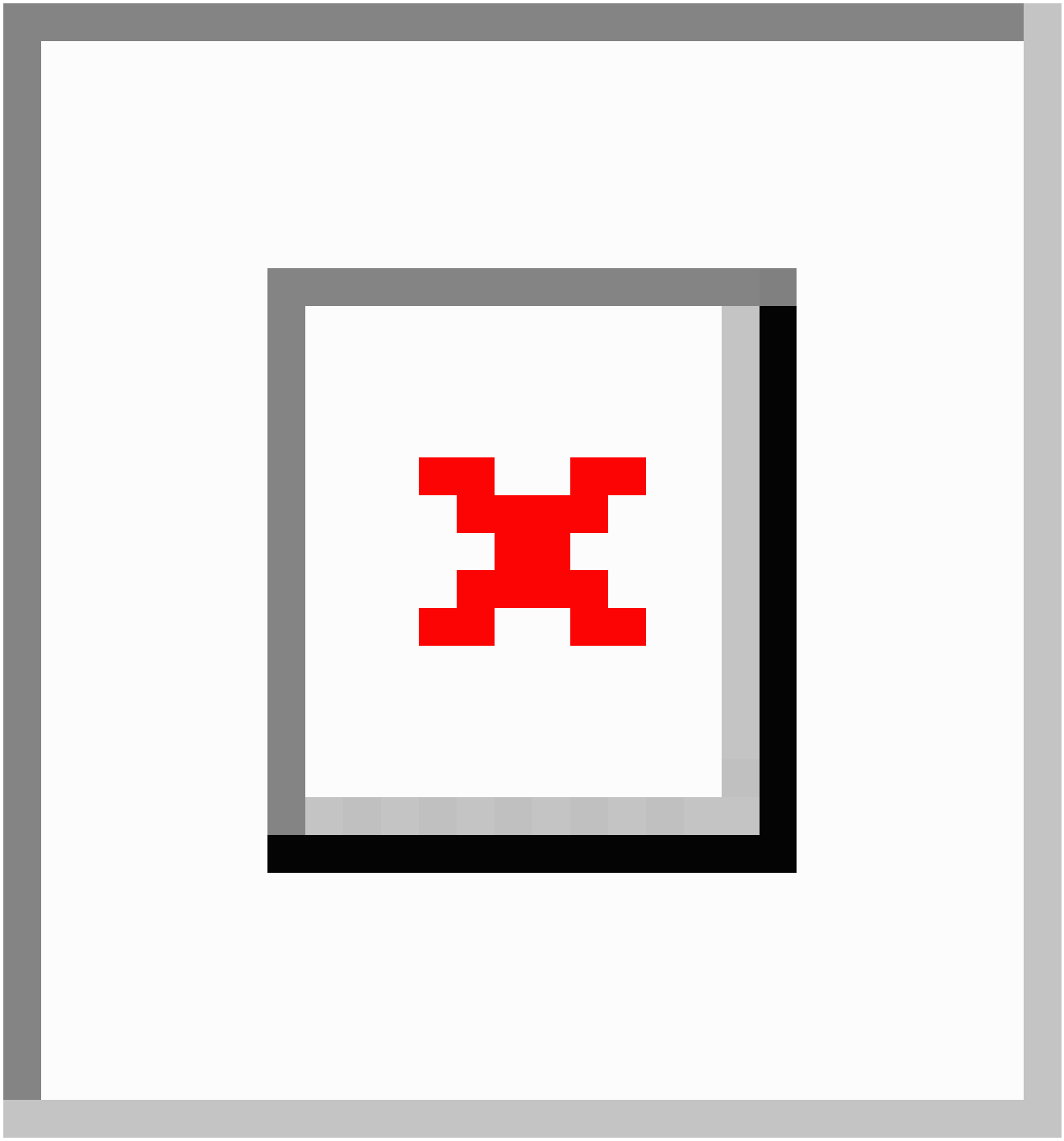
Methods

We conducted a matched cohort study to compare the performance of remote real-time telescreening (hereafter referred to as telescreening) and in-person screening at a single urban academic ED with 67,620 adult patient visits in 2016. This ED is part of a quaternary care, 900-bed, academic medical center serving a mix of predominantly inner-city, suburban, and international patients. At the time, triage was performed by registered nurses using the Emergency Severity Index (ESI) [22]. Patients with ESI levels 1 and 2 were triaged directly to an ED bed, including hallway beds, for a full evaluation and bypassed ED provider screening. Patients with ESI levels 3, 4, and 5 were briefly evaluated or "screened" by a physician, nurse practitioner, or physician assistant (Figure 1).

This analysis was conducted from April to August 2016. These months were chosen because telescreening was initiated in April 2016 with a process similar to that used in in-person screening. In early September 2016, the screening process changed, causing the screening blocks to no longer serve as suitable controls for the telescreening periods. Telemedicine was generally offered during this period, from Tuesday to Friday, 1-3 am, and on Saturdays and Sundays, 7-10 am. Additional hours were included depending on provider availability and need. This period was not previously covered by any screening activity. Telescreening was contiguous with on-site screening, that is, it naturally followed weekday on-site screening and was continued on the weekends. Although these blocks of time were not officially classified as times of "surge," patient volume typically outpaces the capacity of the local ED system, resulting in waiting for all, but the most critically ill patients, similar to surge situations.

All adult patients during these times triaged to ESI levels 3 through 5 were offered telescreening. Since this was not an established practice, written informed consent for telescreening was obtained by certified nursing assistants (CNAs). Patients who did not consent were resigned to the usual protocol available at the time. Non-English–speaking patients and those deemed devoid of mental capacity, including those with an altered mental status, were not eligible to receive telescreening and were relegated to usual care. After registration and triage, appropriate patients proceeded to a screening (tele- or in-person screening) evaluation, which aimed at attending to the patients within 30 minutes of their arrival. The Institutional Review Board exempted this project based on its quality improvement classification. We did not charge professional fees for telemedicine screening.

Figure 1. Process flow for patient intake to the emergency department. RN: registered nurse, ESI: Emergency Severity Index; EKG: electrocardiogram.



Five providers from our institution—three physicians and two physician assistants—who were accustomed to in-person screening in the same ED, received technical training in the telescreening procedures as well as mock standardized patient encounters. These training sessions consisted of technical training on how to use the Clearsteth stethoscope (GlobalMedia Group, LLC, Scottsdale, AZ) and run the accompanying software (Polycom, San Jose, CA). A set of five live patient models were designed to allow providers to practice using the telemedicine equipment as well as write notes and enter orders during the exam. Training sessions were the same for each type of provider and were supervised by one of the authors of this manuscript (JR or NR) for proficiency. Screening was performed by 27 providers (physicians, nurse practitioners, or physician assistants).

We used the Globalmed (GlobalMedia Group, LLC) Clinical Access Station. The customized device utilized two-way high-quality audio and high-resolution cameras attached to a Polycom codec, with pan and zoom controlled by the telescreening provider or the CNAs trained to set up and facilitate use of the Clinical Access Station. Our Clinical Access Station had a personal computer, two monitors, a fiber optic light source, and three peripherals: a high-resolution hand-held camera, a fiber optic otoscope, and a stethoscope. Multimedia Appendix 1 shows our Clinical Access Station. The CNAs were trained in the use and placement of the peripherals to optimize information (images and auscultation) transmitted to the remote

health care providers. The remote health care provider connected with the onsite system using a dual-monitor computer via Polycom for video and Clearsteth for auscultation. Multimedia Appendix 2 shows the interface from the perspective of both the patient and the remote screener.

Software was available for use on an institutional license and provided high-definition, HIPAA (Health Insurance Portability and Accountability Act)-secure, two-way communication between the remote health care provider at their homes and the exam room. All orders and telemedicine screening notes were placed in the existing electronic medical record framework (EPIC, Verona, WI).

Because the telescreening time represented incremental coverage hours, there were no time-matched historical control time periods; therefore, we matched these time periods to equivalent proximate hours to control for day of the week and ED volume. For example, if telescreening was conducted from 1 to 3 am, we matched the time with the most proximate in-person screening, which was 11 pm to 1 am. In addition, to evaluate the effect of telescreening on rates of LWBS as compared to no screening, each telescreening and in-person screening hour was matched to the corresponding day and time in the preceding year (2015). Information on individuals entering the ED during these hours was then abstracted from our EMR; this information included basic demographics, medications ordered, chief complaint, and final disposition.

Scheduled telemedicine shifts that could not be fulfilled for technical or assistant staff's shortfalls were excluded from analysis. However, given that matching was done based on the expected telescreening shifts, the asymmetrical number of hours between the groups can be accounted for by canceled telescreening hours. The matching approach was utilized to control for day of the week and ED volume.

The primary outcome of interest was the number of patients screened per hour. Secondary outcomes of interest included LWBS rates, patients receiving analgesia (ibuprofen, acetaminophen, ketorolac, oxycodone, hydrocodone, hydromorphone, morphine, tramadol, naproxen, dicyclomine, codeine, diclofenac, fentanyl, meloxicam, or methadone), and the ordering of a chest pain bundle.

Patients who LWBS included only patients physically presenting during the evaluation period. For example, if screening occurred from 1 am to 3 am, only patients registering during those times were evaluated to determine the LWBS rates. Rates of analgesia administration were compared as a quality metric. Screeners initiated plans of care; therefore, the time for which the patients were in the waiting room was used to obtain results of laboratory and radiological tests. Similarly, screeners worked toward achieving patient comfort by ordering oral analgesia while the patients waited for a formal evaluation. As pain is one of the most common complaints of patients presenting to the ED, our ability to provide safe and effective palliation is a quality metric. Given that screeners provide basic oral analgesia to those returning to the waiting room to complete their care, it is important that telescreeners provide this care at a similar rate.

As chest pain is a common chief complaint with a high-risk profile, initiation of evaluation of patients with this presenting complaint was used to compare safety and quality between the two modes of screening. The chest pain bundle, which was considered to represent standard orders by health care providers on the research team prior to analysis, included the following items: complete blood count, comprehensive or basic metabolic panel, troponin I levels, electrocardiogram, chest x-ray, and aspirin. If components of the bundle were performed prior to evaluation by the screener, that component was counted as successfully being provided. In some cases, ordering providers placed their orders through an order set designed for patients presenting with chest pain. Manual review of charts of patients presenting with chest pain was performed by one physician-author (NR).

Immediately after a telescreening encounter, patients were given a six-question Likert scale questionnaire to complete. The focus of the questionnaire was patient satisfaction. As such, the questionnaire was not validated, and the response rate is unknown. The results from this survey are shown in Multimedia Appendix 3.

Demographic and clinical characteristics of individuals who received telescreening and in-person screening were compared using chi-square or the Fisher exact test for categorical variables and the Student *t* test for continuous variables, with unequal variances. For the subgroup of individuals presenting with chest pain and screened, receipt of a chest pain bundle was compared between the two screening modes. The mean number of individuals screened per hour, rates of LWBS, and rates of analgesia ordered were compared between telescreening and in-person screening hours in a similar manner. We compared the screening hours and proportions of patients who LWBS between 2016 and 2015 for both telescreening and in-person screening hours. The 95% CI was considered significant, and all analyses were conducted using STATA Version 14.1 (StataCorp, College Station, TX).

This trial is reported in accordance with CONSORT-EHEALTH [23].

Results

From April to August 2016, telescreening was performed for 315 hours and in-person screening was performed for 337 proximate matched hours. During these hours, a total of 3430 individuals were screened, of which 1497 (43.64%) were telescreened and 1933 (56.36%) were screened in person. Demographics, chief complaints, and ESI level of patients who underwent telescreening were comparable to those receiving in-person screening (Table 1). Compared to patients screened in person (46.19%), a greater proportion of individuals telescreened were male (52.22%). Distribution of discharge and disposition status also differed between patients screened in person and those who were telescreened: 65.29% were discharged in the screening group (n=1262) compared to 64.19% in the telescreening group (n=961). A higher proportion of patients presenting during telescreening hours had ESI levels of 3-5 (1904/2341; 81.33%) as compared to the proportion of patients presenting during in-person screening hours (2235/2869;

77.90%; difference=3.43%; 95% CI 1.24-5.62). The total number of telescreened patients and patients screened in person was less than that of patients with ESI levels 3-5 due to the exclusion criteria applied and patient refusal for telescreening. In addition, 24.92% of the hours met our goal door-to-provider time of less than 30 minutes (77/309) for telescreening as compared to 33.23% (111/334) for in-person screening (difference=8.31%; 95% CI 1.33-15.29). The five providers performed 695, 631, 115, 32, and 24 telescreening encounters.

On an average, 4.87 patients received telescreening per hour compared to 5.75 patients in the in-person screening group (difference=–0.87; 95% CI –1.23 to –0.51). Although a statistically significant difference was observed in the number of patients evaluated per hour in the first 3 weeks following implementation of telescreening (5.88 for in-person screening vs 4.40 for telescreening; difference=1.48; 95% CI 0.64-2.33), no differences were observed in the final 3 weeks of the study (5.52 for in-person screening vs 5.49 for telescreening; mean difference=0.03; 95% CI –0.89 to 0.94; Figure 2).

The LWBS rates were higher in the telescreening group than in the in-person screening group (3.8% vs 2.6%; difference=1.2; 95% CI 0.1-1.9). However, while the LWBS rates were not different during periods of in-person screening in 2015 and 2016 (difference: 0.5; 95% CI: –0.7 to 1.6), the LWBS rates in the telescreened hours in 2016 were significantly lower than those in the matched 2015 hours (3.8% vs 8.5%; difference=–4.7; 95% CI –8.6 to –1.0). The difference from 2015 to 2016 was most pronounced in the subgroup receiving telescreening from 1 am to 3 am on weekdays. For this subgroup, the LWBS rate declined from 25.1% to 4.5% (difference=20.7; 95% CI 10.1-31.2; Figure 3).

Table 1. Demographic characteristics of patients in the in-person screening and telescreening groups.

Characteristic	In-person screening (n=1933)[a]	Telescreening (n=1497)[a]
Demographics		
Age (years), mean (SD)	43.44 (16.8)	43.09 (15.7)
Gender, n (%)		
Male	893 (46.19)	782 (52.22)
Female	1040 (53.80)	715 (47.76)
Race, n (%)		
White	443 (22.91)	318 (21.24)
Black	1290 (66.74)	1071 (71.54)
Asian	34 (1.76)	13 (0.87)
Other	145 (7.50)	81 (5.41)
Ethnicity, n (%)		
Hispanic	95 (4.91)	59 (3.94)
Non-Hispanic	1796 (92.91)	1409 (94.12)
Presentation, n (%)		
Disposition		
Admitted	227 (11.74)	124 (8.28)
Discharged	1262 (65.29)	961 (64.19)
Left against medical advice	18 (0.93)	22 (1.47)
Transferred	2 (0.10)	0 (0.0)
Eloped	21 (1.09)	29 (1.94)
Screen and leave	300 (15.52)	288 (19.24)
Other	103 (5.33)	73 (4.88)
Emergency severity index level		
3	1512 (78.22)	1127 (75.28)
4	387 (20.00)	347 (23.18)
5	20 (1.04)	19 (1.27)
Chief complaints		
Abdominal pain	201 (10.40)	162 (10.82)
Chest pain	152 (7.86)	126 (8.42)
Other	1580 (81.7)	1209 (80.76)

[a]Categories may not sum up to the total due to missing data.

On an average, 51.2% of telescreened patients received analgesia, compared to 31.6% of those screened in person (difference=19.6; 95% CI 12.1-27.1). Two of the providers that completed 277 of the 315 (87.9%) telescreening hours did not show a difference in the rates of ordering analgesia (Table 2). However, they ordered more analgesia per patient encounter than the group that performed the in-person screening. Although screeners ordered analgesia for 32% of the patients, the two primary telescreeners ordered 51% and 45% of that proportion when they acted as in-person screeners in a small sample of shifts prior to the implementation of telescreening.

In the subgroup of patients presenting with undifferentiated chest pain, 22.4% (34/152) who received in-person screening and 29.4% (37/126) who received telescreening had all the components of the chest pain bundle provided to them or ordered after their screening encounter (difference=–7.00, 95% CI –17.35 to 3.35). Analysis of individual components of the chest pain bundle revealed that aspirin administration was the only item with a statistically significant difference between the screening methods. The fact that many of the orders were likely placed as part of an order set explains some of the congruency in ordering practice. In addition, 37.3% (47/126) of the telescreened patients received aspirin and 25.0% (38/152) of the patients screened in person received aspirin (difference=12.30; 95% CI 1.41-23.19; Table 3). The results of our patient satisfaction questionnaire can be found in Multimedia Appendix 3.

Figure 2. Weekly trends in patients screened per hour by in-person provider screening and remote telescreening. S: in-person screening; TS: telescreening.

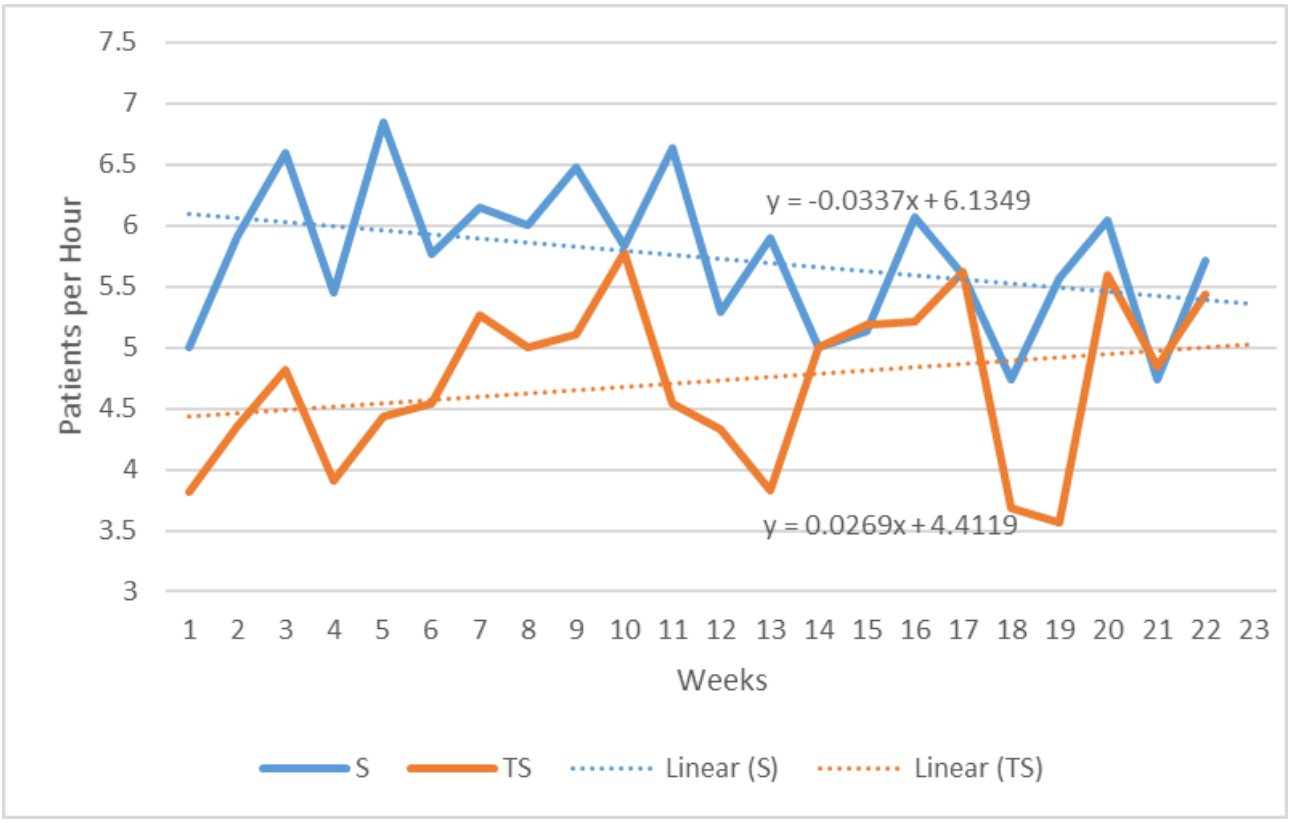


Figure 3. Comparison of screening modes and times between 2015 and 2016. The graph shows a comparison of patients who received in-person screening, matched in-person screening in 2015 and telescreening in 2016, and in-person provider screening (2015) and telescreening from 1 am to 3 am only (2016). LWBS: left without being seen.

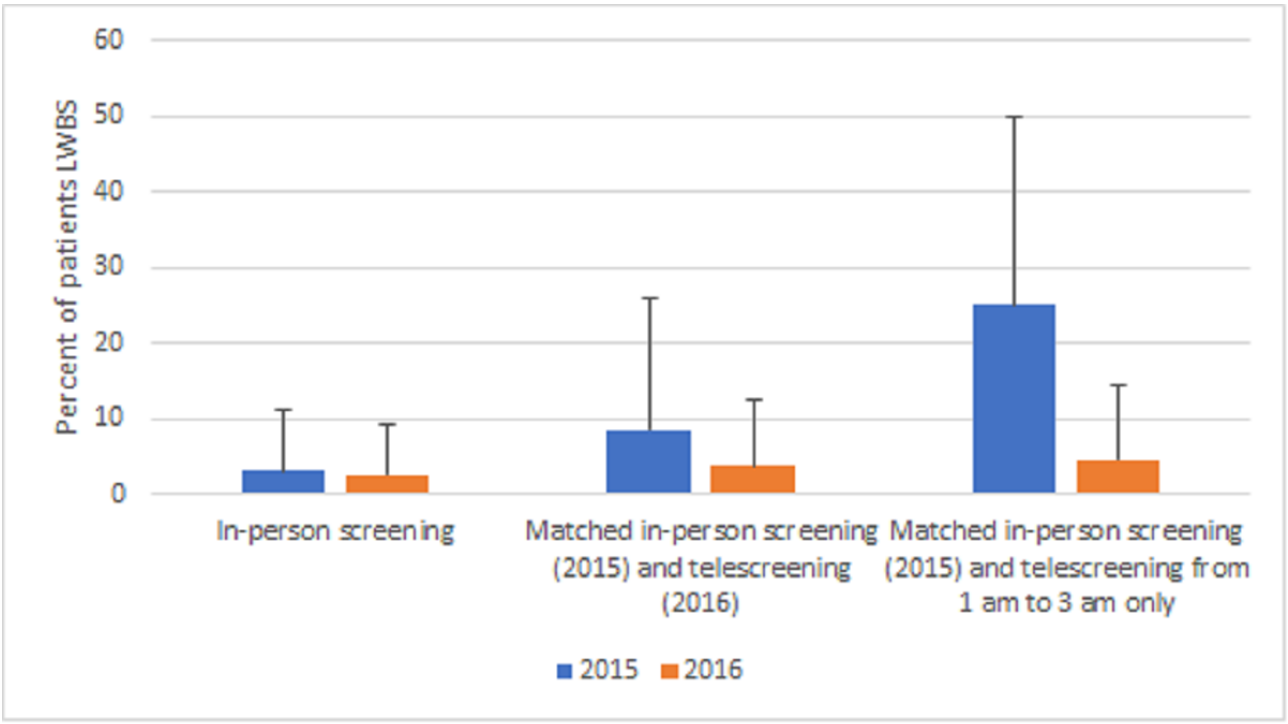


Table 2. Rate of analgesia orders by the two primary telescreeners according to the screening method.

Primary telescreeners	Number of hours	Percentage per hour, mean (SD)		Absolute effect size (95% CI)	Relative effect size (95% CI)
		In-person screening	Telescreening		
Overall	40	48 (12)	48 (16)	–0.24 (–11.43 to 10.96)	1.00 (0.82 to 1.22)
Provider 1	26	51 (13)	44 (14)	–7.61 (–21.70 to 6.48)	0.85 (0.62 to 1.18)
Provider 2	14	45 (10)	57 (18)	11.78 (–7.19 to 30.75)	1.26 (0.91 to 1.75)

Table 3. Comparison of the order rates of the chest pain bundle and its components among patients in the telescreening and in-person screening groups.

Orders	In-person screening (n=152), n (%)	Telescreening (n=126), n (%)	Absolute effect size (95% CI)	Relative effect size (95% CI)
Full chest pain bundle	34 (22.4)	37 (29.4)	7.0 (–5.1 to 19.1)	31.3 (–19.9 to 115.6)
Complete blood count	136 (89.5)	115 (91.3)	1.8 (–20.7 to 24.3)	2.0 (–21.1 to 31.7)
Metabolic panel	135 (88.8)	113 (89.7)	0.9 (–21.5 to 23.2)	1 (–22.1 to 30.61)
Troponin I level	116 (76.3)	104 (82.5)	6.2 (–14.9 to 27.3)	8.1 (–17.8 to 42.2)
Electrocardiography	140 (92.1)	118 (93.7)	1.6 (–21.2 to 24.3)	1.7 (–21.1 to 30.8)
Chest radiograph	119 (77.6)	105 (83.3)	5.0 (–16.2 to 26.3)	6.44 (–18.9 to 39.6)
Aspirin administration	38 (25.0)	47 (37.3)	12.3 (0 to 25.6)	49.2 (–4.8 to 135.2)

Discussion

Principal Findings

This matched cohort study of our pilot telescreening program shows that telescreening can be efficiently and safely used for screening patients presenting to the ED. Although telescreening was initially less efficient than in-person screening, by the final 3 weeks of our analysis, telescreening had achieved efficiency levels similar to those of in-person screening. We included full data without a phase-in period to obtain an estimate of how long it may take for the telemedicine program to reach in-person efficiency.

Importantly, after implementation of telescreening, the LWBS rate dropped from 25.1% (in the 2015 matched weekday 1-3 am time slots) to 4.45%. Although some of these patients simply transitioned from the LWBS category to the “screened and left” category, similar to what has been reported in a recent survey [3], they were evaluated by a health care provider and often had imaging or laboratory tests drawn prior to leaving the ED. One would expect that the population that is screened and leaves is at a lower risk of adverse health outcomes than the population that simply LWBS. This issue and the health outcomes of screening, in general, are research questions worth pursuing.

The screeners’ rates of ordering analgesia were skewed by the individual practice patterns of two telescreeners who worked a majority of the telescreening hours. Additional research should be performed on this topic, especially on the breakdown of analgesic agents.

Except for aspirin administration, the chest pain bundle was completed at a similar rate between the two screening modes. This outcome suggests that telescreeners are able to set a care plan in motion for even high-risk chief complaints. However, as patients with ESI levels 1 and 2 bypassed screening and went directly to the patient care areas, the patients included in this analysis were considered to be only at moderate risk, at best, by the triage nurse. This is a necessary safeguard for a telescreening program, and this analysis does not suggest that those with undifferentiated high-risk chief complaints can safely be cared for when their vital signs or triage assessment considers them to be in danger.

An important area of further investigation is emergency patient and medicine provider satisfaction with telemedicine. Our data (Multimedia Appendix 3) broadly suggest that patients were happy with their experience with telemedicine. Few patients refused telescreening or were unsatisfied with the services; this finding is similar to those of other studies with more formal patient satisfaction surveys [21,24]

We hypothesize that demographic differences between the two groups represent subtle differences in the populations cared for at slightly different hours of the day.

Future research should focus on the use of telemedicine in other areas of emergency medicine practice, such as observation medicine and management of patient boarding in the ED. Similarly, the cost-effectiveness of telemedicine in EDs, especially if a single remote emergency provider can provide coverage to several EDs, is an area of interest. At the policy level, reimbursement of telemedicine services in the ED and ability to practice telemedicine from outside local state jurisdictions remain areas of growing discussion. As a lack of reimbursement continues to prevent wider adoption of telemedicine in the ED, payers should select measurable criteria that would lead them to begin reimbursing the costs of telemedicine, so that we can move toward those metrics.

Limitations

First, our data are matched by date with adjacent, but not exact, time matching and therefore do not control for the time of the day variations in patient populations and presentation patterns. Second, the providers that conduct telescreening were a

relatively select group motivated to carry out telescreening for various reasons. Their general comfort with the use of technology, perhaps, played a role in their willingness to participate and be early adopters of telemedicine. This study cannot estimate the efficiency or quality impact of telemedicine when applied generally to all emergency medicine providers. Moreover, the small sample of providers performing telescreening makes data on items like analgesia and orders, which are part of a chest pain bundle, susceptible to skewing according to their practice patterns. Third, we observed a significant improvement in the efficiency of telescreening as providers became more comfortable with the use of technology, achieving a comparable level of efficiency between in-person and telescreening at week 20 of the program. We did not estimate the number of hours of telescreening per provider required to reach a comparable efficiency level. Similarly, we did not test for changes in the quality-of-care indicators with time.

Fourth, the rates of LWBS were higher during telescreening than during in-person screening hours, but this is likely due to the majority of telescreening hours occurring from 1 am to 3 am, a time period with no direct comparators. The comparator was a time period of in-person screening proximate to the telescreening shift.

Finally, the patient satisfaction questionnaire was not validated prior to administration, and the response rate was unknown, as the initial purpose was to allow patients to provide immediate feedback on this pilot program. In addition, we do not have similar satisfaction data for patients being screened, which may act as a control.

Conclusion

Telescreening is a new tool that can help EDs provide a safe and efficient alternative to in-person screening of patients while allowing a comparable level of efficiency, decreasing rates of LWBS (as compared to periods of time when screening did not previously take place), and providing greater flexibility in the provider's schedules.

Acknowledgments

This project was supported by the Johns Hopkins School of Medicine Biostatistics, Epidemiology and Data Management (BEAD) Core. No additional sponsors financially supported this work, assisted in its development, or reviewed it prior to publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

View of the telemedicine cart from the patient's perspective (left) and photo of the telemedicine system with labeled components (right).

[PNG File, 559KB - medinform_v7i2e11233_app1.png]

Multimedia Appendix 2

Photos from the provider's perspective (left) and certified nursing assistant facilitating a visit (right).

[PNG File, 800KB - medinform_v7i2e11233_app2.png]

Multimedia Appendix 3

Results of the brief patient satisfaction survey.

[PNG File, 288KB - medinform_v7i2e11233_app3.png]

References

1. Moore B, Stocks C, Owens P. HCUP Statistical Brief #227. 2017 Sep. Trends in Emergency Department Visits, 2006-2014 URL: https://www.hcup-us.ahrq.gov/reports/statbriefs/sb227-Emergency-Department-Visit-Trends.pdf [accessed 2018-10-29] [WebCite Cache ID 73X3rk837]
2. Institute of Medicine. Hospital-Based Emergency Care: At the Breaking Point. Washington, DC: The National Academies of Press; 2007:17-32.
3. Peterson S, Harbertson CA, Scheulen JJ, Kelen GD. Trends and Characterization of Academic Emergency Department Patient Visits: A Five-year Review. Acad Emerg Med 2019 Apr;26(4):410-419. [doi: 10.1111/acem.13550] [Medline: 30102817]
4. Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, Society for Academic Emergency Medicine, Emergency Department Crowding Task Force. The effect of emergency department crowding on clinically oriented outcomes. Acad Emerg Med 2009 Jan;16(1):1-10 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00295.x] [Medline: 19007346]

5. Carter EJ, Pouch SM, Larson EL. The relationship between emergency department crowding and patient outcomes: a systematic review. J Nurs Scholarsh 2014 Mar;46(2):106-115 [FREE Full text] [doi: 10.1111/jnu.12055] [Medline: 24354886]
6. Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. Ann Emerg Med 2008 Aug;52(2):126-136. [doi: 10.1016/j.annemergmed.2008.03.014] [Medline: 18433933]
7. Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang L, Han W, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med 2013 Jun;61(6):605-611.e6 [FREE Full text] [doi: 10.1016/j.annemergmed.2012.10.026] [Medline: 23218508]
8. Singer AJ, Thode HC, Viccellio P, Pines JM. The association between length of emergency department boarding and mortality. Acad Emerg Med 2011 Dec;18(12):1324-1329 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01236.x] [Medline: 22168198]
9. Jo S, Kim K, Lee JH, Rhee JE, Kim YJ, Suh GJ, et al. Emergency department crowding is associated with 28-day mortality in community-acquired pneumonia patients. J Infect 2012 Mar;64(3):268-275. [doi: 10.1016/j.jinf.2011.12.007] [Medline: 22227383]
10. Rowe BH, Guo X, Villa-Roel C, Schull M, Holroyd B, Bullard M, et al. The role of triage liaison physicians on mitigating overcrowding in emergency departments: a systematic review. Acad Emerg Med 2011 Feb;18(2):111-120 [FREE Full text] [doi: 10.1111/j.1553-2712.2010.00984.x] [Medline: 21314769]
11. Subash F, Dunn F, McNicholl B, Marlow J. Team triage improves emergency department efficiency. Emerg Med J 2004 Sep;21(5):542-544 [FREE Full text] [doi: 10.1136/emj.2002.003665] [Medline: 15333524]
12. Wechsler LR, Demaerschalk BM, Schwamm LH, Adeoye OM, Audebert HJ, Fanale CV, American Heart Association Stroke Council, Council on EpidemiologyPrevention, Council on Quality of Care Outcomes Research. Telemedicine Quality and Outcomes in Stroke: A Scientific Statement for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke 2017 Dec;48(1):e3-e25. [doi: 10.1161/STR.0000000000000114] [Medline: 27811332]
13. Meyer BC, Raman R, Ernstrom K, Tafreshi GM, Huisa B, Stemer AB, et al. Assessment of long-term outcomes for the STRokE DOC telemedicine trial. J Stroke Cerebrovasc Dis 2012 May;21(4):259-264 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2010.08.004] [Medline: 20851629]
14. Rubin MN, Demaerschalk BM. The use of telemedicine in the management of acute stroke. Neurosurg Focus 2014 Jan;36(1):E4. [doi: 10.3171/2013.11.FOCUS13428] [Medline: 24380481]
15. Demaerschalk BM, Levine SR. Telestroke: Solid support for virtual acute stroke care. Neurology 2016 Sep 27;87(13):1314-1315. [doi: 10.1212/WNL.0000000000003166] [Medline: 27566739]
16. Ward MM, Jaana M, Natafgi N. Systematic review of telemedicine applications in emergency rooms. Int J Med Inform 2015 Sep;84(9):601-616. [doi: 10.1016/j.ijmedinf.2015.05.009] [Medline: 26072326]
17. Langabeer JR, Gonzalez M, Alqusairi D, Champagne-Langabeer T, Jackson A, Mikhail J, et al. Telehealth-Enabled Emergency Medical Services Program Reduces Ambulance Transport to Urban Emergency Departments. West J Emerg Med 2016 Nov;17(6):713-720 [FREE Full text] [doi: 10.5811/westjem.2016.8.30660] [Medline: 27833678]
18. 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 Feb 03;6(2):e009221 [FREE Full text] [doi: 10.1136/bmjopen-2015-009221] [Medline: 26842270]
19. Hernandez M, Hojman N, Sadorra C, Dharmar M, Nesbitt TS, Litman R, et al. Pediatric Critical Care Telemedicine Program: A Single Institution Review. Telemed J E Health 2016 Jan;22(1):51-55. [doi: 10.1089/tmj.2015.0043] [Medline: 26203917]
20. Tolia V, Castillo E, Guss D. EDTITRATE (Emergency Department Telemedicine Initiative to Rapidly Accommodate in Times of Emergency). J Telemed Telecare 2017 May;23(4):484-488. [doi: 10.1177/1357633X16648535] [Medline: 27279469]
21. Powell RE, Stone D, Hollander JE. Patient and Health System Experience With Implementation of an Enterprise-Wide Telehealth Scheduled Video Visit Program: Mixed-Methods Study. JMIR Med Inform 2018 Feb 13;6(1):e10 [FREE Full text] [doi: 10.2196/medinform.8479] [Medline: 29439947]
22. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. Acad Emerg Med 2000 Mar;7(3):236-242. [Medline: 10730830]
23. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
24. Hanna GM, Fishman I, Edwards DA, Shen S, Kram C, Liu X, et al. Development and Patient Satisfaction of a New Telemedicine Service for Pain Management at Massachusetts General Hospital to the Island of Martha's Vineyard. Pain Med 2016 Dec;17(9):1658-1663. [doi: 10.1093/pm/pnw069] [Medline: 27121891]

Abbreviations

ED: emergency department
LWBS: left without being seen
ESI: Emergency Severity Index

CNA: certified nursing assistant

Edited by G Eysenbach; submitted 05.06.18; peer-reviewed by A Joshi, P Chai, A Davoudi; comments to author 05.07.18; revised version received 29.10.18; accepted 29.12.18; published 08.05.19.

Please cite as:
Rademacher NJ, Cole G, Psoter KJ, Kelen G, Fan JWZ, Gordon D, Razzak J
Use of Telemedicine to Screen Patients in the Emergency Department: Matched Cohort Study Evaluating Efficiency and Patient Safety of Telemedicine
JMIR Med Inform 2019;7(2):e11233
URL: http://medinform.jmir.org/2019/2/e11233/
doi:10.2196/11233
PMID:31066698

Original Paper

Use of Electronic Health Records to Develop and Implement a Silent Best Practice Alert Notification System for Patient Recruitment in Clinical Research: Quality Improvement Initiative

Connor Devoe[1], BS; Harriett Gabbidon[1], BA; Nina Schussler[1], BA; Lauren Cortese[1], BA; Emily Caplan[1], BA; Colin Gorman[1], BA; Kamal Jethwani[1,2,3], MPH, MD; Joseph Kvedar[1,2,3], MD; Stephen Agboola[1,2,3], MPH, MD

[1]Partners HealthCare Pivot Labs, Partners HealthCare, Boston, MA, United States

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

[3]Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Connor Devoe, BS
Partners HealthCare Pivot Labs
Partners HealthCare
25 New Chardon Street
Suite 300
Boston, MA, 02114
United States
Phone: 1 6176436368
Email: cdevoe@partners.org

Abstract

Background: Participant recruitment, especially for frail, elderly, hospitalized patients, remains one of the greatest challenges for many research groups. Traditional recruitment methods such as chart reviews are often inefficient, low-yielding, time consuming, and expensive. Best Practice Alert (BPA) systems have previously been used to improve clinical care and inform provider decision making, but the system has not been widely used in the setting of clinical research.

Objective: The primary objective of this quality-improvement initiative was to develop, implement, and refine a silent Best Practice Alert (sBPA) system that could maximize recruitment efficiency.

Methods: The captured duration of the screening sessions for both methods combined with the allotted research coordinator hours in the Emerald-COPD (chronic obstructive pulmonary disease) study budget enabled research coordinators to estimate the cost-efficiency.

Results: Prior to implementation, the sBPA system underwent three primary stages of development. Ultimately, the final iteration produced a system that provided similar results as the manual Epic Reporting Workbench method of screening. A total of 559 potential participants who met the basic prescreen criteria were identified through the two screening methods. Of those, 418 potential participants were identified by both methods simultaneously, 99 were identified only by the Epic Reporting Workbench Method, and 42 were identified only by the sBPA method. Of those identified by the Epic Reporting Workbench, only 12 (of 99, 12.12%) were considered eligible. Of those identified by the sBPA method, 30 (of 42, 71.43%) were considered eligible. Using a side-by-side comparison of the sBPA and the traditional Epic Reporting Workbench method of screening, the sBPA screening method was shown to be approximately four times faster than our previous screening method and estimated a projected 442.5 hours saved over the duration of the study. Additionally, since implementation, the sBPA system identified the equivalent of three additional potential participants per week.

Conclusions: Automation of the recruitment process allowed us to identify potential participants in real time and find more potential participants who meet basic eligibility criteria. sBPA screening is a considerably faster method that allows for more efficient use of resources. This innovative and instrumental functionality can be modified to the needs of other research studies aiming to use the electronic medical records system for participant recruitment.

(JMIR Med Inform 2019;7(2):e10020) doi:10.2196/10020

KEYWORDS

recruitment; silent BPA notifications; research; enrollment; innovation; electronic medical record; COPD

Introduction

Although clinical research is critical to our understanding of disease etiology and the development of novel therapeutics, a commonly encountered problem in clinical trials is the challenge of meeting enrollment targets in the stipulated time. For example, a study of neuroimaging in cognitively impaired geriatric patients found that 58% of potential participants failed to enroll due to a lack of interest [1]. Another study that aimed at screening women aged ≥50 years with ovarian cancer demonstrated similar results [2]; only 54.6% of the eligible candidates who were contacted for the study were willing to participate, and those who were not willing to participate cited reasons such as wanting more information from their doctor, inconvenience, or not believing themselves to be at risk for developing ovarian cancer [2]. Among elderly and inpatient populations, this difficulty in recruiting participants is even more apparent. A study that evaluated cognitive dysfunction in older adults after admission for heart failure reported that potential participants expressed interest in participation at the initial encounter, but later rescinded their interest, often without giving a specific explanation [3]. Participants who provided reasons stated they were too tired, too sick, or no longer interested, among others [3]. Given these recruitment challenges, it is imperative to find novel strategies to facilitate participant recruitment in clinical trials.

Electronic health record (EHR) systems have the potential to facilitate rapid patient recruitment in clinical research [4]. This is primarily due to the widespread use of EHRs in clinical practice since the enactment of the Affordable Care Act. Leveraging various functionalities within the EHR system may facilitate earlier patient identification and increased study enrollment. One such EHR functionality is the Best Practice Alert (BPA) notification systems. BPAs are "...automated alerts within the electronic medical record that help facilitate widespread communication of information to primary care providers..." [4]. They are used clinically to save time, identify patients for follow-up, and increase clinician efficiency [5]. In a pilot study at Yale New Haven Hospital, BPAs were used to identify patients that may be good candidates for a smoking-cessation medication [5]. At the San Francisco Medical Center, University of California, one group implemented BPAs in the Apex EHR for the use of telemetry or continuous cardiac monitoring [6]. This BPA notified clinicians about discontinuing the use of telemetry for patients who exceeded the nationally recommended duration for telemetry [7].

One common reason for not adopting BPA is alert fatigue. In one study at the Stanford Medical Center that attempted to integrate BPA into their EHR to provide clinical decision support in the computerized physician order entry for transfusions, it was found that clinicians continued to order transfusion blood products outside of the recommended guidelines, despite BPAs, exposing several questionable practices surrounding transfusions such as perioperative and periprocedural transfusions or orders anticipating imminent discharge [8]. More recently, the Massachusetts General Hospital, Boston, MA, expanded the use of pop-up BPA notifications to alert providers of patients on the opioid registry in the Epic EHR. Through these alerts, patients receiving outpatient prescriptions for opioids can be monitored (Multimedia Appendix 1).

In this quality-improvement project, the study team was interested in taking this framework of using BPA in clinical settings and tailoring it specifically to increase patient recruitment in a clinical trial. To mitigate concerns about alert fatigue, the study staff implemented and tested a "silent" BPA (sBPA) system. The study team defines sBPA as one in which these notifications do not appear as pop-up messages in the EHR view, but rather in a separate inbox or "in-basket" that can be checked periodically. This system eliminates the need to manually search through inpatient admission data by filtering patient data through an algorithm that identifies candidates who meet preselected screening criteria and subsequently sends the list of identified candidate participants to an in-basket messaging service within the EHR. In this project, our goal was to develop an sBPA system that could be used to efficiently identify potential participants admitted to the hospital in order to expedite recruitment in a difficult-to-recruit, elderly inpatient population.

Methods

Study Population and Settings

The BPA system used in this quality-improvement project was developed to increase recruitment rates for a prospective observational cohort study (Emerald-COPD study), approved by the Partners HealthCare Institutional Review Board. The study aimed to collect objective measures (eg, physical activity and inhaled medication use) and self-reported subjective measures in patients with chronic obstructive pulmonary disease (COPD). The overall goal of the Emerald-COPD study was to determine whether the collected data could be used to predict the patient's health status, such as an acute exacerbation of their COPD or readmission due to COPD. The targeted enrollment sample size was 300, and participants were recruited from the inpatient floors of three Partners Healthcare Hospitals: the Massachusetts General Hospital, Brigham and Women's Hospital, and Brigham and Women's Faulkner Hospital. The study team consisted of five research coordinators that performed all study procedures. Inclusion criteria included sufficient understanding of the English language, willingness to participate in the research, hospitalization within 24 hours of primary or secondary diagnosis of COPD, and discharge from the hospital to home. In this study, it was important to identify potentially eligible participants prior to hospital discharge as dictated by the study protocol.

Traditional Recruitment Method, Workflow, and Challenges

The first step in the recruitment process was identification of potentially eligible participants upon hospital admission.

Traditionally, this was completed by sorting through the Epic Reporting Workbench module. The module is a reporting tool that lives in the Epic Hyperspace (Partners EHR system) and pulls data from the millions of records in the system into a template form, which includes basic demographic information and admission diagnoses. Research coordinators filtered the form by COPD-related admission diagnoses to reduce the number of records flagged for further review. On an average, research coordinators spent 2 hours per day manually sorting through the Epic Reporting Workbenches of the three recruiting hospitals and reviewing admission notes to determine whether potential participants met the study eligibility criteria. Potential participants who met the basic screening criteria were often excluded for various reasons including unstable psychiatric disposition; illicit drug use; or designation to be discharged to rehab, hospice, or long-term care facilities. Following identification and initial screening, research coordinators secured (via email or page) permission from the care provider as per hospital policy prior to approaching the identified potential participants.

Overview of the Silent Best Practice Alert System

The Epic EHR system at Partners HealthCare contains an application programming interface that enables seamless integration of programs like the BPA system. For example, in a clinical decision support BPA for medications, the algorithm would search the EHR for potential prescription discrepancies to provide the best possible clinical recommendations. To reduce the time spent screening for potentially eligible participants, the study staff repurposed a similar BPA system to search the EHR and send an alert to research coordinators notifying them that a patient meeting the preselected study criteria was admitted to the hospital, via an in-basket messaging service accessed through the EHR home screen.

The traditional Epic Reporting Workbench lacked a simplified storage system. Study staff would perform daily screenings and create a finalized list of names, associated admitting diagnoses, and care provider contact information in Microsoft Excel. This enabled research coordinators to mark potential candidates as interested or not interested and to fill in the reason for not participating, if the latter applied. The sBPA system provided a solution to this storage problem, as the in-basket could be managed similar to an email inbox. Additionally, the demographic information, including name, admitting diagnosis, and care provider contact information, was included in the alert and stored in the in-basket. This not only simplified the process of identifying potential study candidates, but also facilitated outreach to patients and their providers (Multimedia Appendices 2-5).

Development: Iteration I

Study staff and collaborators from Partners eCare Research Core (PeRC) built the sBPA functionality in Epic EHR. The real-time alerts were not in the form of pop-up notifications that led to alert fatigue in the original BPA system; instead, the research coordinators received sBPAs in an email-like, in-basket format. The alerts contained a link to relevant patient information such as a patient's current medications and past medical history, which further helped the research coordinator in screening for study eligibility.

The preselected criteria provided to the PeRC team fell into the three distinct categories corresponding to established data capture fields in the Epic EHR: admission diagnosis, Epic problem list, and medication list. In this study, the International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) codes of admission diagnoses related to COPD (eg, shortness of breath, dyspnea, cough, pneumonia, and respiratory failure), COPD appearing in problem list, and medications associated with COPD were all included in the set of multiple nested conditions, which were to be met before an alert was triggered. This initial iteration was completed 1 month after providing the preselected criteria to the PeRC team.

Implementation

During the initial implementation phase, study staff compared the number of potentially eligible participants provided by the newly implemented sBPA system with our traditional screening method in order to refine the sBPA logic and ensure that all potential participants captured in the manual screening method were also captured through the sBPA. From March to October 2017, weekly comparisons between the two screening methods were assessed for yield of potentially eligible participants and time taken to complete daily screenings. This quality testing was performed by the study research coordinators who used the Epic Workbench and were trained to use the sBPA system. The research coordinators first screened for potential participants through the Epic Reporting Workbench and then screened again using the sBPA. Additionally, both methods were timed to assess the effort required for each screening method. By dividing the cost/hour budgeted for research coordinators' time by the total time spent using sBPA, the study staff provided an estimate of any projected change in costs, both in hours and US dollars, should a similar system be utilized.

Refinement: Iterations II, III, and IV

The original sBPA was setup to flag potential participants who either had an admitting diagnosis related to COPD, COPD in their problem list, or a medication associated with COPD treatment as part of their prescribed inpatient or outpatient medication history. In July 2017, the study staff introduced an iteration of the sBPA by revising the trigger conditions provided to the PeRC team. Instead of flagging potential participants who met at least one of the preselected criteria, the sBPA trigger condition aimed at flagging potential participants who met at least two of the preselected criteria. Again, in early August 2017, the study staff revised the trigger conditions provided to the PeRC team and, with this second iteration, potential participants who were prescribed multiple COPD-related medications during their inpatient stay were not flagged as satisfying two of the three criteria set in the July 2017 iteration.

Based on the results of weekly comparisons during the implementation phase, the sBPA trigger logic was adjusted for a third time in late August 2017 to only capture potential participants who met criteria that fell within two of three preselected categories (eg, patient with COPD in the problem

list and COPD-related medication) as opposed to two criteria within the same category (eg, patient with two COPD-related medications). This adjustment was made to account for the number of potential participants flagged by the sBPA who would otherwise not be flagged as eligible in the manual screening method. Another final adjustment was made at the end of the refinement period in October 2017 to strengthen the logic in order to ensure that only potential participants with a hospital inpatient status were flagged as opposed to those who only visited the emergency department but were not transitioned to inpatient admission. This fourth and final iteration added a fourth category of the preselected criteria for the sBPA—inpatient status—yielding the four categories of COPD-related admitting diagnosis, COPD-related medication, COPD in problem list, and inpatient status. Throughout all iterations, the sBPA preselected logic was executed on all hospital admissions across the three participating hospitals.

Results

Development of the sBPA system was completed by the PeRC team 1 month after the study criteria were received from the study staff. A total of 559 potential participants were identified from March 1 to October 2, 2017, from both screening methods. Of these, 418 (of 559, 74.77%) potential participants were identified by both the Epic Workbench method and sBPA (Figure 1 a). Of the potential participants identified from both screening methods, 287 (of 418, 68.66%) were considered eligible. Of the potential participants considered eligible, 60 participants enrolled in the study. Those who did not enroll either declined or were found to be ineligible for a reason other than that listed in the initial screening criteria (eg, active lung cancer or other serious conditions, psychiatric conditions, or language barrier).

Although the sBPA system was being used simultaneously, the Epic Workbench method found additional 99 potential participants who were not identified via the BPA notifications method (Figure 1 a). Of those found by only the Epic Workbench, 12 potential participants (of 99, 12.12%) were determined to be eligible to participate in the study (Figure 1 b). Over the four iterations, the sBPA notifications method found 42 additional potential participants who were not identified by the Epic Workbench method (Figure 1 a). Of these sBPA-only potential participants, 30 participants (of 42, 71.43%) were determined to be eligible by the emailed physicians (Figure 1 c). In summary, although the sBPA method of screening identified fewer potential participants in total over the course of the project, a greater percentage of the potential participants identified were later confirmed to be eligible for study participation. From the overall increase in identified eligible patients, the study staff determined that the system found the equivalent of three additional eligible potential participants per week.

The sensitivity and specificity of the sBPA system in identifying potential participants varied by iteration. The first iteration of the sBPA system flagged for ICD-10 codes for related admitting diagnoses was related to medications and COPD in the problem list. The initial iteration contained trigger logic that was very sensitive, hence including many potential participants that met at least one of the prespecified conditions, but was not specific enough to identify those who were eligible for study participation. Therefore, during much of the initial phases of development and implementation, there was an overpull of potential participants who did not meet the study requirements. Instead of reducing the time spent screening, the study staff spent more time going through potential participants that had met one of three criteria but were not potentially eligible. To resolve this issue, the study staff introduced the later iterations, which stipulate that individuals must meet two of three requirements. Iteration II yielded many names that lacked specificity. Iteration III appeared to be too restrictive and provided far less names than the earlier iterations. Additionally, Iterations I, II, and III generated many names of individuals who were admitted to the emergency department or were under observation. The study team found that the specificity afforded by the modifications that led to Iteration IV optimized the system to generate names of inpatients who met initial screening requirements.

An additional outcome observed when comparing potential participants missed by the Epic Workbench to the sBPA notifications was that the study staff had missed potentially eligible participants who were not identified by the system because their primary diagnosis was not a COPD-related ICD-10 code (ie, COPD exacerbation, dyspnea, shortness of breath, and chest pain). The Epic Workbench model only filtered information based on the primary admitting diagnosis, and therefore, many individuals were missed if they had a primary admitting diagnosis that did not align with these ICD-10 codes. The specified flagged criteria used by the sBPA system proved to be more effective in identifying eligible participants, because it did not rely solely on these potential participants' primary diagnoses.

In addition to the increase in the number of identified eligible potential participants, the sBPA system reduced the screening time. The average screening time for the Epic Workbench screening method was 123 minutes per day, and the average time to complete screening with sBPA notifications was 29 minutes per day. Thus, the sBPA notification method was approximately four times faster than the traditional Epic Reporting Workbench method, or yielded a 76.42% decrease in time spent screening. By dividing the cost/hour budgeted for research coordinators' time by the total time saved with BPA, the cost savings projections, given this increased efficiency, are projected to be US $15,487.50 with over 442.5 hours saved by the end of the study (Table 1). This saving factors in the cost/hour of research coordinators allotted by the study budget (US $35/hour) and the estimated number of hours spent screening over the study duration (approximately 590 hours for the Epic Workbench method and 147.5 hours for the Best Practice Alert notification method).

Figure 1. Comparison of identified potential participants: traditional Workbench method versus silent Best Practice Alert. BPA: Best Practice Alert.

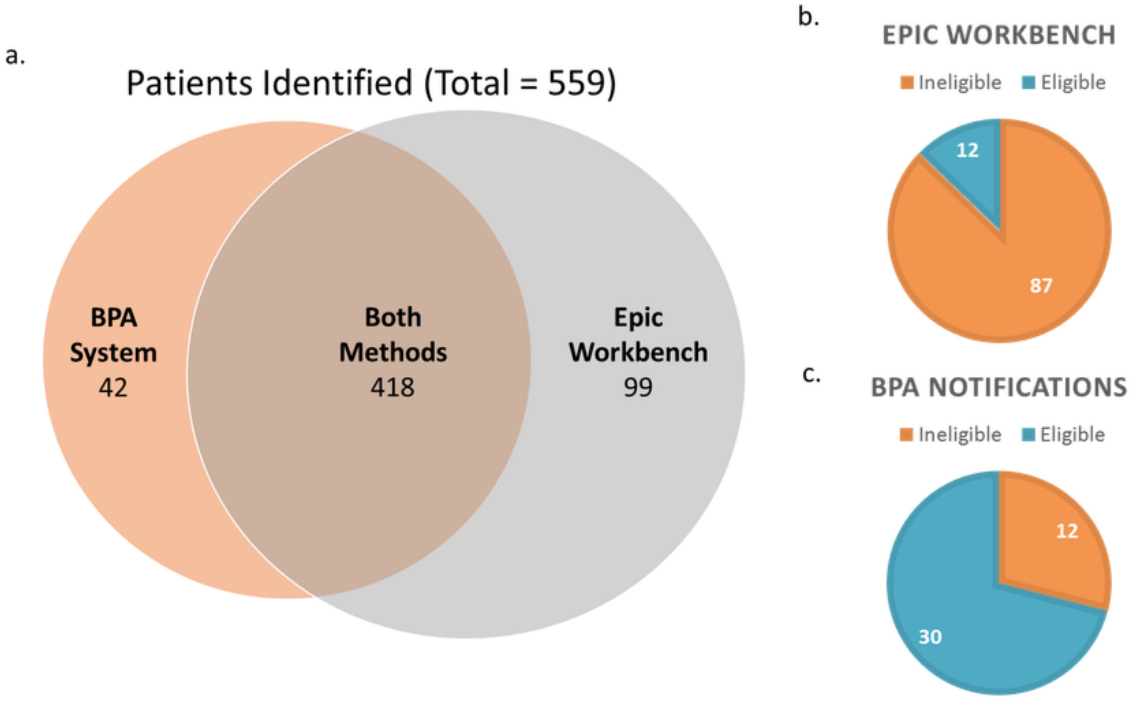


Table 1. Total savings, in hours and US dollars, from the silent Best Practice Alert notifications screening method as compared to the previous method of screening. Expenses are calculated from March 8, 2017, to May 1, 2018 (project completion).

Category	Expenses		
	Cost/hour (US $)	Total hours	Total cost (US $)
Without BPA[a] notifications	35	590.0	20,650.00
With BPA notifications	35	147.5	5162.50
Saved	N/A[b]	442.5	15,487.50

[a]BPA: Best Practice Alert.

[b]N/A: not applicable.

Discussion

Principal Findings

Although BPA notifications are increasingly being used in clinical settings [4-8], this is one of the first studies to implement the application of an sBPA system in the context of clinical research. The primary purpose of this quality-improvement project was to develop, refine, and implement a more efficient and usable version of the BPA system to increase patient recruitment for a clinical trial. Additionally, the study team aimed to determine if automation of the recruitment process through sBPA notifications would not only save time, but also help identify potential participants who were previously missed during our traditional Epic Reporting Workbench screening. Through four iterations, the study team worked to optimize this system through implementation and real-time refinement. Ultimately, sBPA notifications proved to be a considerably faster method of screening. This had a positive impact on the overall success of the study, as a faster recruitment method allowed the study staff to devote more time to other aspects of the research, thus decreasing total hours and, in turn, total cost. Automatic identification of potential participants in real-time through the in-basket reduced the frequency of screening failures and increased the pool of potential participants in a difficult-to-recruit population.

Rapid recruitment and enrollment are vital to the success of any clinical trial, and if the recruitment period must be extended to reach a target sample size, it will delay the trial and result in increased costs [9,10]. This is particularly true for clinical trials involving technology-enabled products, as the spate of innovations in the digital health industry is overwhelming. A device developed a few months prior can become obsolete before the study has had a chance to enroll the targeted enrollment sample size. Nevertheless, digital health solutions, like any other intervention for patient care, need to be rigorously evaluated before broader adoption in clinical settings, and clinical trials remain the gold standard for validating these solutions [9]. Therefore, studies must be designed with methodologies that maintain high internal and external validity and yet allow recruitment to be completed in the shortest time

possible. Application of sBPA notifications in clinical studies can facilitate rapid patient enrollment and help study teams meet their recruitment targets.

sBPA notifications have the potential to increase the overall study cost-efficiency by reducing the number of hours the study staff spend on initial patient identification and screening. It is well known that the cost of conducting clinical trials is rapidly increasing, which has negative implications for the development and evaluation of new interventions for patient care [9]. Patient recruitment time accounts for about 30% of the overall study time and is one of the top reasons for the increasing cost of clinical trials [11]. To plan for unavoidable recruitment factors like ineligibility and lack of interest, it is important to efficiently identify as many candidate participants as quickly as possible. In this quality-improvement project, the study team demonstrated that sBPA notifications and other EHR-based methods that facilitate earlier patient identification and screening may increase the cost-efficiency of clinical trials.

Scalability of similar systems across other EHRs for research screening purposes is becoming more possible through initiatives to increase interoperability of these databases across heterogeneous systems. Much of this interoperability is made possible through the utilization of electronic data capture platforms, such as Vanderbilt University's Research Electronic Data Capture (REDCap) software. Many electronic data capture platforms, including REDCap, have built-in functionalities that facilitate export and import of data from EHRs [12]. In Europe, government initiatives have been implemented to increase shared data across EHR platforms. Electronic Health Records for Clinical Research (EHR4CR) is a €16 million initiative across 35 academic research centers and pharmaceutical companies to create a massive, de-identified EHR data repository in order to assist in prospective eligibility screening and patient recruitment efforts in clinical research [13]. Having secured permissions from patients, the public, and researchers across the continent, the platform also enables researchers to access EHR data from hospitals to determine project feasibility and locate the optimal sites to carry out clinical trials based on their populations [13]. Similar initiatives are seen in the United States, including the Patient-Centered Outcomes Research (PCOR) Institute PCORnet service, a platform that integrates clinical data from 11 clinical data research networks to create sustainable infrastructure for use in comparative effectiveness research [14,15]. These developments are an exciting start; however, more research will be required to assess the functionality of these systems in other modalities such as screening and recruitment.

Limitations

This project was not designed to test a specific study hypothesis. Therefore, some of the project processes are not predefined or do not adhere to any strict study procedures, which raise concerns about the project's reproducibility. However, our goal was to create a system to improve the efficiency of our prescreening process and the developed sBPA system served that purpose. Additionally, there is a challenge of limited generalizability of findings from this quality-improvement project, as it applies to other settings. This method is specific to the Epic Reporting Workbench in an integrated delivery network of hospitals; therefore, the study staff would be required to use the Epic Reporting Workbench to test the sBPA's success in reducing screening time and expanding recruitment. Moreover, information bias may be a problem, because the traditional screening method was performed by one research coordinator who then checked the in-basket messages to compare the patients identified from both methods. With adequate resources, these procedures would ideally be carried out by at least two different screening methods. Although the research team conceived the idea to setup the sBPA system, we depended on another team for actual development and associated timelines. There is also a cost associated with the development, to pay for the developers' efforts. Thus, researchers would need to account for these costs in their study budget.

Conclusions

Utilizing EHR for clinical research and automation of the recruitment workflow process has broad implications for accelerating innovation in health care. The sBPA notifications can help reduce the amount of time spent screening and increase the potential patient pool for study recruitment, resulting in increased cost-efficiency and accelerated study-completion timelines.

Conflicts of Interest

None declared.

Multimedia Appendix 1

An example of a pop-up notification that could be used to inform clinical decision making. To clear this notification, an action is required from the provider.

[PDF File (Adobe PDF File), 53KB - medinform_v7i2e10020_app1.pdf]

Multimedia Appendix 2

All institutional review board–approved research staff receive a best practice alert, real-time notification in their Epic in-basket. This is easily accessed under the "BestPractice" tab in the lower left-hand corner of the Epic homepage.

[PDF File (Adobe PDF File), 90KB - medinform_v7i2e10020_app2.pdf]

Multimedia Appendix 3

The silent Best Practice Alert notifications detect an event of interest (in this case, hospital admissions), using criteria preselected by the Emerald-COPD research staff. COPD: chronic obstructive pulmonary disorder.

[PDF File (Adobe PDF File), 242KB - medinform_v7i2e10020_app3.pdf]

Multimedia Appendix 4

From in-basket view, the research staff can identify patient demographics and reasons for potential eligibility.

[PDF File (Adobe PDF File), 228KB - medinform_v7i2e10020_app4.pdf]

Multimedia Appendix 5

The study staff can directly access the “Encounter” tab for further information regarding patient eligibility.

[PDF File (Adobe PDF File), 259KB - medinform_v7i2e10020_app5.pdf]

References

1. Apostolova I, Lange C, Roberts A, Igel HJ, Mäurer A, Liese S, et al. Challenges in screening and recruitment for a neuroimaging study in cognitively impaired geriatric inpatients. JAD 2017 Jan 12;56(1):197-204. [doi: 10.3233/jad-160797]
2. Sun Z, Gilbert L, Ciampi A, Basso O. Recruitment challenges in clinical research: Survey of potential participants in a diagnostic study of ovarian cancer. Gynecol Oncol 2017 Dec;146(3):470-476. [doi: 10.1016/j.ygyno.2017.06.021] [Medline: 28666542]
3. Arslanian-Engoren C, Giordani BJ, Algase D, Schuh A, Lee C, Moser DK. Recruitment and retention challenges of examining cognitive dysfunction in older adults hospitalized for acute heart failure. Am J Crit Care 2016 Sep;25(5):418-421 [FREE Full text] [doi: 10.4037/ajcc2016305] [Medline: 27587421]
4. Sauers-Ford HS, Gold JM, Statile AM, Tubbs-Cooley HL, Simmons JM, Shah SS, H2O Study Group. Improving recruitment and retention rates in a randomized controlled trial. Pediatrics 2017 May;139(5) [FREE Full text] [doi: 10.1542/peds.2016-2770] [Medline: 28557728]
5. Max J. Yale Center for Clinical Invertigation. 2014. Epic: Changing the Way Medicine is Practiced URL: https://tinyurl.com/y5kx3js6 [accessed 2019-03-11] [WebCite Cache ID 76ngiby1j]
6. Najafi N, Cucina R, Pierre B, Khanna R. ClinicalTrials.gov. Trial of a Best-Practice Alert in the Electronic Medical Record to Reduce Unnecessary Telemetry Monitoring URL: https://clinicaltrials.gov/ct2/show/NCT02529176 [accessed 2019-03-14]
7. Dressler R, Dryer MM, Coletti C, Mahoney D, Doorey AJ. Altering overuse of cardiac telemetry in non-intensive care unit settings by hardwiring the use of American Heart Association guidelines. JAMA Intern Med 2014 Nov;174(11):1852-1854. [doi: 10.1001/jamainternmed.2014.4491] [Medline: 25243419]
8. Chen JH, Fang DZ, Tim Goodnough L, Evans KH, Lee Porter M, Shieh L. Why providers transfuse blood products outside recommended guidelines in spite of integrated electronic best practice alerts. J Hosp Med 2015 Jan;10(1):1-7. [doi: 10.1002/jhm.2236] [Medline: 25044190]
9. Campbell M, Snowdon C, Francis D, Elbourne D, McDonald A, Knight R, STEPS group. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. Health Technol Assess 2007 Nov;11(48):iii, ix-iii,105 [FREE Full text] [Medline: 17999843]
10. Kadam R, Borde S, Madas S, Salvi S, Limaye S. Challenges in recruitment and retention of clinical trial subjects. Perspect Clin Res 2016;7(3):137. [doi: 10.4103/2229-3485.184820]
11. Schulthess D. Opt-in, Opt-out, & Patient Led Databases: Better Patient Outcomes, Faster and Cheaper. 2014. Vital Transformation URL: https://vitaltransformation.com/wp-content/uploads/2014/10/DGS_17-10-Opt-in-Opt-out-Patient-Led-Databases-MAPPs-DG3.pdf [accessed 2019-03-11] [WebCite Cache ID 76nh0KflL]
12. Vawdrey D, Weng C, Herion D, Cimino JJ. Enhancing electronic health records to support clinical research. AMIA Jt Summits Transl Sci Proc 2014;2014:102-108 [FREE Full text] [Medline: 25954585]
13. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015 Feb;53:162-173 [FREE Full text] [doi: 10.1016/j.jbi.2014.10.006] [Medline: 25463966]
14. 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 Jul;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
15. Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: Shining Light on a New Research Paradigm. Ann Intern Med 2015 Oct 20;163(8):635-636. [doi: 10.7326/M15-1460] [Medline: 26301537]

Abbreviations

BPA: Best Practice Alert
COPD: chronic obstructive pulmonary disease
EHR: electronic health record
EHR4CR: Electronic Health Records for Clinical Research
ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th revision
PCOR: Patient-Centered Outcomes Research
PeRC: Partners eCare Research Core
REDCap: Research Electronic Data Capture
sBPA: silent Best Practice Alert

Edited by G Eysenbach; submitted 05.02.18; peer-reviewed by C Knopp, T Guetterman, P Athilingam, A Rodríguez-Molinero, A Argha, J Wang, T Adedeji; comments to author 09.04.18; revised version received 04.09.18; accepted 31.12.18; published 26.04.19.

Please cite as:
Devoe C, Gabbidon H, Schussler N, Cortese L, Caplan E, Gorman C, Jethwani K, Kvedar J, Agboola S
Use of Electronic Health Records to Develop and Implement a Silent Best Practice Alert Notification System for Patient Recruitment in Clinical Research: Quality Improvement Initiative
JMIR Med Inform 2019;7(2):e10020
URL: http://medinform.jmir.org/2019/2/e10020/
doi:10.2196/10020
PMID:31025947

Original Paper

Privacy-Preserving Analysis of Distributed Biomedical Data: Designing Efficient and Secure Multiparty Computations Using Distributed Statistical Learning Theory

Fida K Dankar[1], PhD; Nisha Madathil[1], MS; Samar K Dankar[2], PhD; Sabri Boughorbel[3], PhD

[1]United Arab Emirates University, Abu Dhabi, United Arab Emirates

[2]Independent Scientist, Ottawa, ON, Canada

[3]Sidra Medicine, Doha, Qatar

Corresponding Author:
Fida K Dankar, PhD
United Arab Emirates University
College of IT, Al Ain
Abu Dhabi, 15551
United Arab Emirates
Phone: 971 37673333 ext 5569
Email: fida.dankar@uaeu.ac.ae

Abstract

Background: Biomedical research often requires large cohorts and necessitates the sharing of biomedical data with researchers around the world, which raises many privacy, ethical, and legal concerns. In the face of these concerns, privacy experts are trying to explore approaches to analyzing the distributed data while protecting its privacy. Many of these approaches are based on secure multiparty computations (SMCs). SMC is an attractive approach allowing multiple parties to collectively carry out calculations on their datasets without having to reveal their own raw data; however, it incurs heavy computation time and requires extensive communication between the involved parties.

Objective: This study aimed to develop usable and efficient SMC applications that meet the needs of the potential end-users and to raise general awareness about SMC as a tool that supports data sharing.

Methods: We have introduced distributed statistical computing (DSC) into the design of secure multiparty protocols, which allows us to conduct computations on each of the parties' sites independently and then combine these computations to form 1 estimator for the collective dataset, thus limiting communication to the final step and reducing complexity. The effectiveness of our privacy-preserving model is demonstrated through a linear regression application.

Results: Our secure linear regression algorithm was tested for accuracy and performance using real and synthetic datasets. The results showed no loss of accuracy (over nonsecure regression) and very good performance (20 min for 100 million records).

Conclusions: We used DSC to securely calculate a linear regression model over multiple datasets. Our experiments showed very good performance (in terms of the number of records it can handle). We plan to extend our method to other estimators such as logistic regression.

(JMIR Med Inform 2019;7(2):e12702) doi:10.2196/12702

KEYWORDS

data analytics; data aggregation; personal genetic information; patient data privacy

Introduction

Background and Significance

Human genome research promises to transform health care through personalized medicine. It enables the determination of an individual's unique molecular characteristics, which can be used to diagnose diseases, select individualized treatments (with a higher success rate), and reduce possible adverse reactions [1]. However, before this becomes a reality, more research is needed to understand the complex relationship between genome and health. Such research often requires large cohorts and necessitates the sharing of biomedical data with researchers around the world, which raises many privacy, ethical, and legal concerns.

Traditionally, researchers would strip the data from the identifying information—such as names and identity cards—and apply some privacy-protection techniques—such as generalization or noise addition—before sharing them with each other. However, recent studies have shown that it is possible to deduce the identity of research participants from clinical data that were considered anonymized. DNA sequencing aggravates this problem as the genome is unique to every individual and can be used to predict future ailments for individuals and their blood relatives (such as Alzheimer's or schizophrenia). Such information has the potential to deny jobs and to isolate subjects socially [2]. In the face of these growing concerns, privacy experts are trying to explore alternative approaches to privacy protection. Many of the new strategies are based on cryptography, particularly secure multiparty computations (SMCs). SMC is an attractive approach that allows a set of multiple parties, $S_1,...,S_m$, each holding a private fraction of the data to be analyzed, to collectively carry out a computation f on the overall dataset, without any party having to reveal their own private raw data. Thus, the goal is to compute f efficiently and privately such that no party learns anything aside from the final output of f. Note that the output is computed from the private inputs of the different parties, and as such, it may leak some sensitive information about their input. In fact, SMCs focus on the security of the distributed computation method and do not specify which kind of computations can be performed when privacy is of interest. In other words, it does not specify whether the output of a given computation will leak sensitive information or not, it just guarantees that the computation method itself preserves the privacy of the distributed raw data. Techniques from differential privacy have been used (in combination with SMCs) to prevent leakage of sensitive information from the final output. The discussion of these mechanisms is beyond the scope of the study; for more information, readers are referred to the studies by Beimel et al, Nordholt et al, and Papadimitriou et al [3-5].

Despite the mathematical proofs that have been established, demonstrating the ability of the SMC protocols to protect data, they are still not widely used. This may be because knowledge about their capabilities is still relatively small, they tend to have complex solutions that are not accessible without domain knowledge, they require coordinating analyses among the different sites, or they are not efficient in every setting. In fact, one of the main problems with SMC protocols is efficiency. Communication between the different parties is the main factor driving the inefficiency of SMC protocols [6-8]. In almost all existing research in SMC, one of the main goals is to minimize the total number of messages communicated between the different parties and, thus, minimize the performance gap between secure and regular protocols [9]. One of the approaches taken is to relax security and privacy requirements (such as allowing some information leakage) [10].

Our goal with this line of research is to develop usable and efficient SMC applications that meet the needs of the potential end-users and to raise general awareness about SMC as a tool that supports data sharing. Thus, we proposed a divergence from current research efforts. Instead of lowering the security requirements, we proposed to introduce distributed statistical computing (DSC) into the design of secure protocols. Through DSC, we will study the possibility of conducting computations on each of the parties' sites independently and then combine these (local) computations to form 1 (accurate) estimator for the collective dataset, thus limiting communication to the final step and significantly reducing complexity.

Contributions

The main contribution of this study is introducing a model for privacy-preserving distributed data mining in which local models are produced separately and SMC is used to aggregate the results privately. The study applies these novel ideas to linear regression and introduces the first secure linear regression model that does model selection and parameter estimation efficiently (all previous secure multiparty algorithms perform parameter estimation only). The paper then presents experiments on real and synthetic datasets to demonstrate the accuracy and efficiency of the algorithm.

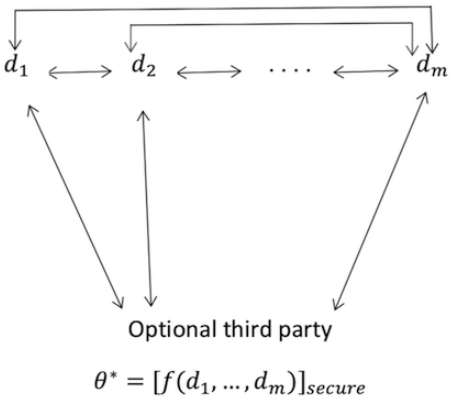
The paper is organized as follows: the next section defines our problem formally and introduces the theory behind DSC; the following section demonstrates the effectiveness of our privacy-preserving model through a linear regression application; and finally, the paper is concluded with a discussion of the results and limitations and a proposal for future research directions.

Methods

Problem Definition

A researcher wants to estimate a population parameter θ by running a computation f over the private inputs of several remote databases, $d_1,...,d_m$ while keeping these inputs private (Figure 1); (where $f(d_1,...,d_m)$ is a mechanism for the estimation of θ). The goal is to achieve a good approximation θ* of θ using as little communication as possible and without any party learning anything about other parties' input aside from the final output θ* (Figure 1).

Figure 1. Traditional secure computation approach. Double-sided arrows indicate required communication channels. All communication should be secure and no party (including the third party) should learn anything about other parties' input aside from the final estimation of θ. θ: population parameter to be estimated; d_i: private dataset owned by site *i* (where *i {1,...,m}*); *f*: a mechanism for the estimation of θ; θ*: the output of *f*; *m*: number of sites.



Interinstitutional data sharing is generally motivated by multiple scenarios such as (1) increasing results' accuracy and lowering bias, (2) performing benchmarking studies, or (3) attaining the cohort required for a study. In what follows, we illustrate each with a scenario:

1. *m* hospitals want to collectively study factors that affect the survival rate for breast cancer patients. Running the regression problem on the *m* datasets will provide better properties by increasing sample size and will reduce data bias (such as environmental and location bias). Sharing data in the open may not be easy as medical data are governed by privacy legislations.
2. Hospitals in a given geographical area are interested in calculating the average rate of hospital-acquired bacterial infection (across all the hospitals in the said area) for the purpose of self-evaluation. In this case, the hospitals have an additional incentive against data sharing as it may implicate them negatively.
3. Monogenic diseases are very rare genetic disorders associated with single gene variations observed in few subjects per 1000 to 10,000 individuals. Some are well-characterized such as cystic fibrosis (frequency of disease 1:2000), sickle cell anemia, phenylketonuria (frequency of disease 1:8000), and some primary immunodeficiency diseases [11]. The study of these rare disorders requires the sharing of data across multiple sources or institutions to enable the collection of more cases for analysis and thus increase the statistical power of the study.

Many protocols have been developed for the above problem in the area of SMC. The most efficient protocols are based on secret sharing [12], oblivious transfer [13], garbled circuits [14], or homomorphic encryption [15]. In addition, there are several hybrid constructions that combine these various models [10]. These protocols have robust mathematical proofs that demonstrate their ability to protect privacy under different assumptions of parties' honesty [10]. However, they mostly involve heavy communication (extensive message passing) between the different concerned parties [9]. To minimize the communication load and decrease the performance gap between secure and regular protocols, researchers tried to relax security and privacy requirements such as relaxing the number and power of dishonest parties or allowing some form of information leakage [10], others use noise addition to intermediate and final results to preserve privacy [16]. In this study, we proposed a change in the methodology by introducing distributed statistical learning into the design of secure computations.

Statistical Learning With Big Data

Overview

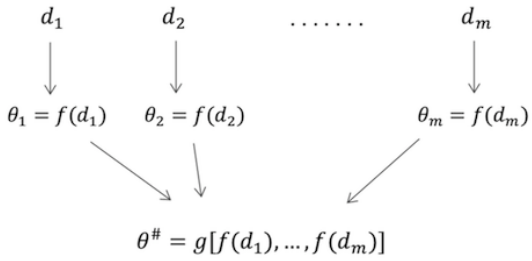
A common approach in statistical learning with big data is to split the data (along observations) into multiple subsets. Each subset conducts the required computation completely independently. The final result is then obtained by combining these local computations. Thus, communication (and sharing of information) is reduced to the final step only. This will significantly reduce the complexity and provide simpler algorithms. The problem is illustrated in Figure 2 and explained below.

A researcher is interested in estimating a population parameter θ from a sample database *D* with *N* records and *p* attributes. Traditionally, θ is estimated from the whole dataset *D* as: $\theta^*=f(D)$ (referred to as the centralized estimator), where *f* is a mechanism for the estimation of θ. In this *split and merge* statistical learning approach, the database *D* is split equally among *m* sites as $d_1,\ldots,d_m$. The number of records in the resulting databases is denoted by $n=N/m$. Each site *i* performs the estimation of θ on its local dataset as: $\theta_i=f(d_i)$, then the final estimate is obtained by combining the local estimates: $\theta^{\#}=g(\theta_1,\ldots,\theta_m)$.

The *N* observations are assumed to be independent and identically distributed. They are evenly and randomly allocated along the *m* different sites.

McDonald et al, who advocated for this *split and merge method* in [17], claim that, given the stated assumptions, it provides a good balance between accuracy and efficiency. As for merging strategies, few were considered in the literature, the most common ones being averaging [18] and median [19].

Figure 2. The split and merge approach. The one-sided arrows indicate message passing. All communication should be secure and no party (including the optional third party) should learn anything about other parties' input aside from the final estimation of θ. θ: population parameter to be estimated; di: private dataset owned by site *i* (where *i {1,...,m}*); *f*: a mechanism for the estimation of θ; θ_i: the output of *f* applied to d_i; *g*: a mechanism for combining local estimates; $\theta^{\#}$: the output of *g*; *m*: number of sites.



Relevant Theory

The *split and merge* strategy is simple to execute and is communication-efficient. Splitting is always done along observations (rather than attributes), and each site performs the estimation on its local dataset. Averaging of the *m* sites estimates is the simplest and most popular strategy. In what follows, we review the available literature while trying to answer the following questions:

- What is the error of the averaged estimator versus the centralized one?
- What affects the optimality of the averaged estimator?
- How many sites to include in a given study? And how many samples to include from these sites?

The accuracy of averaging depends on the relationship between the number of observations (*N*), the number of sites (*m*), and the number of parameters (*p*). As a general insight, averaging provides estimates that are as accurate as the centralized solution when there are many observations per parameter on each local machine [20]. In fact, in [21], the authors proved that when the number of records per site is large, (large $n=N/m$), the mean square error (MSE) of the average estimator (ie, $E||\theta^{\#}-\theta||^2$) is the same as the MSE of the centralized one (ie, $E||\theta^*-\theta||^2$). In [18], Rosenblatt and Nadler proved that the averaged estimator and the centralized one are first order-equivalent, they proved that the leading error term of ($\theta^{\#}-\theta$) and ($\theta^*-\theta$) converge to the same limit at the same rate; however, some accuracy loss is exhibited in higher-order terms for nonlinear models (the second-order term is *m* [number of sites] times larger than the first-order one). The interpretation as given in [18] is that first-order terms generally capture variance, which is reduced by averaging, whereas the second-order term captures bias which is not reduced by averaging. Thus, the old problem of balancing variance and bias comes to light in nonlinear models (where the second-order term can be nonnegligible). Approaches toward this problem can be found in [20-22]. Going further, Rosenblatt and Nadler presented an extensive analysis of the error of the averaging estimator by considering different practical situations [18]:

- For situations where *p* is fixed and *n* is large, they proved that the averaged estimator, $\theta^{\#}$, is asymptotically equivalent to the centralized one, θ*.
- For small and medium *n*, parallelization incurs a non-negligible error for nonlinear models.
- For situations where *N* is fixed, they showed that averaging performs well for small values of *m* and *p*.

The authors presented the exact expression of the estimator errors in all situations and confirmed the results through a series of experiments.

In [20], the authors proved that (for low-dimensional data) it is enough to have a smaller number of sites than local observations ($m \leq n$ or, equivalently, $m \leq N^{1/2}$) to guarantee an MSE that decays at a considerably better rate than the centralized approach. They also showed that when *N* is fixed, the MSE of the averaged estimator increases polynomially with the number of sites *m*, thus echoing previous theoretical results.

Battey et al specified in [23] an exact formulation of the requirements on *m*, *N*, and *p*, for linear models. They proved that when $p < log\ N$ and $m \leq Np/(log\ N)^2$, then the loss incurred by the averaging method is negligible compared with the statistical error incurred by the central one.

As a general insight into the questions raised above and to summarize the above results, we say that when the number of samples per site *n* is large, bigger than the number of features and bigger than the number of sites ($n > p$ and $n > m$) then the averaged machine-wise estimates are as accurate as the centralized estimates. However, when the number of samples per site, *n*, is small and the model is highly nonlinear, the error can be non-negligible. The nice results do not extend to high-dimensional data. When few observations are available per parameter per site ($n < p$), in these cases the accuracy loss increases linearly with the number of sites *m*. Some researchers, such as [24], resorted to schemes other than averaging to obtain well-behaved estimators in specific cases, whereas others showed moderate accuracy loss for averaging in specific cases [18,25,26].

Assumptions and Considerations

The assumptions across the DSC literature are that (1) the $N=mn$ observations are independent and identically distributed according to a distribution *P* and that (2) they are evenly and randomly allocated along the *m* different sites.

An equivalent assumption to (1), that applies to our SMC scenarios, is that of *m* independent sites having observations

that are independent and identically distributed according to the same distribution P.

The (second) assumption of evenness can be easily circumvented by pre-setting the number of samples to consider from each site. However, that is not needed, as the theoretical results presented in the previous section would still apply provided that every site satisfies the required assumptions (ie, the assumptions are true for each n_i, the number of observations for site i).

However, the first assumption is not always realistic and can be overly restrictive for some applications. For instance, if the data are already owned and collected by the different sites, then it may exhibit systematic differences across these sites. For example, if 2 hospitals are considering an analysis involving their cancer patients and if 1 of the hospitals is located in a heavily polluted area whereas the other is not, then the distribution of the local population from which the sites' data are sampled could have significant differences. Pooling the data and redistributing them randomly along the different sites may not be realistic or feasible as the data may be big or private [19].

Going back to the question of the number of sites to include in a study when p is fixed, the authors in [18] distinguished between 2 scenarios N fixed or n fixed. Fixed N captures the case of limited data availability or limited computational power whereas fixed n captures the case of storage restriction or data availability per site. For the SMC problem, where data are already owned by the different sites, fixed n represents the case of a given number of institutions (sites) wanting to run an analysis on their collective data (with n being the minimal number of samples across sites). Fixed N represents the case of a researcher with a requirement on cohort size and is assumed to be able to include as many sites as required to attain the cohort (with each site having at least N/m records). The authors in [18] presented an analysis of (1) the minimal number of sites to attain a desired accuracy in the fixed n scenario and (2) the maximal number of sites to attain a desired accuracy in the fixed N scenario. The objective is to guarantee an MSE that is lower than a preset value as follows: min $\{m;$ MSE$(m)\leq$ with fixed p and $n\}$ and max $\{m;$ MSE$(m)\leq$ with fixed p and $N\}$. For example, in the fixed N scenario with $p=10^3$ if $N=10^6$, m should be ≤ 899 to guarantee an MSE under 0.1 [18].

We distinguish between these 2 strategies when analyzing the performance of our algorithm.

Multiestimators

In many applications, certain inferences require 2 or more estimations. For example, inference for regression typically requires 2 components—feature selection and parameter estimation. When conducting feature selection, the median probability model has been recommended [27]; it consists of all the features that are selected by the majority of the subsets. According to [27], the median model produces the best approximation under some simplifying assumptions, in that it has the highest probability of being equal to the optimal model. Averaging is not recommended as it can lead to a bigger number of nonzero coefficients and, thus, to an inflation in the number of selected features as opposed to median. The median selection model is also less influenced by the heavy presence of outliers when compared with the central selection model, as the effect of the outliers will be waned down over multiple subsets (only a fraction of the subsets will contain a sizable fraction of the outliers) [28].

In [28], the authors present a distributed linear regression algorithm that combines median model for feature selection and simple averaging for parameter estimation. The authors proved that for low-dimensional data, when the features are independent and the number of sites is well chosen (number of sites m chosen so that $m<n$) or when features are correlated and following elliptical distributions (noting that real-world data commonly follow elliptical distributions), the distributed model provides accurate estimates. In fact, they showed that their algorithm can achieve better accuracy in terms of feature selection than the centralized one, which results in a better MSE in general. The authors performed extensive experiments (with $p<N$) that echoed their theoretical results. However, their choice of number of sites versus sample size always satisfied Zhang et al's condition [20], ($m\leq N^{1/2}$).

In the next section, we demonstrate the effectiveness of the privacy-preserving model through a linear regression application. We restrict our application to models with the best theoretical results, that is, linear models with more records than features in every site and with high number of records relative to the number of sites ($n>p$ and $n>m$).

Results

Application: Secure Linear Regression

We introduce the classical setting of a linear regression problem. Let $X=\{x_{i,j}\}$ be an $N\times p$ matrix of features and $Y=(y_1,...,y_N)^T$ a corresponding $N\times 1$ response vector, where N is the number of samples and p is the number of features. Linear regression consists of modeling the relationship between the set of features (also known as independent variables) and the response variable. It assumes that the relationship between the response variable and the independent variables is linear. Fitting a linear regression model consists of feature selection and parameter estimation [29]. Feature selection is the process of constructing a model that includes all relevant predicting variables. In other words, it is the process of determining the subset of features that best predicts the outcome variable, Y, whereas the parameter estimation consists of finding the linear model parameters β where $Y=X\beta+$ [29].

Despite the simplicity of linear regression, it is widely used in various biomedical applications [30]. Although physical and biological processes are inherently nonlinear, linear approximations have been successfully used for centuries to explain phenomena in physics and biology [30] as they present a number of advantages compared with more complex models. Parameters of linear models are usually easy to estimate, the linear models are easy to interpret (coefficient signs and values are indicative for the contribution of the different variables), and many tools have been developed to evaluate the statistical significance of linear models. Linear models are also well-suited for high-dimensional data and are used for association studies such as Genome Wide Association Studies.

Previous Work in Secure Linear Regression

As linear regression is one of the most commonly used statistical tools in data analysis, there are many attempts in the literature at obtaining secure linear regression protocols over distributed databases. Many of these protocols do not offer adequate privacy guarantees [31,32] as they share intermediate results or rely on a trusted third party to handle these intermediate results [33]. In [34], El Emam et al provide some scenarios where privacy can be breached by sharing intermediate aggregate results (refer to [35] for a decomposition of available secure regressions based on privacy and accuracy). The first linear regression algorithm with cryptographic security was developed by Hall et al [15]; it makes heavy usage of SMCs, particularly secure matrix multiplication protocol. The study reported 2 days for solving a linear regression problem of 51k rows and 22 features [15]. In [36], a solution based on homomorphic encryption and garbled circuits is presented. The solution uses 2 noncolluding semihonest third parties. The problem with the approach is that usage of garbled circuits imposed many rounds of interactions and is thus heavy on communication. In a more recent experiment [33], the authors report 8.75 hours for 10^8 records with 20 features and 270 MB of communication. In 2015, Cock et al, presented a method to calculate the parameters of the linear regression by computing $\beta=(X^TX)^{-1}X^TY$ [37]. The algorithm computes β by running Beaver's matrix multiplication protocol many times [38]. Beaver's protocol computes securely, with the help of a trusted initializer, the product of matrices shared by different parties in a way that the result remains shared by the different parties. The theoretical complexity of the algorithm is $O(Np^2)$; however, the protocol is heavy on communication. In fact, the matrix multiplication protocol requires each party to send 2 matrices to every other party (of size, p^2), such protocol is repeated, $O(k)$, where k is the maximal number of bits required to represent the largest integer. However, their algorithm performs better than all previous secure linear regression algorithms [37]. Experiments done by the authors themselves indicated a capacity to handle over 4 million records with 16 features in a range of 3 hours (to provide some perspective, our algorithm requires less than 3 min for the same dataset and same settings and for both feature selection and parameter estimation).

It is very important to note that all cited secure regression algorithms do not perform feature selection. In other words, they use the supplied features set to predict the model [35]. As our algorithm does both, we opted to compare our algorithm with the central version (where data are on the clear).

Our Algorithm

The goal of distributed statistical learning algorithms is to scale up computations by distributing the data over multiple machines. The underlying assumption is that all data are owned by the same organization. Thus, sharing of intermediate and local results between the different machines is allowed.

In our setting, the dataset (X, Y) is owned by $m \geq 2$ data holders (or sites) $S_1,...,S_m$ and the different sites are interested in cooperatively performing linear regression on the union of their datasets; however, they are not willing or able to share their data. Only the final result of the computation should be revealed to all parties. In line with the DSC theory, we assume that all the samples in all sites are independent and identically distributed (randomly drawn from the same [population] distribution). Moreover, if n_{min} is the smallest number of local observations across the different sites, (to guarantee the nice DSC results) we require that the number of features and number of sites are both smaller than n_{min}, that is, we require that $n_{min} \gg p$ and $n_{min} \gg m$.

Formally, the data (X,Y) are divided horizontally into m subsets $\{(X^1,Y^1);...;(X^m, Y^m)\}$, with $X^i=(X_1{}^i,...,X^i_p)$ the $n_i \times p$ feature matrix for subset i (where X^i_j is an $n_i \times 1$ matrix) and $Y^i=(y_1{}^i,...,y^i_{n_i})^T$ the corresponding $n_i \times 1$ response vector. The algorithm then executes the following 2 steps:

1. Each site calculates its local feature selection vector privately, and the local vectors are aggregated securely using a secure median algorithm (in other words, the parties jointly perform the median on their data and obtain the result), without any party revealing any information about their selected features (aside from what can be deduced from the final median output).
2. Next, each site uses the shared selected features to calculate the model parameters locally. These local parameters are then securely averaged using a secure average protocol. Our algorithm is presented in Multimedia Appendix 1. In the algorithm, the secure sum and secure median protocols are based on Paillier homomorphic encryption; however, other secure protocols can be used instead.

Experiments

We evaluated our secure multiparty linear regression algorithm (SMA) by implementing it and analyzing the results using real and synthetic datasets. The real datasets are used to test the accuracy of the algorithm whereas the synthetic datasets are used to analyze its performance. To analyze the accuracy of the algorithm, we needed real datasets that originated from multiple different sources (different data owners). The sources' IDs had to be included to inform the actual data division along the different sites. For the synthetic experiment, data were generated and randomly allocated along the different sites, as the purpose was solely to evaluate the efficiency of the algorithm for various numbers of records and features. We used Python3 as our programming language, which we augmented with the Scikit-learn, numpy, pandas, gmpy2, and phe libraries for functionality such as socket programming, homomorphic encryption, and for dealing with negative and real floating-point arithmetic. We built our system on top of 10 Linux machines with Intel Core i5-4570 CPU, 3.20GHz processor, and 8GB RAM, 4 cores each. To increase the number of possible sites to 20, we installed 2 Linux virtual machines on each machine with 4 GB memory each (note that the Paillier encryption library handles real-number values with arbitrarily high precision).

To test our SMA, we compared its performance with the central algorithm (CA). The CA performs linear regression on 1 machine using the same approach as the SMA (ie, it uses lasso for feature selection and linear least squares method for

parameter calculation [39]). We opted not to test the accuracy or the efficiency of our SMA algorithm against existing secure linear regression algorithms as none of the existing algorithms perform model selection.

Real Datasets

To test the accuracy of our algorithm, we collected real datasets that include information about the original collection site. Then, we treated each site as an independent data owner. We succeeded in finding 4 real datasets: 3 public datasets contained within the University of California Irvine repository (student performance in Portuguese, student performance in Math, and automobile fuel consumption data) and 1 from Cerner clinical database (the Diabetes dataset, where the number of sites included was varied between 3, 6, and 12, and the weight variable was excluded in some experiments because of excessive missing values). The datasets are presented in detail in Multimedia Appendix 2.

In the experiments on real datasets, we randomly divided the datasets into 0.7 training set and 0.3 testing set. A regression model was trained based on the training set and used to predict the outcome variable in the testing set. The average of the square prediction error was used to evaluate the model (MSE). The experiments were repeated 50 times each. Table 1 summarizes the results; as evident from the results, our method does not incur significant loss in accuracy.

Table 1. Performance results for the 4 datasets used.

Dataset	Mean (SD)	m [a]	p [b]	N (values of n)[c]	MSE[d]		R^2		MSE ratio[e]
					CA[f]	SMA[g]	CA	SMA	
Student performance in Portuguese[h]	11.91 (3.23)	2	30	649 (423, 226)	3.364	3.417	0.68	0.675	0.984
Student performance in Math[h]	10.41 (4.58)	2	30	395 (349, 46)	7.554	7.719	0.627	0.62	0.978
AutoMPG[i]	23.45 (7.80)	3	6	392 (245, 68, 79)	13.56	17.563	0.778	0.711	0.77
Diabetes (with weight)[j]	4.848 (3.11)	3	39	267 (129, 72, 66)	8.801	8.59	0.09	0.108	1.025
Diabetes (with weight)[j]	4.41 (3.02)	6	39	456 (68, 130, 57, 73, 55, 73)	7.443	7.733	0.19	0.157	0.962
Diabetes (without weight)[j]	4.39 (3.01)	3	38	8567 (2478, 3936, 2153)	5.558	5.612	0.309	0.303	0.99
Diabetes (without weight)[j]	4.42 (3.00)	6	38	13626 (2478, 3936, 1480, 2153, 2108, 1471)	5.708	5.798	0.345	0.335	0.984
Diabetes (without weight)[j]	4.39 (2.97)	12	38	21205 (2478, 3936, 1480, 2153, 2108, 1471, 1160, 1323, 1524, 1425, 1024, 1122)	5.705	5.87	0.345	0.336	0.971

[a]m: number of sites.

[b]p: number of features.

[c]N (values of n): total number of records and their division along different sites.

[d]MSE: mean square error.

[e]MSE ratio=MSE CA/MSE SMA.

[f]CA: central algorithm.

[g]SMA: secure multiparty linear regression algorithm.

[h]Outcome variable: grade out of 20.

[i]Outcome variable: fuel consumption (miles per gallon).

[j]Outcome variable: length of stay (days).

Synthetic Dataset

Using synthetic data, we performed a scalability analysis to evaluate the time performance of the proposed solution as the data size and the number of parties increase. The synthetic datasets were generated in Python using sklearn.datasets.make_regression. The number of records was varied between 10^5 and 10^8, the number of features between 2 and 50, and the number of sites between 2 and 20. The records were always divided equally between the sites. We distinguished between 2 testing strategies: n fixed (Figures 3-5) and N fixed (Figure 6). The algorithm was compared with the CA (where data are shared in the clear) as there exists no other secure linear regression algorithm that performs model selection.

For the fixed n strategy (with, $p \ll n$), Figures 3 and 4 show the time performance of CA versus SMA as a function of the sample size N. Note that $m=N/n$ is also growing ($m \in [2,20]$ in Figure 3, and $m \in [2,10]$ in Figure 4). As seen from the figures, for large n and p, SMA is scalable, and the security overhead does not affect its performance significantly. Figure 5 shows the time

performance of SMA (with n=10 million) as a function of (1) N (left side) and (2) p (right side). Note that N varies between 20 million and 100 million and that the time performance for N=100 million and p=50 features is under 20 min.

For the fixed N strategy (with $p \ll N$) Figure 6 shows the time performance of SMA as a function of the number of sites (with p=50). Note that the time taken by the CA is constant whereas for the SMA, as the number of sites increases, the time taken by the algorithm decreases. It is important to note that when the number of records per site becomes very small, the communication cost increases, driving the overall computation time with it.

Figure 3. Time performance for central algorithm versus secure multiparty linear regression algorithm (SMA) with 100,000 records per site and varying feature set. As the number of sites increases, the number of records also increases. For small datasets, the time taken by SMA is more than that taken by the central algorithm (CA). This is due to the encrypted communication required by the algorithm. As the number of records and features increases, the time taken by the CA increases rapidly (at 1,500,000 records and 100 features). *n*: number of records per site; *p*: number of features.

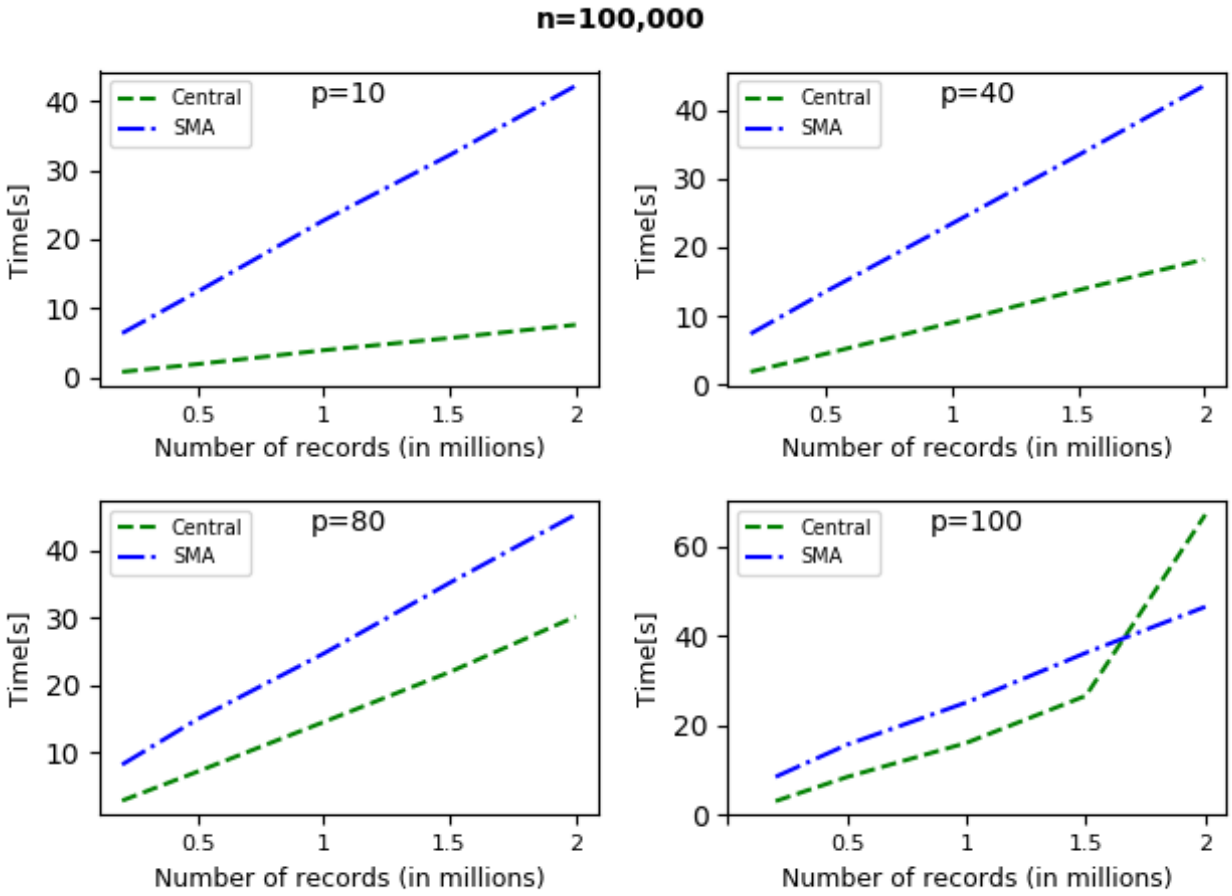


Figure 4. Time performance for central algorithm versus secure multiparty linear regression algorithm (SMA) with 1,000,000 records per site and varying feature set. The time taken by the central algorithm (CA) is greater than that taken by the SMA. For 10 million records, the SMA algorithm takes almost 30 seconds, whereas the CA takes around 18 minutes. *n*: number of records per site; *p*: number of features.

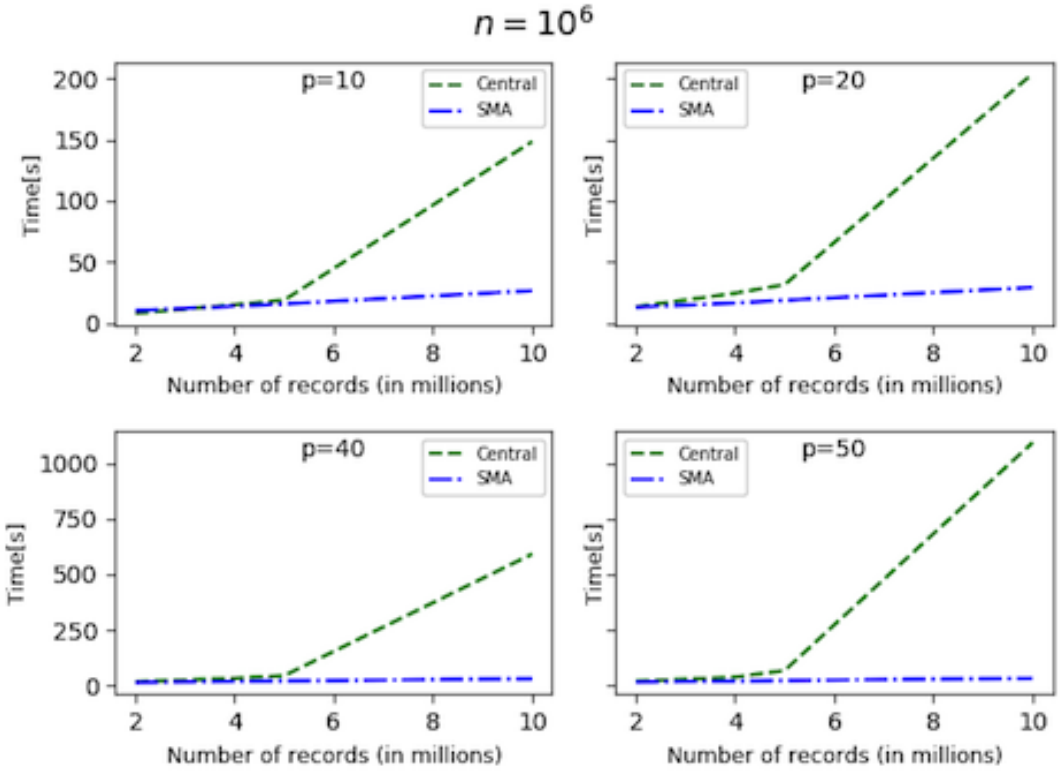


Figure 5. Time performance for secure multiparty linear regression algorithm (SMA) with 10 million records per site and varying feature set. The panel on the left shows the time as a function of the number of features, while the panel on the right shows the time as a function of the number of sites. Note that for *N*=100 million and *p*=50 features, SMA required 20 minutes for execution. *m*: number of sites; *n*: number of records per site; *p*: number of features.

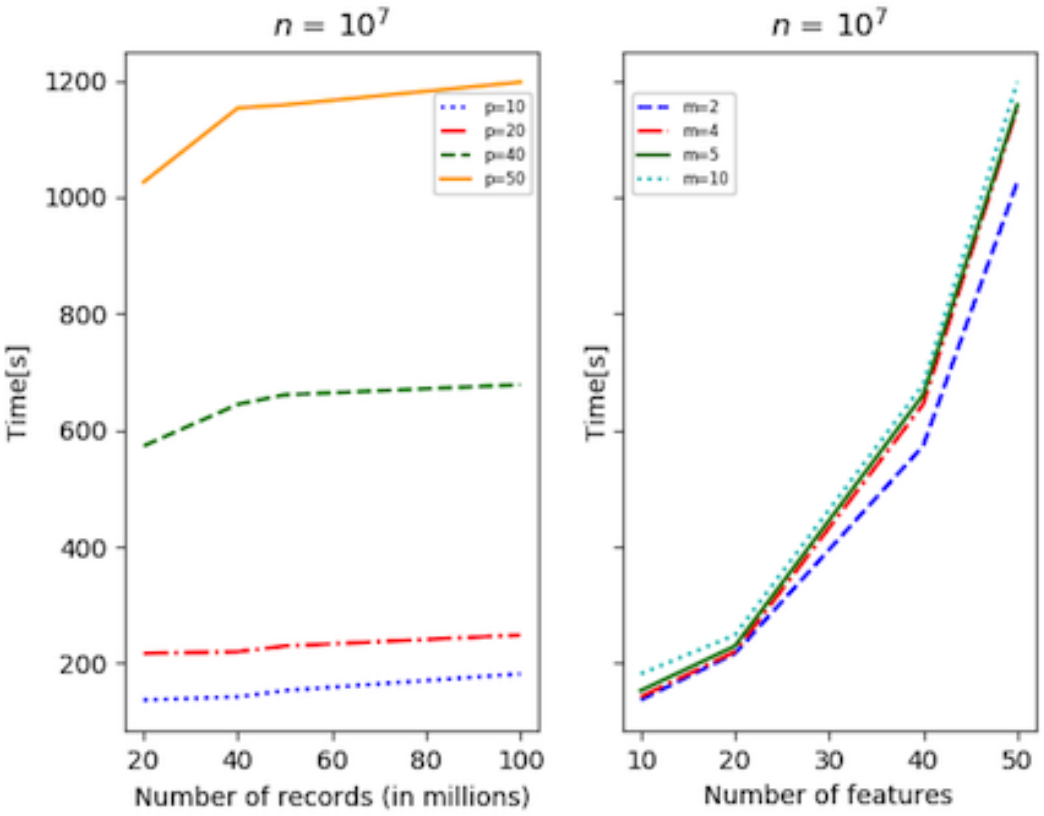
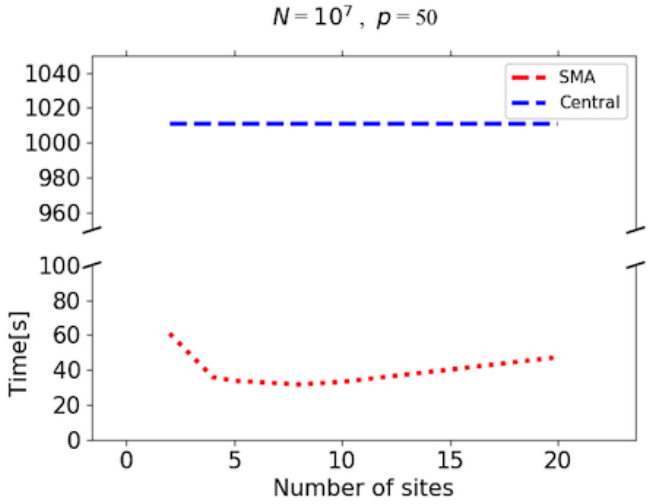


Figure 6. Time performance for central algorithm versus secure multiparty linear regression algorithm (SMA) with total number of records (*N*)=10 million, and features (*p*)=50. For SMA, the records are divided among a varying number of sites (2 to 20). The time taken by the central algorithm (CA) is constant. For SMA, time decreases with the increase in the number of sites, until it reaches *m*=20 (or *n*=50,000). At that point, the communication cost increases and the computation time starts to go up.



Discussion

This study introduced a model for privacy-preserving distributed data mining in which local models are produced separately and SMC is used to aggregate the results privately. Theoretical results from statistical theory were used to design the first secure multiparty linear regression model that does model selection and parameter estimation.

In general, theoretical results from statistical computing say that the averaged local estimates are as accurate as the centralized estimates when $n>p$ and $m<n$. In line with the theoretical results, we conducted computations on the distributed sites independently and then combined the results securely to form 1 estimator for the collective dataset. Experiments were conducted with $n \gg p$ and $m < n$ and they showed the accuracy (using 4 real datasets) and efficiency (using synthetic data) of the algorithm.

The experiments on synthetic data showed very good time performance. When *n* is fixed, as *N* increases, the time taken by the CA increases at a much faster rate than SMA. For big *N* (10^8), the algorithm does model selection, and parameter estimation in under 20 min (the algorithm of [36] does only parameter estimation in the range of 8 hours).

Much of the existing theoretical work in DSC assumes a uniform and random distribution of samples across the different sites or that the *m* independent sites have *n* observations each that are independent and identically distributed according to the same distribution *P*.

This assumption certainly facilitates the mathematical analysis but may not be realistic for some applications. In the SMC applications, data are collected and owned by the different sites and may thus have systematic differences across these different sites, in which case, the assumption could be overly restrictive. Redistributing the samples randomly across the different sites is not an option due to data privacy issues. However, it is worth noting that our experiments on real data (although limited due to lack of access to real data) showed high accuracy compared with the central case. In the future, we intend to relax these assumptions and study their theoretical effect on the accuracy of the results.

Another limitation is the assumption of horizontal distribution among the different sites which should be generalized to vertical divisions (or both).

Moreover, the study demonstrated the theoretical results using a linear model. We plan to extend our results to other estimators such as ridge regression and logistic regression.

Acknowledgments

This work is supported by the United Arab Emirates University research grant #31T084 Research Start-up. The authors would like to thank the anonymous reviewers for their helpful comments and suggestions on earlier drafts of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Secure multiparty algorithm.

[PDF File (Adobe PDF File), 94KB - medinform_v7i2e12702_app1.pdf]

Multimedia Appendix 2

Real datasets.

[PDF File (Adobe PDF File), 53KB - medinform_v7i2e12702_app2.pdf]

References

1. Partners Healthcare. Personalized Genetic Medicine URL: https://www.partners.org/Innovation-And-Leadership/Innovative-Medical-Discovery/Personalized-Genetic-Medicine.aspx [accessed 2019-03-07] [WebCite Cache ID 76hTkCqq0]
2. Dankar FK, Ptitsyn A, Dankar SK. The development of large-scale de-identified biomedical databases in the age of genomics-principles and challenges. Hum Genomics 2018 Dec 10;12(1):19 [FREE Full text] [doi: 10.1186/s40246-018-0147-5] [Medline: 29636096]
3. Beimel A, Nissim K, Omri E. Distributed Private Data Analysis: Simultaneously Solving How and What. : Springer; 2008 Aug 17 Presented at: Annual International Cryptology Conference; 2008; Santa Barbara, CA, USA p. 451-468.
4. Nordholt P, Volgushev N, Fauzi P, Orlandi C, Scholl P, Simkin M, et al. SODA. 2017. State of the Art Analysis of MPC Techniques and Frameworks URL: https://www.soda-project.eu/wp-content/uploads/2017/02/SODA-D1.1-WP1-State-of-the-art.pdf [accessed 2019-03-07] [WebCite Cache ID 76hU8AvP0]
5. Papadimitriou A, Narayan A, Haeberlen A. DStress: Efficient differentially private computations on distributed data. 2017 Apr 23 Presented at: Proceedings of the Twelfth European Conference on Computer Systems; 2017; Belgrade, Serbia p. 560-574.
6. Ishai Y, Mittal M, Ostrovsky R. On the Message Complexity of Secure Multiparty Computation. 2018 Mar 25 Presented at: 21st IACR International Workshop on Public Key Cryptography; 2018; Rio de Janeiro, Brazil p. 698-711.
7. Hirt M, Maurer U, Przydatek B. Efficient Secure Multi-party Computation. : Springer Berlin Heidelberg; 2000 Dec 03 Presented at: 6th International Conference on the Theory and Application of Cryptology and Information Security; 2000; Kyoto, Japan p. 143-161.
8. Dankar F, Brien R, Adams C, Matwin S. Secure Multi-Party linear Regression. 2014 Mar 28 Presented at: EDBT/ICDT Workshops; 2014; Athens, Greece p. 406-414.
9. Boyle E, Jain A, Prabhakaran M. The Bottleneck Complexity of Secure Multiparty Computation. 2018 Jul 09 Presented at: 45th International Colloquium on Automata, Languages, and Programming; 2018; Prague, Czech Republic p. 1-16.
10. Lindell Y, Pinkas B. Secure Multiparty Computation for Privacy-Preserving Data Mining. J Priv Confidentiality 2009 Apr 01;1(1) [FREE Full text] [doi: 10.29012/jpc.v1i1.566]
11. Lee PL, West C, Crain K, Wang L. Genetic polymorphisms and susceptibility to lung disease. J Negat Results Biomed 2006 Apr 11;5:5 [FREE Full text] [doi: 10.1186/1477-5751-5-5] [Medline: 16608528]
12. Beimel A. Secret-sharing schemes: a survey. 2011 May 30 Presented at: International Conference on Coding and Cryptology; 2011; Qingdao, China p. 11-46 URL: http://link.springer.com/chapter/10.1007/978-3-642-20901-7_2
13. Kilian J. Founding crytpography on oblivious transfer. 2008 Aug 17 Presented at: 28th Annual International Cryptology Conference; 2008; Santa Barbara, CA, USA p. 20-31.
14. Bellare M, Hoang V, Rogaway P. Foundations of garbled circuits. 2012 Oct 16 Presented at: ACM conference on Computer and communications security; 2012; Raleigh, North Carolina, USA p. 784-796.
15. Hall R, Fienberg S, Nardi Y. Secure multiple linear regression based on homomorphic encryption. J Off Stat 2011;27(4):669.
16. Dwork C. Differential privacy: a survey of results. 2008 Apr 25 Presented at: 5th International Conference on Theory and Applications of Models of Computation; 2008; Xi'an, China p. 1-19.
17. Mcdonald R, Mohri M, Silberman N, Walker D, Mann G. Efficient large-scale distributed training of conditional maximum entropy models. 2009 Dec 07 Presented at: 22nd International Conference on Neural Information Processing Systems; December 07-10, 2009; Vancouver, British Columbia, Canada p. 1231-1239.
18. Rosenblatt J, Nadler B. On the optimality of averaging in distributed statistical learning. Inf Inference 2016 Jun 09;5(4):379-404. [doi: 10.1093/imaiai/iaw013]
19. Minsker S. Geometric median and robust estimation in Banach spaces. Bernoulli 2015;21(4):2308-2335.
20. Zhang Y, Duchi J, Wainwright M. Divide and conquer kernel ridge regression: a distributed algorithm with minimax optimal rates. J Mach Learn Res 2015;16:3299-3340 [FREE Full text]
21. Liu Q, Ihler A. Distributed estimation, information loss and exponential families. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. 2014 Presented at: NIPS'14; December 08-13, 2014; Montreal, Canada p. 1098-1106.
22. Kim K. Higher order bias correcting moment equation for M-estimation and its higher order efficiency. Econometrics 2016;4:48. [doi: 10.3390/econometrics4040048]
23. Battey H, Fan J, Liu H, Lu J, Zhu Z. Princeton. 2015 Sep 17. Distributed estimation and inference with statistical guarantees URL: https://arxiv.org/pdf/1509.05457.pdf [accessed 2019-03-11] [WebCite Cache ID 76mxxMYmk]
24. Lee J, Liu Q, Sun Y, Taylor J. Communication-efficient sparse regression. J Mach Learn Res 2017;18:1-30.

25. Shamir O, Srebro N, Zhang T. Communication-efficient distributed optimization using an approximate newton-type method. In: Proceedings of the 31st International Conference on International Conference on Machine Learning. 2014 Jan 27 Presented at: ICML'14; June 21-26, 2014; Beijing, China p. 1000-1008.
26. Donoho D, Montanari A. High dimensional robust m-estimation: asymptotic variance via approximate message passing. Probab Theory Relat Fields 2016 Dec 01;166(3-4):935-969. [doi: 10.1007/s00440-015-0675-z]
27. Barbieri M, Berger J. Optimal predictive model selection. Ann Stat 2004;32(3):870-897.
28. Wang X, Peng P, Dunson D. Median Selection Subset Aggregation for Parallel Inference. 2014 Dec 08 Presented at: Advances in Neural Information Processing Systems 27; December 08-13, 2014; Montreal, Canada p. 2195-2203.
29. Montgomery D, Peck E, Vining G. Introduction to linear regression analysis internet. USA: John Wiley & Sons; Apr 09, 2012.
30. Seiler C. On the linearity of nature or the risk of extending regression lines beyond the observed data. Eur J Clin Invest 2011 Nov;41(11):1259-1260. [doi: 10.1111/j.1365-2362.2011.02527.x] [Medline: 21985420]
31. Du W, Han Y, Chen S. Privacy-preserving multivariate statistical analysis: linear regression and classification. 2004 Apr 22 Presented at: Proceedings of the Fourth SIAM International Conference on Data Mining; April 22-24, 2004; Florida, USA p. 222-233. [doi: 10.1137/1.9781611972740.21]
32. Karr A, Lin X, Sanil A, Reiter J. Secure regression on distributed databases. J Comput Graph Stat 2005 Jun;14(2):263-279 [FREE Full text] [doi: 10.1198/106186005X47714]
33. Aono Y, Hayashi T, Phong L, Wang L. National Institute of Information and Communications Technology. 2015. Fast and Secure Linear Regression and Biometric Authentication with Security Update URL: https://eprint.iacr.org/2015/692.pdf [accessed 2019-03-11] [WebCite Cache ID 76mz0vvZP]
34. El Emam K, Samet S, Arbuckle L, Tamblyn R, Earle C, Kantarcioglu M. A secure distributed logistic regression protocol for the detection of rare adverse drug events. J Am Med Inform Assoc 2013 May 01;20(3):453-461 [FREE Full text] [doi: 10.1136/amiajnl-2011-000735] [Medline: 22871397]
35. Dankar F. Privacy preserving linear regression on distributed databases. Transactions on Data Privacy 2015;8(1):3-28.
36. Nikolaenko V, Weinsberg U, Ioannidis S, Joye M, Boneh D, Taft N. Privacy-Preserving Ridge Regression on Hundreds of Millions of Records. 2013 May 22 Presented at: 2013 IEEE Symposium on Security and Privacy; May 19-22, 2013; Berkeley, CA, USA. [doi: 10.1109/SP.2013.30]
37. Cock MD, Dowsley R, Nascimento A, Newman S. Fast, Privacy Preserving Linear Regression over Distributed Datasets Based on Pre-Distributed Data. USA: ACM; 2015 Oct 16 Presented at: Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security; 2015; Denver, Colorado, USA p. 3-14. [doi: 10.1145/2808769.2808774]
38. Beaver D. One-time tables for two-party computation. 2002 Aug 12 Presented at: 4th Annual International Computing and Combinatorics Conference; 1998; Taipei, Taiwan p. 361-370.
39. Fox J. Applied Regression Analysis, Linear Models, and Related Methods. USA: Sage Publications, Inc; 1997.

Abbreviations

CA: central algorithm
DSC: distributed statistical computing
MSE: mean square error
SMA: secure multiparty linear regression algorithm
SMC: secure multiparty computation

Edited by C Lovis; submitted 06.11.18; peer-reviewed by L Arbuckle, G Sakellaropoulos; comments to author 27.12.18; revised version received 14.01.19; accepted 15.02.19; published 29.04.19.

Please cite as:
Dankar FK, Madathil N, Dankar SK, Boughorbel S
Privacy-Preserving Analysis of Distributed Biomedical Data: Designing Efficient and Secure Multiparty Computations Using Distributed Statistical Learning Theory
JMIR Med Inform 2019;7(2):e12702
URL: http://medinform.jmir.org/2019/2/e12702/
doi:10.2196/12702
PMID:31033449

Viewpoint

Consumer-Mediated Data Exchange for Research: Current State of US Law, Technology, and Trust

Yiscah Bracha[1*], MS, PhD; Jacqueline Bagwell[1*], MS, MMCi; Robert Furberg[1*], PhD, MBA; Jonathan S Wald[2*], MD, MPH, FACMI

[1]Division of eHealth Quality and Analytics, RTI International, Research Triangle Park, NC, United States

[2]InterSystems, Cambridge, MA, United States

[*]all authors contributed equally

Corresponding Author:
Yiscah Bracha, MS, PhD
Division of eHealth Quality and Analytics
RTI International
3040 E Cornwallis Rd
PO Box 12194
Research Triangle Park, NC, 27709-2194
United States
Phone: 1 929 741 7255
Email: yiscah.bracha@outlook.com

Abstract

A compendium of US laws and regulations offers increasingly strong support for the concept that researchers can acquire the electronic health record data that their studies need directly from the study participants using technologies and processes called consumer-mediated data exchange. This data acquisition method is particularly valuable for studies that need complete longitudinal electronic records for all their study participants who individually and collectively receive care from multiple providers in the United States. In such studies, it is logistically infeasible for the researcher to receive necessary data directly from each provider, including providers who may not have the capability, capacity, or interest in supporting research. This paper is a tutorial to inform the researcher who faces these data acquisition challenges about the opportunities offered by consumer-mediated data exchange. It outlines 2 approaches and reviews the current state of provider- and consumer-facing technologies that are necessary to support each approach. For one approach, the technology is developed and estimated to be widely available but could raise trust concerns among research organizations or their institutional review boards because of the current state of US law applicable to consumer-facing technologies. For the other approach, which does not elicit the same trust concerns, the necessary technology is emerging and a pilot is underway. After reading this paper, the researcher who has not been following these developments should have a good understanding of the legal, regulatory, technology, and trust issues surrounding consumer-mediated data exchange for research, with an awareness of what is potentially possible now, what is not possible now, and what could change in the future. The researcher interested in trying consumer-mediated data exchange will also be able to anticipate and respond to an anticipated barrier: the trust concerns that their own organizations could raise.

(JMIR Med Inform 2019;7(2):e12348) doi:10.2196/12348

KEYWORDS

health records, personal; electronic health records; patient access to records; research; trust; data collection; consumer health information

Introduction

How Researchers Now Acquire Electronic Health Records

Researchers who need electronic health record (EHR) data [1] for study participants receiving care in the United States commonly acquire them from health care providers who agree to extract the data from their EHR databases. This data acquisition method is convenient and suitable for studies that need data only from participants who receive care from cooperative providers. However, the method is pragmatically infeasible for studies that require complete longitudinal records from all providers caring for participants who individually and collectively receive care from multiple providers. Examples include the National Institutes of Health All of Us Research

Program (formerly the Precision Medicine Initiative) [2] and studies in which participants are patient members of Patient-Powered Research Networks [3]. Researchers could not feasibly establish and maintain data-sharing relationships with all the geographically dispersed providers who could be caring for all participants [4-7], including providers who may not have the capability, capacity, and/or interest in supporting research [8].

Theoretically, researchers could overcome these logistical challenges by obtaining comprehensive longitudinal electronic patient records from a clinical data research network, a clinical data repository, and/or a health information exchange, but practically, both convenience and data completeness challenges remain [9,10]. Currently, these types of organizations assemble records for patients who receive care within a specific geographic region, limiting utility for studies that enroll participants nationally. In the future, the National Trusted Data Exchange Framework (Exchange) may compile records across geographic areas, but this does not help researchers who need data now and will not help researchers who are unable to access the Exchange. Even if completeness and access challenges were resolved, there still is the challenge of linking records for the same patient across multiple providers [11,12].

Consumer-Mediated Data Exchange as a Data Acquisition Alternative

Consumer-mediated data exchange [13,14] may offer researchers a way to acquire the EHR data they need without confronting these logistical barriers. There are 2 approaches. In one, which we call *Download and Send*, study participants use a consumer-facing app to download and aggregate their own health records, which they then contribute to the research database. In the other, which we call *Transmit*, study participants use an app that directs their providers to transmit their data to the research database.

Legal Support for Consumer-Mediated Data Exchange

There has been strong and enduring US federal support for the principle that individuals should have access to their own EHRs and that individuals can share their records with third parties, including researchers [15-18]. Legal support was first codified through a 2011 amendment to the Health Insurance Portability and Accountability Act (HIPAA), which requires US providers and health plans to fulfill patient requests for their own data and to provide the data in electronic form if the patient requests [19]. The 21st Century Cures Act states that consumers must be able to access their own electronic health information "with no special effort" [20]. The current administration's MyHealtheData initiative is predicated on the belief that all individuals should have access to their electronic health information and be empowered to use them however they wish [21,22].

To implement these principles, the Centers for Medicare and Medicaid Services (CMS) had incentivized providers to give patients the ability to view, download, and transmit their own data electronically, originally through the Meaningful Use program [23] and, now, through the Hospital Inpatient Prospective Payment Systems [24] and for every model within the Quality Payment Program, including the Merit-Based Incentive Payment Systems (MIPS), and advanced alternative payment models [25]. As of 2019, CMS requires providers participating in these programs to use health information technology that meets the most recent (2015) certification criteria established by the Office of the National Coordinator for Health Information Technology (ONC). These criteria not only require the technology to allow the patient to view, download, and transmit their own records manually, they also require the use of *open application programming interfaces* (APIs), which allow different technologies to exchange information with each other. APIs offer the technical foundation for patient access to their own electronic health information without special effort.

CMS and ONC took another major step in February 2019 when both agencies released notices of proposed rulemaking, on the same day, intended to accelerate the interoperability of electronic health information in the United States by leveraging consumer-mediated data exchange [26,27]. CMS proposes regulations that would give all publicly insured consumers access to their own claims data. The ONC proposes to change some of the criteria that certify the APIs used in provider-facing technology, where the proposed changes would make it much easier for consumers to access and use their own EHR data with the assistance of any consumer-facing app. Both proposed rules are intended to promote interoperability, and both agencies consider consumer-mediated data exchange to be a linchpin in that effort [28,29]. Although the major objective of interoperability is to improve care and outcomes at a lower cost, federal pronouncements also reference researchers' enhanced ability to acquire the data they need.

In summary, in the United States, there is strong legal and regulatory support for the principle that consumers have a right to access their own electronic health information and that consumers can use their data however they wish, including contributing them to a research database.

Potential Barriers Researchers May Encounter When Using Consumer-Mediated Data Exchange

For consumer-mediated data exchange for research to be viable at scale, the following conditions must exist:

Technical

All study participants must receive care from providers who have the technical capability to respond to patient requests either to download their own data (approach 1: Download and Send) or to direct their own data to a researcher (approach 2: Transmit). In addition, there must be consumer-facing technologies to support the participant.

Utility

The data obtained through this method must be useful for the study.

Trust

The participants, providers, researcher's organization and Institutional Review Board (IRB) must have any potential trust concerns allayed. This paper focuses on trust concerns that the researcher's organization or IRB could raise, because if they

are not addressed, they may refuse to approve the study. (If providers have trust concerns, they will not make this service available to their patients, which for this paper is indistinguishable from a technical barrier. If prospective participants have trust concerns, they would decline to participate, and presumably, there would be others without such concerns for the study to proceed).

Structure of This Paper

This paper first reviews the contemporary technology landscape to assess the level of technical support for both approaches to consumer-mediated data exchange, along with the types of data that researchers could expect to receive. The paper does not address data completeness, provenance, harmonization, and other utility barriers in using EHR data for research, as these have been documented elsewhere [1,8,30-33].

The paper will explain why technical viability in theory is now strong for the *Download and Send* approach and weak but rapidly growing stronger for the *Transmit* approach. The paper outlines the trust concerns that could arise with *Download and Send* and offers suggestions to researchers for responding to them. The paper then describes regulatory and technology developments currently underway that should make the *Transmit* approach more viable in the future, which would mitigate many trust concerns that currently exist.

The paper concludes with a summary, limitations, and a review of potential future developments.

Technical Support for Consumer-Mediated Data Exchange

Prevalence Among Providers of Necessary Provider-Facing Technology

Approach 1: Download and Send

For the *Download and Send* approach to be viable, all US providers must be able to give patients the ability to download their records electronically, if the patient requests. This means that, at a minimum, providers have the necessary technology to recognize and respond to a patient-initiated *download* request.

The ONC had estimated that, in 2015, 87% of hospitals and 41% of office-based physicians gave patients the ability to download their own medical records [34,35], representing a sharp and steady increase from previous years. In 2012, only 14% of hospitals reported having this ability [35,36]. The change in prevalence of download capabilities for office-based physicians is less clear as, in 2013, the ONC reported that 33% offered patients all 3 *view, download, and transmit* capabilities [37], without reporting the percentage that only had download capabilities.

Since the 2015 reports, estimates of provider prevalence with capabilities have continued to increase. In 2018, the ONC reported [38] that among MIPS-eligible hospitals and clinicians, more than 90% of hospitals and more than 80% of US clinicians were using health information technology systems, meeting the most recent (2015) certification requirements [39]. These requirements include giving patients the ability to download their own records, where the data types must include those in the Common Clinical Data Set [40,41].

In a 2017 report to the Congress, the US Government Accountability Office reported that most providers achieved the view, download, and transmit functionality through patient portals [42]: The patient logs into the portal with a username and password and from the portal clicks a *download* button. The patient can download a PDF of their data to a drive, or if they are being assisted by a personal health records app (more below), they can download the data to the app which may offer added services such as configuring the data for visualization and manipulation. From a configured portal, the patient could also click a *transmit* button to send records to another provider, the approach to be considered next.

On the basis of ONC estimates, it is likely that most, if not all, US providers either currently (March 2019) have the *download and send* capability or will soon have it. On the basis of the Government Accountability Office report, it is likely that, through 2016, providers offered these capabilities through patient portals. We have discussed the use of APIs as an alternative technology in more detail below.

Approach 2: Transmit

For the *Transmit* approach to be viable, all US providers must give patients the ability to electronically direct their providers to transmit their data to a designated recipient.

In 2015, the ONC reported that 66% of hospitals [43] and 19% of physician offices [35] had the technical ability to transmit patient records to other providers; the ONC did not comment upon providers' ability to transmit records to nonproviders such as researchers. We expect the number of providers with transmit capabilities to increase because of the rules that CMS finalized in the fall of 2018: as of 2019, all providers participating in CMS value-based purchasing programs must use health information technology meeting ONC's most recent (2015) certification requirements [25]. Such technology not only gives patients the ability to download their records but it also gives them the ability to direct their providers to *transmit* their records to a third party, where the transmission can occur using secure Direct Messaging, or by email if the patient requests [39].

Transmission by insecure email would likely be unacceptable to researchers. Transmission by Direct Messaging, however, requires the electronic address of the recipient, which the patient must know and be able to share with the transmitting provider.

This is problematic for usability. Hospitals' most commonly reported barrier to electronic information exchange was difficulty locating the electronic addresses of the recipient [44], a finding supported by an ONC user experience study [45]. Ideally, there would be a national directory that consumers and providers could search to locate the recipient address and just click to make it happen, and ideally, researchers would be able to enter their study names and addresses into this directory.

Provider directories containing digital contact information exist in both the private and public domains. The nonprofit organization, DirectTrust, maintains a directory [46], which we

will not discuss further because there is no federal agency with authority over its structure or contents.

Publicly, CMS maintains a directory called the National Plan and Provider Enumeration System (NPPES) [47-49], originally established as a mechanism for HIPAA-covered entities to exchange information with each other. CMS assigns a unique numeric National Provider Identifier to every individual clinician and facility submitting claims to CMS, and these numeric identifiers represent the provider entries in the NPPES. Provider entries contain name, taxonomy code and description, mailing address and practice address. In response to requirements in the 21st Century Cures Act, CMS modified the NPPES [47-49] to accept digital contact information called *endpoint identifiers* [50,51]. However, not all providers list their endpoint identifiers in the NPPES. In its February 2019 proposed rule, CMS proposes to address this problem by creating and publicly disseminating a list of providers *without* endpoint identifiers in the NPPES, to exert publish pressure upon them.

With respect to researcher inclusion in the NPPES, CMS directions for obtaining an identifier, and for becoming listed in the NPPES, convey through use of examples such as "physician" and "hospital" that CMS uses the word "provider" to refer to an individual or organization that delivers medical care. However, some individuals and organizations now listed in the NPPES have research-related taxonomy codes. Due to the novelty of consumer-mediated data exchange, we doubt that these providers are receiving consumer-directed electronic health information. Rather, we believe that their presence in the NPPES facilitates administrative mechanisms required to charge a research study when its participants receive clinical care. However, the fact that research-related individuals and organizations already appear in the NPPES demonstrates that researchers interested in receiving consumer-directed electronic health information do have a mechanism for listing themselves in this public directory.

Usability challenges do remain but could be addressed by the ONC. For example, the ONC already publishes a Patient Engagement Playbook [52] for providers. Chapter 2 offers providers a step-by-step guide for setting up a user-friendly patient portal that patients can use to download their health information. The Playbook is now silent about how providers can set up their portals to enable patients to *transmit* their health information, but it could be amended to include this information. Not only would this facilitate consumer-directed interoperability, it also would facilitate consumer-directed data exchange for research.

This review suggests that although provider-facing technology should support transmit functionality, usability is a barrier, particularly for consumer attempts to direct the transmission of data using portal-based technology. It is possible that these usability issues will be addressed, and it also is possible that the introduction and use of new technology will address usability issues in other ways. Between 2015 and 2019, technology evolved through the maturation and testing of standards that were in development at the time that ONC finalized its 2015 certification requirements. The emerging technology rapidly gaining adoption involves the exchange of electronic health information through *FHIR-based APIs*, using the *OAuth 2.0* authentication protocol. *FHIR* means *Fast Healthcare Interoperability Resources*; this is a relatively new standard for exchanging health care information electronically [53-56]. As noted earlier, an *API* is a technology that permits 2 applications to communicate or exchange electronic information between them [57]. *OAuth 2.0* allows consumers to give their apps access to their private health information without exposing their login credentials, and thus, creating an additional layer of security.

When both provider- and consumer-facing technologies use the FHIR APIs, the OAuth 2.0-protected consumer can easily access his or her data from the provider and then transmit the data to a target of the user's choice. In theory, the target could be another provider, to facilitate interoperability. The target could be the consumer's device, which is functionally equivalent to the *download* capability. The target could be a research database: the consumer would still need to know its digital endpoint, and presumably, the research team would give that to its study participants. In other words, consumers could use the new technology to direct their providers to transmit their data to a research database without the need for the researchers to list themselves in the NPPES or any other directory.

For this emerging technology to support research, all US providers must have technology that can support FHIR APIs with OAuth 2.0. The rule that ONC proposed in February 2019 [58,59] is intended to close the remaining gaps to achieve ubiquity. On their own, without regulatory requirements, many vendors in the industry have already adopted these standards. The ONC reports in both its proposed rule [58] and in a 2018 blog [60] that as of mid-September 2018, 51% of certified vendors were already using some version of FHIR APIs and OAuth 2.0 together. On the basis of these vendors' market shares, the ONC concluded that approximately 87% of hospitals and 57% of MIPS-eligible clinicians had technology with these capabilities. If ONC's proposed certification criteria for provider-facing APIs [61] are finalized, all vendors serving providers would have these capabilities and ubiquity among providers would be possible by 2021 or 2022.

In summary, to support patient care, most providers now have technology that supports *Transmit*, but *usage* of that technology to transmit electronic health information is underdeveloped. Social and organizational processes could change if all providers listed their digital endpoint identifiers in the NPPES, and once those processes emerged, they could be used to transmit data for research as well as for patient care, if researchers also presented their digital endpoint identifiers in the NPPES. However, researcher use of *Transmit* might be easier if their study participants use consumer-facing technology that leverages FHIR APIs, under the assumption that provider-facing technology uses FHIR APIs as well. The ONC's 2019 proposed rule is intended to ensure that all providers have FHIR API technology by 2021 or 2022, a very promising development.

Summary Regarding Prevalence Among Providers of Necessary Technology to Support Consumer-Mediated Data Exchange for Research

If ONC estimates are correct, it is likely that, as of early 2019, most US providers already have technology capable of supporting the *Download and Send* approach to consumer-mediated data exchange for research. In this approach, study participants would download their own data from all their providers and transfer the aggregated records to the research database.

It is likely that, in a few years, all providers will have technology that has been certified to support the *Transmit* approach to consumer-mediated data exchange for research, where the transmission method relies upon FHIR APIs.

Existence of Necessary Consumer-Facing Technology

Approach 1: Download and Send

The *Download and Send* approach requires the existence of a consumer-facing app that can collect and compile data from multiple providers and continue to collect these data automatically, if the consumer gives this direction. Such apps are often called *untethered* or *standalone* personal health records [62-65]. The word "untethered" means that they operate independently from a particular provider and/or EHR vendor (see Textbox 1). These apps are existent: *App stores* return numerous results when entering the words "personal health records" into the search fields. We call attention to some that are notable.

At least two vendors of consumer-facing apps, both relying on portal download technology, promote their research utility. The vendor Carebox boasts that its app can collect medical records automatically, once the user establishes the connection [66], and reports that the Leukemia and Lymphoma Society National Patient Registry [67] is using its technology to enable patients to download and aggregate their records for contribution to the registry. Hugo, a commercial app developed by the vendor Me2Health was specifically designed to enable consumers to download and aggregate their own EHRs and send them to researchers [68,69]. Investigators at the Yale Center for Clinical Investigation and the Mayo Clinic are conducting a pilot to explore Hugo's utility for postmarket surveillance of medical devices [70]. Both Carebox and Me2Health take responsibility for transferring the aggregated records to the researchers, using whatever transfer protocols the researcher requires. Both apps give users the ability to visualize their own data on the device, to automatically update with new records whenever they appear, and to sever the connection whenever they wish.

Consumer-facing apps that use FHIR APIs conceivably can access the consumers' electronic health information from providers and then send them to a designated target, which includes the consumer's own device, which is functionally equivalent to the *download* technology. In January 2017, the vendor PatientLink Enterprises won a 2016 ONC competition [71,72] for its product, MyLinks [73], which uses the FHIR API to aggregate data from multiple health care systems. A year later, in January 2018, Apple unveiled a beta version of a personal health record iPhone app using an FHIR API [74], an event that attracted a great deal of attention among analysts who watch the health information technology industry [75-78]. Data exchanged using Apple's app do not traverse Apple's network, a feature that Apple cites prominently on its website [79]. This is a key differentiator compared with portal download technology and is discussed in this paper in more detail below, in the section titled "Trust Concerns that Research Organizations or IRBs Could Raise."

Textbox 1. Personal health records: coevolution of technology and lexicon.

When patient portals were first introduced, they were called personal health records. Later, when consumer-facing applications that gathered data from multiple sources entered the market, these too were called personal health records. To differentiate between them, market analysts and some researchers began referring to the patient portal as a tethered personal health record and referring to the consumer-facing application as an untethered personal health record. As patient portals became more common, and included convenience functions such as making appointments online, paying bills, sending secure messages to providers, the vocabulary changed again, and writers simply referred to patient portals as "patient portal". Once the word "tethered" was less often used to refer to patient portals, its antonym, "untethered", was less often used to refer to consumer-facing applications that can compile data from multiple sources.

Approach 2: Transmit

None of the apps that we reviewed, which rely on portal technology, advertise any ability to help the user direct their providers to transmit their data to a third party, either other providers or a research database. Apps that use FHIR APIs could authorize providers' technology to transmit the data to other providers or to a research database. There are growing numbers of consumer-facing apps that use FHIR APIs and one has been designed to support data transmission for research. This is the app designed and built for the Sync for Science project, on behalf of the All of Us Research Program. We have discussed Sync for Science in more detail below, in the context of future possibilities.

Technology Summary

As shown in Table 1, as of March 2019, there is strong technical support for the *Download and Send* approach and limited but growing support for the *Transmit* approach. The next sections focus on the trust concerns associated with *Download and Send.*

Table 1. State of provider- and consumer-facing technology to support consumer-mediated data exchange for research, as of March 2019.

Approach	Provider-facing technology	Consumer-facing technology
Approach 1: Download and send	Likely ubiquitous or nearing ubiquity among US providers, using download capabilities from the patient portal.	Many consumer-facing applications on the market tha give users ability to download records and compile them. Two vendors are known to take responsibility for transmitting users' records to researchers, with user consent.
Approach 2: Transmit	Capabilities from the patient portal exist but insufficient use of digital contact information poses usability barriers. Many providers have technology that uses FHIR APIs[a], and this technology would support the approach, but there are social, business and usability obstacles. These could be removed within several years if ONC's[b] February 2019 proposed rule is finalized.	There are no indications that applications exist in the consumer-facing market that support "transmit" from the portal technology. A growing number of applications, with Apple as a market leader, use the FHIR API technology which could be deployed for "transmit". Sync-For-Science is a prominent pilot testing the technology for research use, but scale is currently limited to four EHR[c] vendors and approximately 12 providers. This holds great promise for the future, but current opportunities are limited.

[a]FHIR API: Fast Healthcare Interoperability Resources open application programming interface.

[b]ONC: Office of the National Coordinator for Health Information Technology.

[c]EHR: electronic health record.

Trust Concerns That Research Organizations or Institutional Review Boards Could Raise

Definition of "Trust"

As a social concept, *trust* conveys perceptions of safety, reliability, risk, and vulnerability across a wide range of contexts. It has been modeled as a 3-part function in which *A* (trustor) trusts *B* (trustee) to fulfil *C* (task) [80]. Within clinical information systems, technical representations of social trust facilitate the exchange of protected health information [81], such as credentialing for access control. Trust is necessary to facilitate participants' involvement in contemporary clinical research studies, such as the *All of Us Research* Program [82].

For this paper, there are 3 actors (participant, researcher, and vendor of the consumer-facing app), and thus, there are 3 trust relationships:

1. Relationship 1: *The study participant trusts the researcher to keep his or her data safe, and not allow the data to be used in any way other than what the participant intends.*
2. Relationship 2: *The researcher trusts the vendor of the consumer-facing app to not permit the data to be used in any way other than what the researcher intends*.
3. Relationship 3: *The study participant trusts the vendor of the consumer-facing app to not permit the data to be used in any way other than what the participant intends*.

All relationships could be violated if the vendor uses participant data in ways that neither the researcher nor the patient authorized [83].

Why Research Organizations and Institutional Review Boards May Not Trust Consumer-Mediated Data Exchange

In the most common current use of existing technology, which does not rely on FHIR-based APIs, consumers who use commercial personal health record apps to download their data from providers must give the app necessary credentials for accessing their data, such as a patient portal username and password. In addition, the vendor will be exposed to the data during the process of transferring them from the provider to the patient's device. The vendor will also be exposed if the vendor takes responsibility for transferring the records from the consumer's device to the research database. Thus, the developers of the consumer-facing apps have access to personal health information that they have the potential to abuse, raising concerns about the integrity of the trust relationships.

These concerns have been enduring [84]. A 2007 study commissioned by the ONC reviewed privacy policies of 30 vendors, finding that none had policies that named the vendors' data partners or other secondary data users nor did any of the policies explicitly describe what data elements might be shared [85]. Moreover, 5 years later, in 2012, newly published studies showed that personal health record apps available at the time frequently lacked basic security features, with highly variable privacy policies [86,87].

As there is little disagreement about the basic privacy and security protections to which personal health record apps should adhere [88], in 2018, the ONC disseminated an updated model privacy notice [89,90], encouraging vendors to adhere to it. The astute reader will notice the use of the word "encourage" rather than "require". This is because the ONC has no authority to *require* commercial vendors of consumer-facing products to comply with anything.

No other US federal agency has this authority either. The Office of Civil Rights and the Federal Trade Commission (FTC) are the 2 federal agencies with the closest authority. The Office of Civil Rights investigates every reported HIPAA violation and has the authority to impose fines or other punishments for violations. However, by definition, commercial vendors of untethered health record apps are not HIPAA-covered entities [91], and so they fall outside the authority of the Office of Civil Rights [92-94].

The FTC has authority to investigate complaints of consumer harm, although unlike the Office of Civil Rights, which investigates every reported HIPAA violation, the FTC chooses which complaints to investigate, typically on the basis of the magnitude of harm [86]. The FTC also has authority to impose punishments. However, the FTC only has authority to investigate if a vendor fails to follow its own privacy policies. It does not have the authority to require the vendor to have a policy nor can it dictate what the policy should be [86].

Will There Be Changes in US Law Anytime Soon?

Some believe that the FTC and the Office of Civil Rights already have the authority they need [95] and argue that until serious harms have been demonstrated, new legislation might stifle the innovation necessary in a time of rapidly evolving technology [96]. In particular, there has been opposition to expanding HIPAA's scope to cover personal health record app developers. Spokespersons from the Center for Democracy and Technology and others argue that HIPAA contains built-in limitations because of its original intent and that extensions of HIPAA to cover personal health record apps will be inadequate [91,96-99]. Although there are European [100,101] and other models [84] that US legislators and regulators could theoretically adopt, there have been no bills introduced in the US Congress that follow these leads. The closest example was the 2013 changes to the Omnibus HIPAA bill, which extended HIPAA to business associates of covered entities. However, most untethered personal health record apps do not seek formal business association with covered entities; they direct their attention to consumers instead.

As there are unlikely to be changes in US law that would address these concerns, there now are activities underway to better leverage the FTC's existing authority. The CARIN Alliance is a prominent multisector group within the health care industry that works collaboratively with US government leaders to promote consumers' ability to access their electronic health information via open APIs [102]. In November 2018, the CARIN Alliance published a *trust framework and voluntary code of conduct* directed toward developers and vendors of consumer-facing apps [103]. Under the trust framework, when app developers place their products on app stores, they would attest that they adhere to the CARIN Code (Code). The Code is based on the internationally recognized standards and practices for sharing consumer information, including the Code of Fair Information Practices. Developers who publicly attest to adhering to the Code, and then violate it, expose themselves to FTC accountability [104].

The press release announcing the framework and Code says:

> *For the first time, health care organizations and other organizations can have an enforceable code of conduct for third-party applications not covered by HIPAA to self-attest to in order to access health care data on behalf of consumers.*

The CARIN Code has stakeholder support from consumers and caregivers, health information networks, former regulators, app vendors, health care providers, medical home networks, and health plans [105]. The CARIN Alliance is actively working with developers to encourage them to adopt the Code as a part of their process of registering their apps and rolling them out to consumers.

Will the Office of the National Coordinator for Health Information Technology's Trusted Exchange Framework and Common Agreement (Exchange) Resolve Concerns?

We believe this is unlikely. The Exchange is intended to be the implementation of a provision of the 21st Century Cures Act which directs the ONC to create a framework and agreement for the exchange of electronic health information between health information networks. As noted above, in January 2018 the ONC released a draft of the Trusted Exchange Framework and Common Agreement [106]. This draft affirmed that consumers seeking their own data could access the Exchange with a commercial app as long as the app complied with Framework provisions. ONC reaffirmed these individual access provisions in its second draft, which it released in April 2019 [107], by creating a distinct category called "Individual Access Services", defined as services which enable individuals to access and obtain copies of their own electronic health information. In the second draft, ONC says that entities that wish to offer Individual Access Services and thus participate in the Exchange must agree to provisions that are aligned with HIPAA, even if they are not HIPAA-covered entities. This includes a requirement that the entity must publish its privacy practices, with a notice that reflects ONC's Model Privacy Notice [89].

However, these requirements that ONC would impose upon non-covered entities apply only to entities that wish to participate in the Exchange by offering Individual Access Services. ONC still does not impose any requirements upon non-covered entities that choose not to participate in the Exchange, including noncovered apps that help consumers access their data directly from their providers, rather than health information networks.

Potential Impact of Trust Erosion on Research Organizations and Institutional Review Boards

The Office of Civil Rights, in early 2016, issued strongly worded Guidance directed toward providers that (1) reaffirmed that providers are legally required to provide patients access to their own EHRs upon request and (2) stated that providers were not liable if, upon complying with a patient request, the patient subsequently placed the privacy and/or security of their own records at risk [108].

Research organizations undoubtedly would welcome comparable guidance that specifies liability if studies *receive* potentially exposed participants' private health information. However, at

the moment, there is no such guidance directed toward research organizations or their IRBs nor is there guidance regarding organizations' culpability if a researcher uses an unregulated commercial app to gather data from study participants.

This silence can provoke uncertainty for research organizations' legal and risk departments, and they may contemplate *what-if* scenarios: What if the app's vendor, a small start-up with few resources, has inadequate security and hackers gain access to participant data? What if the vendor sells consumers' access credentials and/or their data to a third party? A participant who learns of these violations could seek redress from the research organization, attempting to bring a civil suit for monetary damages or bringing a suit in the court of public opinion, placing the organization's reputation at risk. In addition, IRBs reviewing the study may question whether participants fully understand the risks associated with using unregulated commercial apps to manage and transfer their personal health information [87,109]. They may refuse to approve studies that use consumer-mediated data exchange because they believe that participants are incapable of granting truly informed consent.

The most likely source of guidance would come from the organization, *Public Responsibility in Medicine and Research* [110]. Its goals are to create a strong and vibrant community of ethics-minded research administration and oversight personnel and to provide educational and professional development opportunities to raise the bar of research administration beyond basic regulatory compliance. It also has formalized professional standards and is active in public policy, offering expert opinion to rule-making and advisory bodies governing the research enterprise. This organization may not now be aware of the potential offered to researchers by consumer-mediated data exchange, as well as its attendant trust concerns, as a January 2019 search in its Knowledge Center on the term "personal health records" returned no results.

In conclusion, research organizations or IRBs could be concerned about studies that acquire EHR data using unregulated vendors. The professional society, *Public Responsibility in Medicine and Research*, is likely to hear of these concerns when more researchers take interest in the promise offered by consumer-mediated data exchange. As it develops best practice guidance for research ethics given these new technology developments, the organization will be able to rely upon the CARIN Alliance trust framework, and its associated Code of Conduct, particularly as the framework and Code gains traction with app developers.

Researcher Options

Responses to Trust Concerns About the Download and Send Approach

The trust concerns associated with the *Download and Send* approach are focused on the integrity of the entity that manages the consumer-facing app that participants use to download their data from all their providers and then send the compiled record to the research database.

Option 1: Build Trust With a Commercial Developer

The research team could proactively build trusted relationships with commercial developers through the following mechanisms:

- Only consider vendors that have adopted the ONC's Model Privacy Notice [89] or the CARIN Voluntary Code of Conduct [103]. Developers who pledge to abide by the CARIN Code risk being held accountable by the FTC if they violate that pledge.
- Collaborate with the CARIN Alliance to implement the third phase of its Trust Framework, in which third parties certify apps and their vendors for their adherence to the Code and can also create other certification criteria. Research organizations, for example, could develop criteria related to the vendor's ability to support research needs, including—if applicable—its willingness and ability to securely transfer user data to the research team if the user consents, and to configure these data so that they are analytically digestible.
- Impose the organization's security and privacy requirements upon the vendor contractually, detailing the consequences to the vendor if the organization learns that the vendor has compromised participant data.
- Establish monitoring systems for the vendor's privacy policies and the movement of data through the vendor's system, intervening if there is suspicious activity.

Option 2: Build Your Own Personal Health Record App

A research organization could create its own app that its study participants could use, which would take the vendor out of the mix of trusted relationships, so that trust would only be established and maintained between the study participant and the research team. This increases the burden on the research team, which now has to build and maintain a consumer app. The burden is lower if the study is retrospective, only requiring participant records that exist at the time the study begins. The research team's burden increases if the study is prospective, meaning that new records must be obtained as they become available, which then means that the researcher will have to maintain the app over time and potentially release upgrades that are compatible with changes in the broader environment. The burden is particularly high if the researcher would like to use the app as an incentive to attract participants, offering them the personal data management capabilities that commercial vendors offer.

If the research team is not troubled by these potential burdens, then building and managing its own consumer-facing app could potentially mitigate trust concerns and allow the study to proceed. A faculty member at the Yale Center for Clinical Investigation did just that, which is how the Hugo app was developed [111].

Lower Expectations About the Comprehensiveness of the Data to Be Received

If neither of the options mentioned above are viable, the researcher could use an app that relies on the FHIR API technology so that the data can be transmitted from the provider to the user's device without traversing the vendor's network. Among the consumer-facing apps that say they use FHIR APIs,

Apple is the most well-known, and the ONC cited Apple by name in its 2019 proposed rule. However, using Apple could impose limitations on the study. Apple's app will only work with an iPhone, which means that the study can enroll only iPhone users. In addition, Apple's app only downloads data from providers who are members of Apple's partnership. The partnership has grown steadily since Apple first launched the app in January 2018 [112], and as of April 2019, there were more than 370 partnering providers [113]. The partners are diverse, including individual physicians, small specialty practices, and large systems such as Kaiser in Northern California. This is impressive growth, but the partnership still does not include all providers in the United States. So even if all study participants had iPhones, it is possible and probably likely that for each study participant, there would be incomplete data.

Watchful Waiting Until the Environment Becomes More Favorable

Researchers unable to work within the existing limitations could prepare themselves for a future that may hold one or more of the following:

Apple and Equivalents Enroll All US Providers in "the Partnership"

If Apple's partnership eventually includes all US providers, then limitations in data only coming from participating providers would be removed, leaving only the limitation of iPhone use. This limitation will also evaporate as other vendors develop and promote consumer personal health record products using FHIR APIs. In the future, all consumers may be able to download their electronic health data to their devices using FHIR APIs, and researchers could adapt to whatever device and app and prospective study the participant prefers.

This still would leave open the challenge of getting the data from the participants' devices to the research database. It is technically feasible for a consumer-facing research app to use the FHIR API standard to request and receive data from a consumer-facing personal health record app, and although we are unaware of apps that support this now, an enterprising developer could create one in the future.

The CARIN Alliance Trusted Framework Becomes Normative

If the CARIN Alliance succeeds in fostering its Trust Framework, and vendors of personal health records self-attest to adhering to its Code of Conduct, and symbols of this attestation appear on app stores and product labels, then the symbol could serve as a *trust flag* that helps consumers and others such as researchers differentiate between products on the basis of their adherence to internationally accepted norms regarding the sharing of consumer information.

The Federal Trade Commission Asserts an Intent to Investigate if Vendors Violate a Pledge to Adhere to the Code

The norms around vendor behavior would be strengthened if it were publicly known that the FTC will investigate vendors that violate pledges to adhere to the Code. Research organizations could then be assured that there would be externally imposed consequences upon vendors who fail to uphold the standards of conduct.

Emergent Technology Tuned for Research

In the *Transmit* approach to consumer-mediated data exchange, health care providers' technology send the study participant's patient data directly to a research database upon receiving an order to do so from the participant. This approach reduces, if not eliminates, trust concerns because there is no consumer-facing app that is exposed to the participants' private health information. If the approach relies on the Direct Messaging protocol from the patient portal, technical support may exist but usability is now weak. If the approach relies on the emerging technology of FHIR APIs coupled with OAuth 2.0 authentication, usability may soar, making the *Transmit* approach much more viable.

The Sync for Science Pilot

In the spring of 2016, the National Institutes of Health, in collaboration with the ONC, contracted with Harvard Medical School to create a standards-based open source technology framework called Sync for Science [114,115]. The vision was that EHR vendors could use the framework to augment their patient portals to be able to respond to consumer-facing research apps that ask the provider to transfer the patient's data to a designated research study. Apps that comply with Sync for Science requirements would not need patients' portal login credentials nor would they handle personal health information. Instead, the study participant would enter a provider-specific code into the app, which would then route the participant to the provider's compliant patient portal, which would be able to recognize and act upon data transfer requests.

After several years of design, development, and consultation with IRBs, in September 2018, a pilot began to test the Sync for Science transmit technology on behalf of the All of Us Research Program [2]. The pilot involves 4 EHR vendors and approximately a dozen providers. Participating providers are recruiting up to 100 of their patients first to enroll into All of Us and then to consent to use a consumer-facing research app built by the Sync for Science team. The study participant uses the app to identify his or her provider, and then the app asks the provider's EHR system to transfer the participant's EHR data to the study. The provider's EHR system, which the piloting vendor has modified for Sync for Science, uses a 2-step authentication process, first to validate that the request is coming from a registered app and then to validate that a legitimate patient issued the request. If the authentication process succeeds, the EHR system issues an *access token* to the app that enables it to transfer the data.

At no point does the study participant give the research app his or her portal access credentials nor does the app ever see or manage the participant's private health information. The app merely facilitates, on behalf of the patient/consumer/participant, the exchange of data between the HIPAA-covered entity and research team. Thus, the approach eliminates the primary sources of trust concerns that the researcher's organization or IRB may have.

Scaling Sync for Science

The pilot described above is intended to test the technology with a limited number of EHR vendors and providers. It will only enroll patients who receive care from one of the pilot's providers. If that participant happens to receive care from other providers, not involved in the pilot, the EHR data from the other providers will not be available. It is a technical proof of concept, rather than a test of how the *Transmit* approach could work for any study, recruiting any participant and receiving care from any provider(s) in the United States.

For this process to work at scale for any study, all US providers must be served by health information technology that complies with the Sync for Science technical framework. The research team must use a consumer-facing app that has the same technical capabilities as the app built by the Sync for Science team, and the apps' developers must have registered their products with all Sync for Science-compliant health information technology vendors in the United States. The vendors must extend the product's registration to all the providers that the technology vendor supports [115]. These systemic requirements may seem onerous. The next section describes how the ONC's 2019 proposed rule could remove apparent barriers.

US Federal Regulations Intended to Stimulate and Support Scale

The Sync for Science technical framework is none other than the FHIR API coupled with OAuth 2.0 authentication. If the ONC's February 2019 proposed rule is finalized, all vendors of provider-facing health information technology that meet ONC certification requirements will support this framework. CMS will likely require all providers participating in its value-based purchasing programs to use health information technology that satisfies the ONC's now-proposed certification requirements, and this is the mechanism through which all US providers would acquire the necessary technology. The ONC's 2019 proposed rule, if finalized, also would remove known technical and business barriers to widespread use of the technical framework. The ONC proposes to require vendors of provider-facing technology to (1) respond within 5 business days to consumer-facing apps' registration requests, (2) publish technical documentation that enables consumer-facing apps to build connections to the vendors' APIs, and (3) refrain from charging fees to apps or providers other than the fees designed to differentiate themselves in the market with value-added services. With these and other proposed certification criteria, the ONC intends to break down barriers that now inhibit this technically sophisticated approach to consumer-mediated data exchange.

When Will the Future Arrive?

The ONC proposes to require vendors to meet the updated certification requirements no later than 24 months after the rule is finalized. If finalization occurs in mid-2019 with these proposed requirements intact, by the middle of 2021, the provider-facing technology should be available to support the *Transmit* approach to consumer-mediated data exchange, for which the Sync for Science is the most visible example.

With regard to the existence of consumer-facing apps that support the *Transmit* approach for research, the prospects are also promising. Apple and other developers of consumer-facing products already are using the FHIR API standards, currently transmitting EHR data from the provider to the user's device. It should be feasible to use the same technology to transmit EHR data from the provider to a research database: Only the target has changed.

About That Directory...

Previously, we discussed directories containing providers' digital contact information in the context of the portal-supported *Transmit* approach to consumer-mediated data exchange for research. This approach would require the researchers or studies to list their digital endpoint information in the NPPES, so that providers could transmit data to researchers the same way that they would transmit data to other providers, to support patient care.

Although it is administratively and technically possible for researchers to list themselves in the NPPES, it may not be necessary if the study participant uses a research app with a FHIR API with OAuth 2.0 as its core technology. With this technology, the research app already will tell the provider's technology where to send the requested data, and so neither the provider nor user will have to look up an address in a directory.

However, while the FHIR API technology may eliminate the need for researchers or their studies to be listed in the NPPES, a directory of providers will still be necessary because the study participants will need to tell the research app who their providers are. In the Sync for Science pilot now underway, a small directory exists that only includes the providers participating in the pilot. In the future, at scale, there will need to be a directory that lists every US provider with the technology to respond to a request to transfer data using an FHIR API.

The NPPES may not be a good vehicle for meeting these research needs. Although it is publicly accessible, it was designed to facilitate exchanges of information between HIPAA-covered entities; it was not designed to meet consumer needs. It contains records for health plans and for individual clinicians and facilities. For each entry, there could be several types of electronic addresses used [47,51]. The data transfer technology will most likely exist at an organizational level, embracing hundreds or even thousands of clinicians and perhaps scores of facilities. It is not clear how a study participant would know how to select the appropriate *provider* for the purposes of transferring data.

However, there is an alternative, in which the research app that study participants will use contain its own provider directory, with a user interface designed to help participants search for their providers and select them when they find them. The apps' developers could use the content of the NPPES (which is publicly available via download) to populate the consumer-friendly research app directory.

In summary, the FHIR API technology eliminates the need for a directory containing digital contact information for the research database, because the target that receives the data will be identified within the research app that the study participants

will use. The FHIR API technology does not eliminate the need for a directory containing a list of US providers, but the task of designing and populating a directory that consumers can use will probably be assumed by the developers of the research apps.

Conclusions

There are 2 approaches to consumer-mediated data exchange for research. In one approach, which we call *Download and Send*, study participants use a consumer-facing app to download and aggregate their EHR data from all their providers and then send the aggregated record to the research database. In the other approach, which we call *Transmit*, study participants use a consumer-facing app to direct their providers to transmit their data to a research database, where the researcher then aggregates records coming from multiple providers.

As of early 2019, technical support for the *Download and Send* approach is presumed to be strong and usability has been demonstrated. If researchers wish to proceed now using the *Download and Send* approach supported by portal technology, they may encounter trust concerns among their organizations or IRBs, focused on the role of the unregulated consumer-facing app. This paper offers ways in which researchers can respond to concerns that their own organizations may raise, but if the organization is very risk-averse, even these responses may not satisfy them. If this occurs, the researcher may need to relax the study's requirements for complete data from all possible participants.

Alternatively, the researcher could wait until the environment becomes more favorable. Activities are underway now that could produce a much more favorable environment within several years. These include CARIN Alliance's efforts to establish a universally accepted trust framework, in which the FTC could investigate actors who publicly commit to adhering to it and then violate that pledge. In addition, the ONC has proposed rules which, if finalized, would stimulate universal adoption among providers of next-wave technologies that support the *Transmit* approach, which mitigates many of the current trust concerns.

For the researcher interested in using consumer-mediated data exchange, there are potential limitations other than those that we already addressed:

- The paper assumes that ONC estimates about the widespread prevalence of the necessary provider-facing technology are correct. As the ONC estimates are based on provider self-reporting and vendor attestation, rather than actual patient experience, this assumption could be incorrect. In addition, the ONC estimates rely on the presence of the necessary technology, not whether the provider is actually using it, or that the patients can use it as well [116].
- The paper assumes that study participants are willing and able to use the necessary consumer-facing technology and that they have credentials that allow them to access their own data from their providers. Although there is evidence that consumers are increasingly taking advantage of patient portals [117], these assumptions may be incorrect.
- The paper assumes that data obtained through consumer-mediated data exchange will contain the data elements in the Common Clinical Data Set [40] because that is what the current regulations require. It is possible that researchers need data elements not in the common set, which would limit their utility to the study. In the ONC's February 2019 proposed rule, the required data would expand to include clinical notes and provenance, and the rule also establishes a predictable process through which further expansions would occur, so these limitations would likely relax over time. What is more troubling is the possibility that the data which providers actually make available are something less than the regulations require [118].

In the absence of the limitations described above, researchers could be using the *Download and Send* approach now to obtain EHR data for their study participants, assuming they are able to manage the trust concerns that their own organizations or IRBs could raise. It is quite likely that within a few years, researchers could be using the *Transmit* approach, which should mitigate these concerns. Hopefully, this paper gives researchers ways to respond to trust concerns if they arise now and prepare themselves for a future in which the concerns are eased because of the anticipated widespread adoption of emerging technologies.

Acknowledgments

Funding was provided through internal research and development funds at RTI International. The authors acknowledge input from David Kreda representing Sync for Science, and input from Fred Trotter, at CareSet Systems, who provided invaluable insight into the NPPES.

Authors' Contributions

All authors contributed to this manuscript by conducting background research, writing, and editing drafts.

Conflicts of Interest

None declared.

References

1. Holve E, Segal C, Hamilton Lopez M. Opportunities and challenges for comparative effectiveness research (CER) with Electronic Clinical Data: a perspective from the EDM forum. Med Care 2012 Jul;50 Suppl:S11-S18. [doi: 10.1097/MLR.0b013e318258530f] [Medline: 22692252]
2. National Institutes of Health. 2018. All of Us Research Program URL: https://allofus.nih.gov/ [accessed 2019-04-13] [WebCite Cache ID 77b55nQQt]
3. Fleurence RL, Beal AC, Sheridan SE, Johnson LB, Selby JV. Patient-powered research networks aim to improve patient care and health research. Health Aff (Millwood) 2014 Jul;33(7):1212-1219. [doi: 10.1377/hlthaff.2014.0113] [Medline: 25006148]
4. Pitts SR, Carrier ER, Rich EC, Kellermann AL. Where Americans get acute care: increasingly, it's not at their doctor's office. Health Aff (Millwood) 2010 Sep;29(9):1620-1629. [doi: 10.1377/hlthaff.2009.1026] [Medline: 20820017]
5. Unni N, Peddinghaus M, Tormey CA, Stack G. Record fragmentation due to transfusion at multiple health care facilities: a risk factor for delayed hemolytic transfusion reactions. Transfusion 2014 Jan;54(1):98-103. [doi: 10.1111/trf.12251] [Medline: 23711236]
6. Grinspan ZM, Abramson EL, Banerjee S, Kern LM, Kaushal R, Shapiro JS. People with epilepsy who use multiple hospitals; prevalence and associated factors assessed via a health information exchange. Epilepsia 2014 May;55(5):734-745 [FREE Full text] [doi: 10.1111/epi.12552] [Medline: 24598038]
7. Bell RA, Quandt SA, Arcury TA, Snively BM, Stafford JM, Smith SL, et al. Primary and specialty medical care among ethnically diverse, older rural adults with type 2 diabetes: the ELDER Diabetes Study. J Rural Health 2005;21(3):198-205 [FREE Full text] [Medline: 16092292]
8. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 2013 Aug;51(8 Suppl 3):S30-S37 [FREE Full text] [doi: 10.1097/MLR.0b013e31829b1dbd] [Medline: 23774517]
9. 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]
10. Sung NS, Crowley WF, Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. J Am Med Assoc 2003 Mar 12;289(10):1278-1287. [Medline: 12633190]
11. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. Health Serv Res 2010 Oct;45(5 Pt 2):1468-1488 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01142.x] [Medline: 21054367]
12. Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I, et al. Data linkage: a powerful research tool with potential problems. BMC Health Serv Res 2010 Dec 22;10:346 [FREE Full text] [doi: 10.1186/1472-6963-10-346] [Medline: 21176171]
13. Williams C, Mostashari F, Mertz K, Hogin E, Atwal P. From the Office of the National Coordinator: the strategy for advancing the exchange of health information. Health Aff (Millwood) 2012 Mar;31(3):527-536 [FREE Full text] [doi: 10.1377/hlthaff.2011.1314] [Medline: 22392663]
14. Mandl KD, Kohane IS. Time for a patient-driven health information economy? N Engl J Med 2016 Jan 21;374(3):205-208 [FREE Full text] [doi: 10.1056/NEJMp1512142] [Medline: 26789868]
15. U.S. Department of Health and Human Services. 2019 Feb 25. Get it. Check it. Use it. Access to your health information is your right URL: https://www.hhs.gov/hipaa/for-individuals/right-to-access/index.html [accessed 2019-04-13] [WebCite Cache ID 77b6Qny5X]
16. DeSalvo K, Samuels J. Health IT Buzz. 2016 Jun 02. Empowering Patients: New Videos to Promote Access to Electronic Health Information URL: https://www.healthit.gov/buzz-blog/consumer/empowering-patients-videos-promote-access-electronic-health-information [accessed 2019-04-13] [WebCite Cache ID 77b6dG3Md]
17. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
18. Coelius Rebecca Mitchell. Health IT Buzz. 2014 Jan 31. Get the facts regarding view, download and transmit 2014 requirements URL: https://www.healthit.gov/buzz-blog/meaningful-use/view-download-transmit-facts [accessed 2019-04-13] [WebCite Cache ID 77b7wTGpa]
19. U.S. Government Publishing Office. 2011 Oct 01. 45 CFR 164.524 - Access of individuals to protected health information URL: https://www.govinfo.gov/app/details/CFR-2011-title45-vol1/CFR-2011-title45-vol1-sec164-524 [accessed 2019-04-13] [WebCite Cache ID 77b7BrS7Q]
20. U.S. Government Information. 2016 Dec 13. Public Law 114 - 255 - 21st Century Cures Act URL: https://www.govinfo.gov/app/details/PLAW-114publ255 [accessed 2019-04-13] [WebCite Cache ID 77b7kKmJx]
21. CMS. Centers for Medicare & Medicaid Services. 2018 Mar 06. Trump Administration Announces MyHealthEData Initiative to Put Patients at the Center of the US Healthcare System URL: https://www.cms.gov/newsroom/press-releases/

trump-administration-announces-myhealthedata-initiative-put-patients-center-us-healthcare-system [accessed 2019-04-13] [WebCite Cache ID 77b8AiD3V]
22. Centers for Medicare & Medicaid Services. 2018 Mar 16. Remarks by Administrator Seema Verma at the ONC Interoperability Forum in Washington, DC URL: https://www.cms.gov/newsroom/press-releases/speech-remarks-administrator-seema-verma-onc-interoperability-forum-washington-dc [accessed 2019-04-13] [WebCite Cache ID 77b8Os0BM]
23. Centers for Medicare & Medicaid Services. 2015 Oct 16. Medicare and Medicaid Programs. Electronic Health Record Incentive Program—Stage 3 and Modifications to Meaningful Use in 2015 Through 2017. 42 CFR Parts 412 and 495 URL: https://www.govinfo.gov/content/pkg/FR-2015-10-16/pdf/2015-25595.pdf [accessed 2019-04-13] [WebCite Cache ID 77b8jhOof]
24. Centers for Medicare & Medicaid Services. 2018 Aug 17. Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2019 Rates; Quality Reporting Requirements for Specific Providers; Medicare and Medicaid Electronic Health Record (EHR) Incentive Programs (Promoting Interoperability Programs) Requirements for Eligible Hospitals, Critical Access Hospitals, and Eligible Professionals; Medicare Cost Reporting Requirements; and Physician Certification and Recertification of Claims: A Rule by the Centers for Medicare and Medicaid Services on 8/17/2018 URL: https://www.federalregister.gov/documents/2018/08/17/2018-16766/medicare-program-hospital-inpatient-prospective-payment-systems-for-acute-care-hospitals-and-the [accessed 2019-04-13] [WebCite Cache ID 77b96VnZx]
25. Center for Medicare and Medicaid Services. 2018 Nov 23. Final Rule. Medicare Program; Revisions to Payment Policies under the Physician Fee Schedule and Other Revisions to Part B for CY 2019; Medicare Shared Savings Program Requirements; Quality Payment Program; and Medicaid Promoting Interoperability Program URL: https://www.federalregister.gov/documents/2018/11/23/2018-24170/medicare-program-revisions-to-payment-policies-under-the-physician-fee-schedule-and-other-revisions [accessed 2019-04-12]
26. HHS Press Office. U.S. Department of Health & Human Services. 2019 Feb 11. HHS Proposes New Rules to Improve the Interoperability of Electronic Health Information URL: https://www.hhs.gov/about/news/2019/02/11/hhs-proposes-new-rules-improve-interoperability-electronic-health-information.html [accessed 2019-04-13] [WebCite Cache ID 77b9ZfMbV]
27. Blumenthal D. Harvard Business Review. 2019 Mar 15. A Big Step Toward Giving Patients Control Over Their Health Care Data URL: https://hbr.org/2019/03/a-big-step-toward-giving-patients-control-over-their-health-care-data [accessed 2019-04-13] [WebCite Cache ID 77b9fnJvt]
28. Centers for Medicare & Medicaid Services. 2018 Aug 06. Remarks by Administrator Seema Verma at the ONC Interoperability Forum in Washington, DC URL: https://www.cms.gov/newsroom/press-releases/speech-remarks-administrator-seema-verma-onc-interoperability-forum-washington-dc [accessed 2019-04-13] [WebCite Cache ID 77b9yz3ED]
29. Monica K. EHR Intelligence. 2018 Mar 07. CMS Shares Plans to Enable Interoperability, Value-Based Care URL: https://ehrintelligence.com/news/cms-shares-plans-to-enable-interoperability-value-based-care [accessed 2019-04-13] [WebCite Cache ID 77bA4W4se]
30. Weiskopf NG, Rusanov A, Weng C. Sick patients have more data: the non-random completeness of electronic health records. AMIA Annu Symp Proc 2013;2013:1472-1477 [FREE Full text] [Medline: 24551421]
31. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 1;20(1):144-151 [FREE Full text] [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
32. Stevenson F. The use of electronic patient records for medical research: conflicts and contradictions. BMC Health Serv Res 2015 Mar 29;15:124 [FREE Full text] [doi: 10.1186/s12913-015-0783-6] [Medline: 25886266]
33. Liu C, Talaei-Khoei A, Zowghi D, Daniel J. Data completeness in healthcare: a literature survey. PAJAIS 2017 Jun;9(2):75-100 [FREE Full text]
34. Office of the National Coordinator for Health Information Technology. 2016 Sep. U.S. Hospital Adoption of Patient Engagement Functionalities URL: https://dashboard.healthit.gov/quickstats/pages/FIG-Hospital-Adoption-of-Patient-Engagement-Functionalities.php [accessed 2019-04-13] [WebCite Cache ID 77bB1tmX3]
35. Office of the National Coordinator for Health Information Technology. 2016 Dec. Office-based Physician Electronic Patient Engagement Capabilities URL: https://dashboard.healthit.gov/quickstats/pages/physicians-view-download-transmit-secure-messaging-patient-engagement.php [accessed 2019-04-13] [WebCite Cache ID 77bBDaNpv]
36. American Hospital Association. 2016 Jul. Individuals' ability to electronically access their hospital medical records, perform key tasks is growing URL: https://www.aha.org/system/files/research/reports/tw/16jul-tw-healthIT.pdf [accessed 2019-04-13] [WebCite Cache ID 77bBK6OVK]
37. Heisey-Grove D, Patel V, Searcy T. The Office of the National Coordinator for Health Information Technology. 2015 Sep. Physician Electronic Exchange of Patient Health Information, 2014 URL: https://dashboard.healthit.gov/evaluations/

data-briefs/physician-electronic-exchange-patient-health-information.php [accessed 2019-04-22] [WebCite Cache ID 77ovInAMJ]
38. Office of the National Coordinator for Health Information Technology. 2018 Mar. 2015 Edition Market Readiness for Hospitals and Clinicians. Health IT Quick-Stat #55 URL: https://dashboard.healthit.gov/quickstats/pages/2015-edition-market-readiness-hospitals-clinicians.php [accessed 2019-04-13] [WebCite Cache ID 77bBcVhwH]
39. Office of the National Coordinator for Health Information Technology. 2015 Oct 16. Final Rule. 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications URL: https://www.federalregister.gov/documents/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base [accessed 2019-04-13] [WebCite Cache ID 77bBlzLAB]
40. Office of the National Coordinator for Health Information Technology. 2015 Oct 16. Table 8. Common Clinical Data Set. in Final Rule: 2015 Edition Health Information Technology (Health IT) Certification Criteria URL: https://www.federalregister.gov/documents/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base [accessed 2019-04-13] [WebCite Cache ID 77bCOfL42]
41. Office of the National Coordinator for Health Information Technology. 2018 Feb 09. 2015 Edition Common Clinical Data Set - 45 CFR 170.102 URL: https://www.healthit.gov/sites/default/files/topiclanding/2018-04/2015Ed_CCG_CCDS.pdf [accessed 2019-04-13] [WebCite Cache ID 77bCiJmbc]
42. United States Government Accountability Office. 2017 Mar. HHS Should Assess the Effectiveness of Its Efforts to Enhance Patient Access to and Use of Electronic Health Information URL: https://www.gao.gov/assets/690/683388.pdf [accessed 2019-04-13] [WebCite Cache ID 77bCwtNj8]
43. Charles D, Gabriel M, Henry J. Office of the National Coordinator for Health Information Technology. 2015 Oct. Electronic Capabilities for Patient Engagement among U.S. Non-Federal Acute Care Hospitals: 2012-2014 URL: https://www.healthit.gov/sites/default/files/briefs/oncdatabrief29_patientengagement.pdf [accessed 2019-04-13] [WebCite Cache ID 77bDAFHzU]
44. Pylypchuk Y, Johnson C, Henry J, Ciricean D. Office of the National Coordinator for Health Information Technology. 2018 Nov. Variation in Interoperability among U.S. Non-federal Acute Care Hospitals in 2017 URL: https://www.healthit.gov/sites/default/files/page/2018-11/Interop%20variation_0.pdf [accessed 2019-04-13] [WebCite Cache ID 77bDIpL8l]
45. Office of the National Coordinator for Health Information Technology. 2017 Jun 01. Improving the Health Records Request Process for Patients: Insights from User Experience Research URL: https://www.healthit.gov/sites/default/files/onc_records-request-research-report_2017-06-01.pdf [accessed 2019-04-13] [WebCite Cache ID 77bDVHpJG]
46. DirectTrust. 2017 May 05. DirectTrust Overview URL: https://www.directtrust.org/wp-content/uploads/2017/05/DirectTrust-One-Pager-05.12.2017.pdf [accessed 2019-04-13] [WebCite Cache ID 77bDyKZTy]
47. Centers for Medicare & Medicaid Services. 2019 Mar 04. Proposed Rule. Interoperability and Patient Access for MA Organization and Medicaid Managed Care Plans, State Medicaid Agencies, CHIP Agencies and CHIP Managed Care Entities, Issuers of QHPs in the Federally-facilitated Exchanges and Health Care Providers (CMS-9115-P) URL: https://www.federalregister.gov/documents/2019/03/04/2019-02200/medicare-and-medicaid-programs-patient-protection-and-affordable-care-act-interoperability-and [accessed 2019-04-13] [WebCite Cache ID 77bE8Czqd]
48. National Plan & Provider Enumeration System. 2016. Other Identifiers Page URL: https://nppes.cms.hhs.gov/webhelp/nppeshelp/OTHER%20IDENTIFIERS%20PAGE.html [accessed 2019-04-13] [WebCite Cache ID 77bEHMCGQ]
49. Centers for Medicare & Medicaid Services. 2018 May 23. Announcing Changes to the National Plan and Provider Enumeration System (NPPES) Downloadable File URL: https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand/DataDissemination.html [accessed 2019-04-13] [WebCite Cache ID 77bElGMYy]
50. FHIR. 2018 Dec 28. 8.9 Resource Endpoint - Content URL: https://www.hl7.org/fhir/endpoint.html [accessed 2019-04-13] [WebCite Cache ID 77bEx3KKl]
51. Trotter F. CareSet Systems. 2018 Jul 18. New NPPES features will improve interoperability efforts URL: https://careset.com/new-nppes-features-will-improve-interoperability-efforts/ [accessed 2019-04-13] [WebCite Cache ID 77bF4ZPFZ]
52. Office of the National Coordinator for Health Information Technology. 2019 Jan 30. Patient Engagement Playbook URL: https://www.healthit.gov/playbook/pe/ [accessed 2019-04-13] [WebCite Cache ID 77bFEp7by]
53. FHIR. 2018 Dec 28. 2.13 FHIR Overview URL: https://www.hl7.org/fhir/overview.html [accessed 2019-04-13] [WebCite Cache ID 77bFXqTeM]
54. Rouse M, Sutner S, Vallejo M. HealthIT. 2017 Oct. Definition: FHIR (Fast Healthare Interoperability Resources) URL: https://searchhealthit.techtarget.com/definition/FHIR-Fast-Healthcare-Interoperability-Resources [accessed 2019-04-13] [WebCite Cache ID 77bFfv37R]
55. Bresnick J. Health IT Analytics.: xtelligent HEALTHCARE media; 2016 Mar 22. 4 Basics to Know about the Role of FHIR in Interoperability URL: https://healthitanalytics.com/news/4-basics-to-know-about-the-role-of-fhir-in-interoperability [accessed 2019-04-13] [WebCite Cache ID 77bFqxsXY]

56. Olschesky M. Datica. 2017 Mar 15. FHIR is No Longer Just a Concept URL: https://datica.com/blog/fhir-is-no-longer-just-a-concept/ [accessed 2019-04-13] [WebCite Cache ID 77bFzZ47y]
57. OpenAPI Initiative. 2019. URL: https://www.openapis.org/about [accessed 2019-04-15] [WebCite Cache ID 77e1JIg3G]
58. Office of the National Coordinator for Health Information Technology. 2019 Mar 04. Proposed Rule. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program URL: https://www.govinfo.gov/content/pkg/FR-2019-03-04/pdf/2019-02224.pdf [accessed 2019-04-14] [WebCite Cache ID 77dZW7snZ]
59. Office of the National Coordinator for Health Information Technology. 2019 Apr 05. Notice of Proposed Rulemaking to Improve the Interoperability of Health Information URL: https://www.healthit.gov/topic/laws-regulation-and-policy/notice-proposed-rulemaking-improve-interoperability-health, [accessed 2019-04-14] [WebCite Cache ID 77dZj1Gb4]
60. Posnack S, Barker W. HealthIT Buzz. 2018 Oct 01. The U.S. is Poised to Catch FHIR in 2019 URL: https://www.healthit.gov/buzz-blog/interoperability/heat-wave-the-u-s-is-poised-to-catch-fhir-in-2019/: [accessed 2019-04-14] [WebCite Cache ID 77dZo7dhc]
61. Posnack S. Office of the National Coordinator for Health Information Technology. 2019 Feb. API Conditions of Certification (and more!). Two Statutory Sections Implemented Together URL: https://www.healthit.gov/sites/default/files/page/2019-02/PosnackAPICoCFeesIB.pdf [accessed 2019-04-14] [WebCite Cache ID 77dZySJI0]
62. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
63. Roehrs A, da Costa CA, Righi RD, de Oliveira KS. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
64. Vance B, Tomblin B, Studeny J, Coustasse A. Benefits and barriers for adoption of personal health records. 2015 Mar Presented at: 2015 Business and Health Administration Association Annual Conference, at the 51st Annual Midwest Business Administration Association International Conference; March 2015; Chicago, IL URL: https://mds.marshall.edu/mgmt_faculty/135/
65. Pushpangadan S, Seckman C. Consumer perspective on personal health records: a review of the literature. OJNI 2015 Feb 04;19(1) [FREE Full text]
66. Carebox. 2019. Automated Medical Record Collection URL: http://careboxhealth.com/ [accessed 2019-04-14] [WebCite Cache ID 77daecTg6]
67. Leukemia & Lymphoma Society. 2019. The LLS National Patient Registry Project URL: https://www.lls.org/article/the-lls-national-patient-registry-project
68. HITC Staff. HIT Consultant. 2016 May 10. Yale School of Medicine Launches Cloud-Based PHR Platform Hugo URL: https://hitconsultant.net/2016/05/10/yale-school-medicine-launches-cloud-based-phr-platform-hugo/ [accessed 2019-04-14] [WebCite Cache ID 77dauf4u5]
69. Manos D. Healthcare Information and Management Systems Society. 2017 Jan 24. Yale researchers tapping into emerging secure cloud platform for sharing patient data URL: https://www.healthcareitnews.com/news/yale-researchers-building-secure-cloud-platform-sharing-patient-data, [accessed 2019-04-14] [WebCite Cache ID 77db2dpOB]
70. Raths D. Healthcare Innovation. 2018 Aug 13. Patient-Centered PHR Pulls Together Four Data Sources for Post-Market Surveillance of Medical Devices URL: https://www.healthcare-informatics.com/article/mobile/patient-centered-phr-pulls-together-four-data-sources-post-market-surveillance, [accessed 2019-04-14] [WebCite Cache ID 77db7pfdK]
71. Bresnick J. HealthIT Analytics. 2016 Mar 01. ONC Launches Two Interoperability App Challenges URL: https://healthitanalytics.com/news/onc-launches-two-fhir-interoperability-app-challenges [accessed 2019-04-14] [WebCite Cache ID 77dbFVmwV]
72. U.S. Department of Health & Human Services. 2017 Jan 12. ONC announces winners of consumer and provider app challenges to improve health information access and use URL: http://wayback.archive-it.org/3926/20170127095313/https://www.hhs.gov/about/news/2017/01/12/onc-announces-winners-consumer-and-provider-app-challenges-improve-health-information-access-and-use [accessed 2019-04-14] [WebCite Cache ID 77dbJQk70]
73. Willis D. MyLinks. 2018. Patients having their aggregate health data in their own hands will change healthcare! URL: https://www.mylinks.com/patients-having-their-aggregate-health-data-in-their-own-hands-will-change-healthcare/ [accessed 2019-04-14] [WebCite Cache ID 77dbbt9hC]
74. Apple. 2018 Jan 24. Apple announces effortless solution bringing health records to iPhone URL: https://www.apple.com/newsroom/2018/01/apple-announces-effortless-solution-bringing-health-records-to-iPhone/ [accessed 2019-04-14] [WebCite Cache ID 77dbjDBWj]
75. Blumenthal D, Chopra A. Harvard Business Review. 2018 Mar 23. Apple's Pact with 13 Healthcare Systems Might Actually Disrupt the Industry URL: https://hbr.org/2018/03/apples-pact-with-13-health-care-systems-might-actually-disrupt-the-industry [accessed 2019-04-14] [WebCite Cache ID 77dbnBpxH]

76. Comstock J. Healthcare IT News. 2018 Jan 24. Apple to launch Health Records app with HL7's FHIR specifications at 12 hospitals URL: https://www.healthcareitnews.com/news/apple-launch-health-records-app-hl7s-fhir-specifications-12-hospitals [accessed 2019-04-14] [WebCite Cache ID 77dbsjrG6]
77. Ravindranath M. Politico. 2018 Mar 23. Apple's EHR play could expand APIs Politico March 23, 2018 URL: https://www.politico.com/story/2018/03/23/apples-ehr-play-could-expand-apis-482187 [accessed 2019-04-14] [WebCite Cache ID 77dc3MSOK]
78. Miliard M. HealthcareIT News. 2018 Apr 18. How Apple's Health Records could reshape patient engagement URL: https://www.healthcareitnews.com/news/how-apples-health-records-could-reshape-patient-engagement [accessed 2019-04-14] [WebCite Cache ID 77dc9CjCU]
79. Apple Inc. 2019. Empower your patients with Health Records on iPhone URL: https://www.apple.com/healthcare/health-records/ [accessed 2019-04-14] [WebCite Cache ID 77dcDwZ99]
80. Baier A. Trust and antitrust. Ethics 1986 Jan;96(2):231-260. [doi: 10.1086/292745]
81. Blaze M, Ioannidis J, Keromytis A. Columbia University. 2003. Experience with the KeyNote Trust Management System: ApplicationsFuture Directions URL: https://www1.cs.columbia.edu/~angelos/Papers/newkeynote.pdf [accessed 2019-04-22] [WebCite Cache ID 77owfFGOS]
82. Adjekum A, Ienca M, Vayena E. What is trust? Ethics and risk governance in precision medicine and predictive analytics. OMICS 2017 Dec 01;21(12):704-710 [FREE Full text] [doi: 10.1089/omi.2017.0156]
83. De Mooy M. Center for Democracy and Technology. 2016 Apr 15. Re: Request for Information on Updates to ONC's Voluntary Personal Health Record Model Privacy Notice. Center for Democracy and Technology URL: https://cdt.org/ [accessed 2019-04-14] [WebCite Cache ID 77dczOLb2]
84. Flaumenhaft Y, Ben-Assuli O. Personal health records, global policy and regulation review. Health Policy 2018 Aug;122(8):815-826. [doi: 10.1016/j.healthpol.2018.05.002]
85. Lecker R, Armijo D, Chin S, Christensen J, Desper J, Hong A, et al. Altarum. 2007 Jan 05. Review of the Personal Health Record (PHR) Service Provider Market: Privacy and Security URL: https://www.cdt.org/files/healthprivacy/PHRs_Altarum_2007.pdf [accessed 2019-04-14] [WebCite Cache ID 77ddLXHin]
86. Office of the National Coordinator for Health Information Technology. 2012 Dec 13. Non-HIPAA Covered Entities: Privacy and Security Policies and Practices of PHR Vendors and Related Entities Report URL: https://www.healthit.gov/sites/default/files/maximus_report_012816.pdf: [accessed 2019-04-14] [WebCite Cache ID 77ddVE41r]
87. Carrión SI, Fernández-Alemán JL, Toval A. Are personal health records safe? A review of free web-accessible personal health record privacy policies. J Med Internet Res 2012;14(4):e114 [FREE Full text] [doi: 10.2196/jmir.1904]
88. Connecting for Health Work Group on Consumer Access Policies for Networked Personal Health Information. Markle. 2008 Jun. Overview and Principles URL: https://www.markle.org/sites/default/files/CF-Consumers-Full.pdf [accessed 2019-04-14] [WebCite Cache ID 77dsnGlrm]
89. Office of the National Coordinator for Health Information Technology. 2018 Feb 06. Model Privacy Notice (MPN) URL: https://www.healthit.gov/topic/privacy-security-and-hipaa/model-privacy-notice-mpn [accessed 2019-04-15] [WebCite Cache ID 77f5qjkgB]
90. HealthIT. 2018. 2018 Model Privacy Notice URL: https://www.healthit.gov/sites/default/files/2018modelprivacynotice.pdf [accessed 2019-04-14] [WebCite Cache ID 77dt0gUbd]
91. Center for Democracy and Technology. 2009 Jun 09. Personal Health Records Need a Comprehensive and Consistent Privacy and Security Framework URL: https://cdt.org/insight/personal-health-records-need-a-comprehensive-and-consistent-privacy-and-security-framework [accessed 2019-04-14] [WebCite Cache ID 77dt7hJKy]
92. Gellman R. World Privacy Forum: Legal and Policy Analysis. 2008 Feb 20. Personal Health Records: Why Many PHRs Threaten Privacy URL: http://www.worldprivacyforum.org/wp-content/uploads/2012/04/WPF_PHR_02_20_2008fs.pdf [accessed 2019-04-14] [WebCite Cache ID 77dtCuI6w]
93. Glenn T, Monteith S. Privacy in the digital world: medical and health data outside of HIPAA protections. Curr Psychiatry Rep 2014 Nov;16(11):494. [doi: 10.1007/s11920-014-0494-4] [Medline: 25218603]
94. U.S. Department of Health and Human Services. U.S. Department of Health and Human Services. 2016 Jul 19. Examining Oversight of the Privacy & Security of Health Data Collected by Entities not Regulated by HIPAA URL: https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf [accessed 2019-04-14] [WebCite Cache ID 77dtTrwtz]
95. FTC Staff Report. Federal Trade Commission. 2015 Jan. Internet of Things: Privacy and Security in a Connected World URL: https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf [accessed 2019-04-14] [WebCite Cache ID 77dtpuatY]
96. Wang CJ, Huang DJ. The HIPAA conundrum in the era of mobile health and communications. J Am Med Assoc 2013 Sep 18;310(11):1121-1122. [doi: 10.1001/jama.2013.219869]
97. Center for Democracy and Technology. 2009 Jun 12. Personal Health Records: Who Are You Going to Trust? URL: https://cdt.org/blog/personal-health-records-who-are-you-going-to-trust/ [accessed 2019-04-14] [WebCite Cache ID 77duCemZU]

98. McGraw D. Center for Democracy and Technology. 2008 Sep. Why the HIPAA Privacy Rules Would Not Adequately Protect Personal Health Records URL: https://www.cdt.org/files/healthprivacy/HIPAA-PHRs.pdf [accessed 2019-04-14] [WebCite Cache ID 77duH8xm1]
99. McGraw D. e-CareManagement Blog. 2009 Apr 21. Privacy Law Showdown? Legal and Policy Analysis URL: http://e-caremanagement.com/privacy-law-showdown-legal-and-policy-analysis [accessed 2019-04-14] [WebCite Cache ID 77duVIeOb]
100. European Commission. 2018 Jan 24. Stronger protection, new opportunities - Commission guidance on the direct application of the General Data Protection Regulation as of 25 May 2018 URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0043&from=EN
101. European Commission. 2018 Jan 24. 2018 Reform of EU Data Protection Rules URL: https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en [accessed 2019-04-14] [WebCite Cache ID 77ducrbEw]
102. The CARIN Alliance. 2019. Enabling consumers and their authorized caregivers to access more of their digital health information with less friction URL: https://www.carinalliance.com/about-us/ [accessed 2019-04-14] [WebCite Cache ID 77dul0X3B]
103. The CARIN Alliance. 2018 Nov 26. Trust Framework and Code of Conduct URL: https://www.carinalliance.com/our-work/trust-framework-and-code-of-conduct/ [accessed 2019-04-14] [WebCite Cache ID 77dv3Jfek]
104. Raths D. Healthcare Innovation. 2018 Nov 28. CARIN Alliance Creates Code of Conduct for Third-Party Apps URL: https://www.hcinnovationgroup.com/cybersecurity/news/13030922/carin-alliance-creates-code-of-conduct-for-thirdparty-apps [accessed 2019-04-14] [WebCite Cache ID 77dv79zAl]
105. Medical Home Network. 2018 Nov 27. Voluntary code of conduct developed by more than 60 industry stakeholders can help facilitate health data exchange with entities not covered by HIPAA URL: https://medicalhomenetwork.org/news/voluntary-code-of-conduct-developed-by-more-than-60-industry-stakeholders-can-help-facilitate-health-data-exchange-with-entities-not-covered-by-hipaa [accessed 2019-04-14] [WebCite Cache ID 77dvFFObr]
106. Office of National Coordinator for Health Information Technology. 2018 Jan 05. Trusted Exchange Framework: Draft for Public Comment URL: https://www.healthit.gov/sites/default/files/draft-trusted-exchange-framework.pdf [accessed 2019-04-14] [WebCite Cache ID 77dvLr1gP]
107. Office of National Coordinator for Health Information Technology. 2019 Apr 19. Trusted Exchange Framework and Common Agreement (TEFCA): Draft 2 URL: https://www.healthit.gov/sites/default/files/page/2019-04/FINALTEFCAQTF41719508version.pdf [accessed 2019-05-19]
108. US Department of Health & Human Services. 2016 Feb 25. Individuals' Right under HIPAA to Access their Health Information 45 CFR § 164.524 URL: https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html [accessed 2019-04-22] [WebCite Cache ID 77oxwl2qd]
109. Furano R, Kushniruk A, Bernett J. Deriving a Set of Privacy Specific Heuristics for the Assessment of PHRs (Personal Health Records). In: Lau F, Bartle-Clar J, Bliss G, Borycki E, Courtney K, Kuo A, editors. Building Capacity for Health Informatics in the Future. Amsterdam, Netherlands: IOS Press BV. Amsterdam: IOS Press; 2017:125-130.
110. Public Responsibility in Medicine and Research. 2019. About Us URL: https://www.primr.org/about/ [accessed 2019-04-15] [WebCite Cache ID 77dwb66xK]
111. Yale School of Medicine. Yale School of Medicine. 2019. Looking Ahead. Engaging Patients Through a Personal Health Record URL: https://medicine.yale.edu/ycci/researchspectrum/informatics/hugo.aspx [accessed 2019-04-15] [WebCite Cache ID 77dwks9S9]
112. Sweeney E. FierceHealthcare. 2018 Mar 30. 27 more health systems join Apple Health Records platform URL: https://www.fiercehealthcare.com/tech/apple-health-records-dignity-health-patient-access [accessed 2019-04-15] [WebCite Cache ID 77dwrxXRD]
113. Apple. 2019 Apr 11. Institutions that support health records on iPhone (beta) URL: https://support.apple.com/en-us/HT208647 [accessed 2019-04-15] [WebCite Cache ID 77dzTsHga]
114. White J, Briggs J, Mandel J. Health IT Buzz. 2016 Mar 21. NIH and ONC Launch the Sync for Science (S4S) Pilot: Enabling Individual Health Data Access and Donation URL: https://www.healthit.gov/buzz-blog/health-innovation/nih-and-onc-launch-the-sync-for-science-pilot [accessed 2019-04-15] [WebCite Cache ID 77dzdcgA7]
115. Kreda D. UCI Institute for Future Health: Distinguished Seminars. 2018 May 15. Sync for Science (S4S): A Wholly Digital Way for Patients to Share their EHR Data with Researchers URL: https://futurehealth.ics.uci.edu/wp-content/ifh/uploads/David-Kreda-May2018.pdf [accessed 2019-04-15] [WebCite Cache ID 77dzpDA6z]
116. Lye CT, Forman HP, Gao R, Daniel JG, Hsiao AL, Mann MK, et al. Assessment of US hospital compliance with regulations for patients' requests for medical records. JAMA Network Open 2018 Oct 5;1(6):e183014-e183e14. [doi: 10.1001/jamanetworkopen.2018.3014]
117. Murphy-Abdouch K, Dolezel D, McLeod A. Perspectives in Health Information Management.: American Health Information Management Association (AHIMA); 2017. Patient Access to Personal Health Information: An Analysis of the Consumer's Perspective URL: https://media.proquest.com/media/hms/PFT/1/pmyE2?_s=oi0UdEaEyJuVUowhjIRpfOurkuc%3D [accessed 2019-04-15] [WebCite Cache ID 77e0UvHoc]

118. Murphy-Abdouch K. Patient access to personal health information: regulation vs. reality. Perspect Health Inf Manag 2015;12:1c [FREE Full text] [Medline: 26807075]

Abbreviations

API: application programming interface
CMS: Centers for Medicare and Medicaid Services
EHR: electronic health record
FHIR: Fast Health Care Interoperability Resources
FTC: Federal Trade Commission
HIPAA: Health Insurance Portability and Accountability Act
IRB: Institutional Review Board
MIPS: Merit-Based Incentive Payment System
NPPES: National Plan and Provider Enumeration System
ONC: Office of the National Coordinator for Health Information Technology

Edited by G Eysenbach; submitted 28.09.18; peer-reviewed by D Klein, K Goodman, L Buis, R Righi; comments to author 20.12.18; revised version received 01.02.19; accepted 24.03.19; published 03.06.19.

Please cite as:
Bracha Y, Bagwell J, Furberg R, Wald JS
Consumer-Mediated Data Exchange for Research: Current State of US Law, Technology, and Trust
JMIR Med Inform 2019;7(2):e12348
URL: http://medinform.jmir.org/2019/2/e12348/
doi:10.2196/12348
PMID:30946692

Original Paper

Medication Adherence Prediction Through Online Social Forums: A Case Study of Fibromyalgia

Kyle Haas[1*], MSc; Zina Ben Miled[1*], PhD; Malika Mahoui[2*], PhD

[1]Department of Electrical and Computer Engineering, School of Engineering and Technology, Indiana University Purdue University, Indianapolis, IN, United States

[2]Eli Lilly and Company, Indianapolis, IN, United States

[*]all authors contributed equally

Corresponding Author:
Zina Ben Miled, PhD
Department of Electrical and Computer Engineering
School of Engineering and Technology
Indiana University Purdue University
SL 160
723 W Michigan St
Indianapolis, IN, 46202
United States
Phone: 1 3172783317
Email: zmiled@iupui.edu

Abstract

Background: Medication nonadherence can compound into severe medical problems for patients. Identifying patients who are likely to become nonadherent may help reduce these problems. Data-driven machine learning models can predict medication adherence by using selected indicators from patients' past health records. Sources of data for these models traditionally fall under two main categories: (1) proprietary data from insurance claims, pharmacy prescriptions, or electronic medical records and (2) survey data collected from representative groups of patients. Models developed using these data sources often are limited because they are proprietary, subject to high cost, have limited scalability, or lack timely accessibility. These limitations suggest that social health forums might be an alternate source of data for adherence prediction. Indeed, these data are accessible, affordable, timely, and available at scale. However, they can be inaccurate.

Objective: This paper proposes a medication adherence machine learning model for fibromyalgia therapies that can mitigate the inaccuracy of social health forum data.

Methods: Transfer learning is a machine learning technique that allows knowledge acquired from one dataset to be transferred to another dataset. In this study, predictive adherence models for the target disease were first developed by using accurate but limited survey data. These models were then used to predict medication adherence from health social forum data. Random forest, an ensemble machine learning technique, was used to develop the predictive models. This transfer learning methodology is demonstrated in this study by examining data from the Medical Expenditure Panel Survey and the PatientsLikeMe social health forum.

Results: When the models are carefully designed, less than a 5% difference in accuracy is observed between the Medical Expenditure Panel Survey and the PatientsLikeMe medication adherence predictions for fibromyalgia treatments. This design must take into consideration the mapping between the predictors and the outcomes in the two datasets.

Conclusions: This study exemplifies the potential and limitations of transfer learning in medication adherence–predictive models based on survey data and social health forum data. The proposed approach can make timely medication adherence monitoring cost-effective and widely accessible. Additional investigation is needed to improve the robustness of the approach and extend its applicability to other therapies and other sources of data.

(JMIR Med Inform 2019;7(2):e12561) doi:10.2196/12561

KEYWORDS

medication adherence; fibromyalgia; Medical Expenditure Panel Survey; social forum; random forest; transfer learning

Introduction

Medication nonadherence is one of the most expensive medical expenditures. As of 2015, the cost of patient nonadherence in the United States reached US $290 billion [1]. The majority of the cost of nonadherence arises from prescriptions that are either never filled or medications that are not taken as prescribed [2]. Although the financial losses are staggering, the most prominent motivation for better adherence is saving patients whose conditions worsen due to poor compliance. Indeed, close to 125,000 deaths related to inadequate adherence were reported in the United States [3].

Identifying patients at risk and the reasons for medication nonadherence can help guide the development of remedial and preventive plans. For many years, researchers have stipulated that several factors can influence nonadherence including poor patient-doctor interactions and a lack of overall health understanding [3]. On one hand, the multitude of factors and their potential interdependence make profiling patients at risk of nonadherence difficult [3,4]. On the other hand, the digitization efforts in the health sector over the past decade have resulted in the availability of various data sources that can support the design of medication adherence–prediction models. Examples of such sources include reimbursement claims data from insurance companies, dispensed medication data from pharmacies, and medication prescription data from health providers. Using these proprietary data, several recent medication adherence–predictive models were developed and deployed. For instance, Express Scripts developed a model with 300 predictors [5]. These predictors include the patient's demographic as well as clinical and genomic indicators. The Express Scripts model was reported to have a prediction accuracy of over 90% with a lead time of 6-12 months. Similar proprietary models were also developed by Allazo Health and FICO [6,7].

These models, although successful, primarily rely on proprietary data accessible to the health provider. These data tend to be structured and relatively accurate, providing models with high predictive accuracy [5]. However, the proprietary nature of both the data and the predictive models hinders their widespread use by other health service providers. Due to these limitations, several research efforts started to explore the use of data from social media for large-scale analysis of trends in population health. For instance, social media data were used to build a machine learning model that can predict stress [8]. Twitter data were used to study allergen effects and monitor adverse events of pharmaceutical products across the United States [9,10]. Social media was also used as a mechanism for engaging patients in order to improve compliance [11,12] and assist nonadherent patients [13]. Recently, a medication-adherence model using Twitter was proposed by Klein et al [14]. This model identifies the medication intake from tweets that mention at least one of 55 different medications.

The abovementioned studies highlight the fact that medication-adherence models, in particular, and population health models, in general, have been progressing along parallel but completely disjointed paths. The first path draws its advantage from the accuracy and validity of the data collected in a controlled environment at the expense of limited applicability to the wider population, whereas the second path leverages widespread accessibility, but suffers from reduced data accuracy or lack of verifiable model validity. The objective of this paper is to answer the question: *Can machine learning models trained using data from a controlled environment be used to predict medication adherence for health social forum users?* If they can be used, this approach can bridge the abovementioned parallel paths and help combine the benefits of the two environments.

In machine learning, transfer learning is used to improve modeling in various domains including social media data [15,16,17]. This technique is similar to the ability of a human to transfer knowledge from one context to another, thereby reducing the learning efforts required with every new context. This approach has not been previously explored for medication adherence. The aim of this paper is to investigate the applicability of this approach to medication adherence among patients with fibromyalgia. Fibromyalgia was selected to demonstrate the proposed approach because of its high incidence rate and the fact that it is subject to strict medication regimens with severe consequences of nonadherence [18]. The proposed medication-adherence models for this disease are trained using the Medical Expenditure Panel Survey (MEPS) dataset [19]. Although it is not proprietary, the MEPS dataset is used in this paper as a proxy for datasets collected in a controlled environment. The target domain for knowledge transfer is the social health forum PatientsLikeMe [20]. This paper investigates the accuracy of MEPS-trained models when used to predict the adherence of PatientsLikeMe users in the case of fibromyalgia. The mapping between the variables in the source and target datasets and its impact on the prediction accuracy of the proposed models are also analyzed.

Methods

The Machine Learning Model

Typically, a machine learning model is an agent trained with a set of predictors to generate a target outcome. This model varies based on the dataset used for the training and validation as well as the technique used to train each model. Moreover, traditionally, each model is trained and validated using data from a single source, since the learning and application are confined to a single domain. In this paper, and because we are learning from one domain and applying this knowledge to a different domain, two datasets are needed. These datasets are derived from MEPS and PatientsLikeMe.

Data Extraction and Cross-Domain Variable Mapping

The MEPS database is provided by the Agency for Healthcare Research and Quality [19]. It is a collection of surveys from a nationally representative population of individuals. The survey participants provide responses in a series of five rounds over a 2-year interval. During each round, participants are asked to answer a survey questionnaire that focuses on their health status, medical conditions, prescribed medications, and insurance coverage. Each year, a new panel of participants is enrolled in the study, while the previous year's panel finishes the final

second year. This panel overlap provides an insight into nationwide dynamic changes. For the purpose of this study, patient records were extracted from panels 17-19, which span the period from 2012 to 2015.

The second data source is PatientsLikeMe, which is a social health forum where patients post, discuss, and review many of their current medications and conditions [20]. Some of the users of this forum make their data publicly available. The data were collected from treatment evaluations as of March 2017. These evaluations are available in a structured format that includes self-reported adherence to treatment.

Patients from MEPS and PatientsLikeMe were selected if they were receiving a treatment associated with the target disease fibromyalgia. Treatments were included in the list if they were taken by at least a single patient from PatientsLikeMe. This list includes duloxetine, gabapentin, pregabalin, tramadol, and zolpidem.

The model predictors that were extracted for each patient from both data sources are type of medication, years taking treatment, daily intake, dosage, age at the end of the study or last known age, sex of the patient, out-of-pocket expense, and region of living of the patient (ie, Northeast, Midwest, West, or South). These were the only predictors available in both MEPS and PatientsLikeMe.

As previously mentioned, typical machine learning models rely on a dataset from a single domain for training and validation. However, because we want to transfer knowledge from one domain to another, mapping is needed from the variables in the source dataset to their counterpart in the target dataset. This mapping is straightforward (ie, one to one) in the case of the first six predictors (ie, type of medication, years taking treatment, daily intake, dosage, age, and sex).

However, more elaborate mapping was needed for out-of-pocket expense and region of living. Although MEPS provides the exact amount paid for each medication, PatientsLikeMe lists only ranges for the approximate expenses each month. Therefore, out-of-pocket expense payments in MEPS were categorized using the out-of-pocket expense ranges provided in PatientsLikeMe (Table 1). Similarly, the residence of each MEPS patient is provided according to the appropriate US census region, while for PatientsLikeMe, the residence of the patient is provided at the state level. Again, since one-to-one mapping between the two datasets is needed for knowledge transfer, the value of region of living for PatientsLikeMe was mapped to the census region based on the state of residence of the patient (eg, Indiana is mapped to the Midwest).

Table 1. Categorical ranges for out-of-pocket expenses.

Actual expenses (US $)	Out-of-pocket expense category
<25	0
25-50	1
50-100	2
100	3
>200	4

The mapping for the out-of-pocket expense and region of living predictors between the two datasets is relatively simple. The challenge is mapping the outcome of the model. The target outcome for the model is medication adherence. In PatientsLikeMe, patients self-report a selected adherence value from four categories (ie, *Always taken as prescribed*, *Usually taken as prescribed*, *Sometimes taken as prescribed*, or *Never taken as prescribed*). The MEPS data do not include a direct measure of adherence; therefore, this measure had to be derived. In a previous study, Hess et al evaluated 11 different medication-adherence metrics and recommended the use of medication-refill adherence (MRA), which is defined as the total number of days of medication supply divided by the number of days of study participation multiplied by 100 [21]. For example, a patient with a total of 200 days of supply over a period of 365 days will have an MRA of 55%.

The MRA value from MEPS and the adherence classes from PatientsLikeMe have to be mapped to a common scale. This scale consists of two classes: adherent and nonadherent. For PatientsLikeMe, the *Always taken as prescribed* category is mapped to the adherent class and the remaining three categories (ie, *Usually taken as prescribed*, *Sometimes taken as prescribed*, and *Never taken as prescribed*) are mapped to the nonadherent class. In the case of MEPS, four different MRA thresholds are considered. For each threshold, if the MRA is greater than or equal to the threshold value, the outcome is mapped to the adherent class; otherwise, the outcome is mapped to the nonadherent class. The threshold is varied in order to understand the differences in the interpretation of adherence between the MEPS and the PatientsLikeMe datasets. This difference can be due to the fact that adherence is quantitative in MEPS and qualitative in PatientsLikeMe. Moreover, using MRA as an adherence measure in MEPS does not account for scenarios where patients are proactive in refilling their prescriptions or accidentally misplace medications. Finally, adherence in PatientsLikeMe is self-reported and may therefore be subjective [22]. Understanding the differences between the variables in the two datasets and calibrating the associated mapping is a necessary enabler for transfer learning.

Model Training and Validation

The model proposed for prediction of medication adherence is based on the random forest (RF) tool [23]. Other machine learning techniques (eg, neural networks and support vector machine) are available [24,25]. Although models based on these techniques can be considered for medication adherence,

RF was selected for this study because (1) it can handle variables with missing and categorical values, a characteristic inherent to social forum data [23]; (2) it facilitates the comparative analysis of two models trained by using different datasets including the evaluation of the importance of each predictor in each model [26], which is needed for the validation of transfer learning; and (3) it was successfully used in previous health-related models including models to predict the response of patients to various drugs and models to predict patients with liver disease [27,28]. Previous studies [27,28] showed that RF outperformed other machine learning techniques including neural networks and support vector machine.

RF consists of an ensemble of decision trees, where each tree contributes a vote to the overall decision of the RF. A majority vote of adherent or nonadherent classifies the patient as adherent or nonadherent, respectively. The uniqueness of each tree is ensured through a two-step randomization process. The first step is called bagging or bootstrap aggregation and is responsible for randomizing the patients [29]. For each tree in the RF, a predefined number of patients are selected randomly from the training dataset with replacement. Based on this selection, a given patient may be selected more than once in a given tree, while other patients may not be selected. The second randomization step selects predictors and occurs during the construction of each tree. Only a random subset of the predictors is considered at each node. In this paper, the size of this subset was set to a typical standard of Standalone Equation 1, where n is the total number of predictors in the dataset. Randomized selection of both patients and predictors from the training dataset helps generate multiple unique decision trees in the RF ensemble.

The best predictor among the Standalone Equation 1 predictors considered at each node of the decision tree is selected based on the greatest reduction in impurity [30]. The parent node in the tree always has a higher impurity (less homogenous set of patients) than its children. A homogenous set of patients corresponds to the case where all the patients belong to the same class (ie, adherent class or nonadherent class). The impurity of a node is defined by $I=1-(A_+)^2-(A_-)^2$, where A_+ and A_- are the percentage of adherent and nonadherent patients, respectively, presented to the node in a given tree [30]. The change in impurity between the parent node (p) and its left (l) and right (r) child nodes is given by $\Delta I= I_p- P_lI_l- P_rI_r$, where P_l and P_r represent the percent of the total number of patients in the parent node that are mapped to the left and right branches, respectively.

In order to split the patients into the appropriate right or left branches at each node, the selected predictor requires a reference value. All possible values for a given predictor are iteratively evaluated until the best split is found (ie, branches with the lowest impurity). In general, predictors can be numeric or categorical. For instance, the predictor age is numeric. When age is used as a predictor for a given node in the tree, patients with an age value greater than the reference value are assigned to the right branch of the node, and the remaining patients are assigned to the left branch. For the categorical predictors, such as region of living, patients that have the same value as the reference value are assigned to the right branch, while the remaining patients are assigned to the left branch.

The abovementioned procedure describes the traditional learning process used for decision trees [23]. One of the limitations of this process is that it does not adequately handle patients with missing predictor values. As previously mentioned, missing predictor values are prevalent in social media data. Specifically, in the PatientsLikeMe dataset, 40% of the patients with fibromyalgia had at least one missing predictor value. Several approaches can be considered to handle missing predictor values: (1) A default split can be adopted, where the patient with a missing predictor value is arbitrarily yet consistently assigned to a given branch. (2) The corresponding patient record can be removed. (3) All numeric predictors are binned using k-means clustering [27]. An additional bin is then added to represent the case where the predictor value is missing.

All of the abovementioned approaches were investigated and discarded because of their limitations. The default split is arbitrary and leads to poor accuracy. The second approach excludes approximately half of the patients from PatientsLikeMe. The third approach translates all numerical values to categorical values. In order to overcome the limitations of these previous approaches, a new technique for learning with missing values in RF models is proposed.

Traditional RF models use decision trees that are binary trees, where each node has one left and one right child. The proposed model uses ternary decisions trees, where each node has three children: a left child, a middle child, and a right child. Patients with missing predictor values are assigned to the third child. The underlying learning algorithm is modified in order to accommodate the additional child and to ensure that the missing value is never selected as a reference in the split at any node.

RF models for fibromyalgia are trained using the abovementioned approach. The training dataset is composed of patient records that include the predictors and the target outcome (ie, adherent/nonadherent). It is also balanced and consists of an equal number of adherent/nonadherent patients. This class balance eliminates the potential of bias in the model toward the larger represented class. The model is then validated using a testing dataset that is completely independent from the training dataset. Two metrics are used to quantify this validation: (1) Accuracy: The ratio of the number of records that are correctly predicted to the total number of records in the testing dataset and (2) F_1 score: A composite metric that represents a weighted balance between the recall and the precision of the models. Recall accounts for the number of correctly classified adherent patients compared with the total number of actually adherent patients in the testing dataset. Precision is the total number of correctly classified adherent patients against the total number of patients classified as adherent.

Another measure is used to evaluate the importance of each predictor in the model. Predictor importance (PI) is defined as the ratio of the number of times a predictor is traversed to the total number of times all the predictors are traversed in the RF model when processing the testing dataset. As described earlier, the decision trees in the RF model are built by selecting

predictors that provide the greatest reduction in impurity. Although there are measures to ensure that the same predictor is not selected repeatedly at each node, the number of times a predictor is selected is indicative of its relative entropy compared to other predictors. Therefore, the higher the PI, the more important the predictor is in the model.

Transfer Learning

In an ideal case, a model trained on MEPS patients should be able to predict medication adherence for patients from PatientsLikeMe with the same level of accuracy. As an analogy, a human trained to drive a given vehicle should be able to transfer this knowledge to the driving of a different vehicle. However, transferring this knowledge is not simple in either case. The success of this transfer is subject to disparities in data-collection methods, variable mapping, and population distribution between the source and target domains.

Despite the abovementioned difficulties, transfer learning offers several benefits. For instance, the approach has been used to transfer user behavior and content knowledge from one social network to another [15,16]. The approach has also been used for the predictive modeling of the relationship between transcription factor-binding sites and gene expression from one human cell line to another [17]. Transfer learning can help relax the accuracy requirements that are often associated with traditional machine learning methods [31]. Specifically, it can make a medication-adherence model that is developed and validated in a controlled environment accessible to a large-size population with a limited financial burden through, for example, social health forum services. In general, the proposed approach exemplifies the transfer of health models from the proprietary domain to the public domain.

Working toward this objective, a medication-adherence model was initially trained on MEPS patients using all the predictors extracted from the dataset. It was then tested on both the MEPS and the PatientsLikeMe patients. Ideally, the model should be able to predict medication adherence for both population groups with the same level of accuracy. However, this transfer was dependent on the adequacy of the mapping between the predictors and the outcomes in the two domains. This aspect is particularly important in this study owing to the absence of an absolute ground truth for medication adherence and the fact that adherence is equated to medication refills in MEPS and to a subjective self-reported assessment by the patient in PatientsLikeMe.

In order to understand the potential and limitation of transfer learning for medication adherence between MEPS and PatientsLikeMe, the following procedure was adopted:

- The prediction accuracy results for both domains were compared. A significant difference in the accuracy between the two domains is indicative of predictors that fail to transfer from the source domain (MEPS) to the target domain (PatientsLikeMe).
- A secondary model was developed by using the target domain dataset (in this case, PatientsLikeMe). The purpose of this model is to provide an understanding of the differences in the importance of each predictor across the two domains.
- Guided by the abovementioned analysis and using a reductionist approach, a set of predictors that do not transfer from MEPS to PatientsLikeMe was removed.
- A revised model was then developed with the remaining set of predictors, which are deemed transferrable between the two domains. The transfer capabilities of the revised model were then re-evaluated using both the MEPS and the PatientsLikeMe datasets.

Results

Overview

The findings derived from the application of the methodology described in the previous section are presented below. These results highlight the salient characteristics of the datasets, the predictive accuracy of the model developed by using the traditional machine learning approach, and the potential applicability of this model to a different domain when appropriate transfer learning requirements are taken into consideration.

Data Extraction and Cross-Domain Variable Mapping

The demographic breakdown of the MEPS and PatientsLikeMe cohorts is shown in Table 2. The average PatientsLikeMe patient was about 10 years younger than the average MEPS patient. With respect to the region of residence, the largest difference was observed in the southern region. The MEPS patient population in this region accounted for approximately 40% of the total population. However, the PatientsLikeMe dataset had a southern patient population that barely exceeded 30%. Moreover, compared to male patients, female patients accounted for the majority of the cohort for both datasets. However, the female population was significantly larger in the PatientsLikeMe (91%) dataset compared to the MEPS (63.9%) dataset.

In addition to understanding the differences in the demographic distribution of the patients across the two domains, an understanding of the distribution of the patients into adherent/nonadherent classes is crucial. As previously mentioned, the MEPS dataset does not contain a direct adherence metric. Therefore, based on previous studies [21], adherence was derived from the MRA. Since the MRA threshold that distinguishes between adherent/nonadherent patients is not known a priori, several values (ie, 35%, 45%, 65%, and 80%) for the threshold were considered. The distribution of the adherent/nonadherent MEPS patients for each threshold is shown in Table 3. PatientsLikeMe patients self-reported adherence (adherent: n=281 [79%]; nonadherent: n=76 [21%]).

Table 2. Demographics of MEPS[a] and PatientsLikeMe patients with fibromyalgia.

Characteristic	MEPS patients (N=3044)	PatientsLikeMe patients (N=357)
Age (years), mean (SD)	58.0 (14.9)	49.1 (10.7)
Sex, n (%)		
Male	995 (32.7)	27 (7.6)
Female	1945 (63.9)	325 (91.0)
Unknown	103 (3.4)	5 (1.4)
Region, n (%)		
Northeast	399 (13.1)	55 (15.4)
Midwest	651 (21.4)	84 (23.5)
South	1309 (43.0)	110 (30.8)
West	560 (18.4)	92 (25.8)
Unknown	125 (4.1)	16 (4.5)

[a]MEPS: Medical Expenditure Panel Survey

Table 3. Distribution of adherent and nonadherent Medical Expenditure Panel Survey patients for varying medication refill–adherence thresholds.

MRA[a] threshold (%)	Adherent patients, n (%)	Nonadherent patients, n (%)
80	714 (23)	2330 (77)
65	906 (30)	2138 (70)
45	1314 (43)	1730 (57)
35	1571 (52)	1473 (48)

[a]MRA: medication-refill adherence.

Model Training and Validation

Models for fibromyalgia were trained and tested using the MEPS dataset at different MRA threshold values. For each model, the dataset was split into a training and a testing dataset following an 80/20 split. Moreover, patients were randomly removed from the higher represented class in each training dataset until a 50/50 balance between adherent and nonadherent patients was obtained. For instance, 1616 nonadherent patients selected at random were removed from the model at the 80% MRA threshold.

The results in Table 4 show that the highest predictive accuracy was obtained when the MRA threshold was set to 35%. This indicates that the MEPS models are better at differentiating between extremely nonadherent patients and moderately or highly adherent patients. Based on this result, the MRA threshold of 35% was adopted in the remainder of the study. An intended direction for future work is investigating multiclass adherence models that can differentiate between highly and moderately adherent patients.

Transfer Learning

The MEPS fibromyalgia model developed in the previous section was tested using the PatientsLikeMe dataset. The prediction accuracy for this dataset was 54.9%, which is significantly lower than the accuracy obtained with the MEPS dataset (76.2%; Table 4). This significant difference indicates that some of the predictors did not adequately transfer from MEPS to PatientsLikeMe. To investigate this result, the PI values for the predictors in the MEPS model were compared to the PI values for the predictors of a secondary model that was trained using the PatientsLikeMe dataset (Table 5).

Predictors with large differences in PI values across the two domains suggest that a predictor has a higher significance in one domain than in the other domain. Based on the results of Table 5, the PI values of both daily intake and out-of-pocket expense differ by approximately 4% across the two domains, while none of the other predictors show a difference of more than ~1%.

Guided by this result, a reduced MEPS model was created after the elimination of the two predictors daily intake and out-of-pocket expense. The performance of the reduced model for both testing datasets (ie, MEPS and PatientsLikeMe) is reported in Table 6. Removing the two predictors significantly improved the accuracy and the F_1 score for the PatientsLikeMe dataset. As expected, however, for the MEPS testing dataset, a slight reduction in accuracy was observed in the reduced model as compared to the original model. Typically, in traditional machine learning methods, more predictors yield higher accuracy models. However, for transfer learning, these predictors must also map adequately from the source domain to the target domain.

Table 4. Accuracy and F_1 scores of fibromyalgia medication-prediction models that were trained and tested using the Medical Expenditure Panel Survey dataset. Models were created for varying medication refill–adherence thresholds.

MRA[a] threshold (%)	Patients in the training dataset, n (%)	Patients in the testing dataset, n (%)	Model accuracy (%)	F_1 score
80	1143 (80)	285 (20)	70.2	72.5
65	1450 (80)	362 (20)	69.9	70.0
45	2103 (80)	525 (20)	73.0	74.6
35	2436 (80)	609 (20)	76.2	77.1

[a]MRA: medication-refill adherence.

Table 5. Predictor importance for each predictor in the MEPS[a]–trained model and the PatientsLikeMe-trained model. Both models were tested using the MEPS dataset.

Predictor	Predictor importance	
	Model trained by MEPS	Model trained by PatientsLikeMe
Type of medication	24.2	24.7
Years taking treatment	11.9	12.5
Daily intake	11.8	5.3
Dosage	14.4	14.4
Out-of-pocket expense	9.7	13.6
Region of living of the patient	8.5	9.6
Age at the end of the study or last known age	15.2	16.1
Sex of the patient	4.3	3.8

[a]MEPS: Medical Expenditure Panel Survey.

Table 6. Accuracy and F_1 score of the reduced MEPS[a]–trained model for the MEPS and PatientsLikeMe testing datasets. For comparison, the second row category of the table repeats the results previously obtained for the original model with all predictors from Table 4.

Model	MEPS dataset	PatientsLikeMe dataset
Reduced model (without daily intake and out-of-pocket expense)		
Accuracy (%)	73.2	67.8
F_1 score	73.9	79.7
Original model (with daily intake and >out-of-pocket expense)		
Accuracy (%)	76.2	54.9
F_1 score	77.1	65.5

[a]MEPS: Medical Expenditure Panel Survey.

Additional investigation is needed to analyze why some of the predictors transfer, whereas others do not. We speculate that some of the root causes can be attributed to potential over/underreporting by the patients [22] and to differences in the sociodemographic distribution of the patients. For instance, 80% of the MEPS patients had an out-of-pocket expense <US $25 each month, whereas less than 40% of the patients in PatientsLikeMe reported an out-of-pocket expense <US $25.

Discussion

Principal Results

This study shows that it is possible to develop and validate a model for fibromyalgia medication adherence in a controlled environment and then apply it widely through social health forums. A model trained using MEPS patients with fibromyalgia was able to predict adherence with an accuracy of 73.2% and an F_1 score of 73.9 for other MEPS patients. Traditionally, for this model to benefit a wider population of patients, these patients would have to be enrolled in the MEPS survey, an approach that is neither practical nor cost-effective. This paper shows that the MEPS model can be transferred to the social health forum PatientsLikeMe with careful mapping between the variables in each domain. The proposed approach was tested with PatientsLikeMe patients with fibromyalgia, and the MEPS-trained model was able to predict adherence for these patients with an accuracy of 67.8% and an F_1 score of 79.7.

Limitations

An initial design of the model showed that two of the predictors (daily intake and out-of-pocket expense) in MEPS failed to adequately transfer to PatientsLikeMe. Additional investigation is needed to understand the root cause of this lack of transfer. Access to additional demographic information about the patients may help with this investigation. However, information including race, education, and social status was not available in this study. Furthermore, although MRA was previously shown to provide a good estimate of adherence, there are certain cases where this threshold could misclassify a patient as adherent, since there are no assurances that the medication is actually being taken by the patient.

Future Work

Future work will consider transfer learning in the context of multiclass adherence models (ie, always, usually, sometimes, or never taken as prescribed). In addition, we would like to study transfer learning for other diseases and other datasets including other social media sources. Finally, understanding the impact of missing values, an unavoidable characteristic of social media data, and development of learning techniques that can handle missing values are areas open for continued research.

Conclusions

The transferability of a model developed and validated in a controlled environment to the wider public provides tremendous possibilities for improved population health. One can imagine a model for medication adherence derived by a health institution deployed in PatientsLikeMe and enabling the users of this forum to receive alerts or targeted educational material when they are flagged to be at risk of nonadherence.

Transfer learning from one domain to another can be extended, perhaps to disease prediction or other health-related models. This research showed that robust models that can systematically transfer from one domain to another are possible and that it is important to understand the limitations of this transfer. We showed that transfer learning between MEPS and PatientsLikeMe produced similar accuracy of medication-adherence prediction. This approach can have a significant advantage, even if this advantage comes at a slight reduction in prediction accuracy compared with costly, institution-specific machine learning models.

Conflicts of Interest

KH and MM are currently employed by Eli Lilly and Company. ZBM was previously funded by Eli Lilly and Company and Merck Sharp & Dohme Corp, a subsidiary of Merck & Co, Inc.

References

1. Philipson T. Forbes. 2015 May 08. Non-adherence in health care: Are patients or policy makers ill-informed? URL: https://tinyurl.com/y67nbyr6 [accessed 2018-09-20] [WebCite Cache ID 73HljNhJm]
2. Iuga AO, McGuire MJ. Adherence and health care costs. Risk Manag Healthc Policy 2014;7:35-44 [FREE Full text] [doi: 10.2147/RMHP.S19801] [Medline: 24591853]
3. Martin LR, Williams SL, Haskard KB, Dimatteo MR. The challenge of patient adherence. Ther Clin Risk Manag 2005 Sep;1(3):189-199 [FREE Full text] [Medline: 18360559]
4. Friedman RH, Kazis LE, Jette A, Smith MB, Stollerman J, Torgerson J, et al. A telecommunications system for monitoring and counseling patients with hypertension. Impact on medication adherence and blood pressure control. Am J Hypertens 1996 Apr;9(4 Pt 1):285-292. [Medline: 8722429]
5. Express Scripts. 2013 Apr 24. Predicting adherence URL: http://lab.express-scripts.com/lab/insights/adherence/infographic-predicting-rx-nonadherence [accessed 2018-09-20] [WebCite Cache ID 73Hm26jbS]
6. AllazoHealth. New York: AllazoHealth Driving Behavioral Change to Improve Patients Outcomes Through Precision Engagement URL: https://allazohealth.com [accessed 2018-10-19] [WebCite Cache ID 73HmOZpXt]
7. FICO. FICO medication adherence score URL: https://www.fico.com/en/products/fico-medication-adherence-score [accessed 2018-10-19] [WebCite Cache ID 73HmhLDii]
8. Lin H, Jia J, Guo Q, Xue Y, Li Q, Huang J, et al. User-level psychological stress detection from social media using deep neural network. : ACM; 2014 Nov 03 Presented at: International Conference on Multimedia; November 3-7, 2014; Orlando. [doi: 10.1145/2647868.2654945]
9. Paul MJ, Dredze M. You are what you tweet: analyzing twitter for public health. : AAAI; 2011 Jul 17 Presented at: AAAI Conference on Weblogs and Social Media; July 17, 2011; Barcelona URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2880/3264
10. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug Saf 2014 May;37(5):343-350 [FREE Full text] [doi: 10.1007/s40264-014-0155-x] [Medline: 24777653]
11. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
12. Knight E, Werstine RJ, Rasmussen-Pennington DM, Fitzsimmons D, Petrella RJ. Physical therapy 2.0: leveraging social media to engage patients in rehabilitation and health promotion. Phys Ther 2015 Mar;95(3):389-396 [FREE Full text] [doi: 10.2522/ptj.20130432] [Medline: 24627429]

13. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using Artificial Intelligence to Reduce the Risk of Nonadherence in Patients on Anticoagulation Therapy. Stroke 2017 Dec;48(5):1416-1419 [FREE Full text] [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
14. Klein A, Sarker A, Rouhizadeh M, O'Conner K, Gonzalez G. Detecting personal medication intake in twitter: an annotated corpus and baseline classification system. In: BioNLP.: Association for Computational Linguistics; 2017 Aug 04 Presented at: BioNLP; October 4, 2017; Vancouver. [doi: 10.18653/v1/w17-2316]
15. Sun J, Staab S, Kunegis J. Understanding Social Networks Using Transfer Learning. Computer 2018 Jun;51(6):52-60. [doi: 10.1109/MC.2018.2701640]
16. Roy SD, Mei T, Zeng W, Li S. SocialTransfer: cross-domain transfer learning from social streams for media applications. In: ACM.: ACM; 2012 Oct 29 Presented at: Multimedia; October 29, 2012; Nara. [doi: 10.1145/2393347.2393437]
17. Zou N, Baydogan M, Zhu Y, Wang W, Zhu J, Li J. A Transfer Learning Approach for Predictive Modeling of Degenerate Biological Systems. Technometrics 2015;57(3):362-373 [FREE Full text] [doi: 10.1080/00401706.2015.1044117] [Medline: 27330227]
18. Cutler RL, Fernandez-Llimos F, Frommer M, Benrimoj C, Garcia-Cardenas V. Economic impact of medication non-adherence by disease groups: a systematic review. BMJ Open 2018 Dec 21;8(1):e016982 [FREE Full text] [Medline: 29358417]
19. US Department of Health & Human Services. Medical Expenditure Panel Survey URL: https://meps.ahrq.gov/mepsweb/ [accessed 2018-10-19] [WebCite Cache ID 73HmqVH3R]
20. PatientsLikeMe. URL: https://www.patientslikeme.com [accessed 2018-10-19] [WebCite Cache ID 73HmwVe31]
21. Hess LM, Raebel MA, Conner DA, Malone DC. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. Ann Pharmacother 2006;40(7-8):1280-1288. [doi: 10.1345/aph.1H018] [Medline: 16868217]
22. Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME. It's not that I don't have problems, I'm just not putting them on Facebook: Challenges and opportunities in using online social networks for health. : ACM; 2011 Mar 19 Presented at: Computer supported cooperative work; March 19, 2011; Hangzhou p. 341-350. [doi: 10.1145/1958824.1958876]
23. Breiman L. Random forests. In: Machine Learning. The Netherlands: Kluwer Academic Publishers; 2001:5-32.
24. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
25. Hearst M, Dumais S, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Their Appl 1998 Jul 10;13(4):18-28. [doi: 10.1109/5254.708428]
26. Gunning D. DARPA. 2017. Explainable Artificial Intelligence URL: https://www.darpa.mil/attachments/XAIProgramUpdate.pdf [accessed 2019-01-04] [WebCite Cache ID 75BGK9zPw]
27. Dittman D, Khoshgoftaar TM, Wald R, Napolitano A. Random forest: A reliable tool for patient response prediction. : IEEE; 2011 Nov 12 Presented at: International Conference on Bioinformatics and Biomedicine Workshops; November 12, 2011; Atlanta. [doi: 10.1109/bibmw.2011.6112389]
28. Gulia A, Vohra R, Rani P. Liver patient classification using intelligent techniques. International Journal of Computer Science and Information Technologies 2014;5(4):5110-5115 [FREE Full text]
29. Breiman L. Bagging predictors. Mach Learn 1996;24(2):123-140. [doi: 10.1007/BF00058655]
30. Raileanu LE, Stoffel K. Theoretical Comparison between the Gini Index and Information Gain Criteria. Annals of Mathematics and Artificial Intelligence 2004 May;41(1):77-93. [doi: 10.1023/B:AMAI.0000018580.96245.c6]
31. 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]

Abbreviations

MEPS: Medical Expenditure Panel Survey
MRA: medication-refill adherence
PI: predictor importance
RF: random forest

Edited by M Focsa; submitted 20.10.18; peer-reviewed by T Kass-Hout, M Warner; comments to author 26.11.18; revised version received 20.01.19; accepted 21.01.19; published 04.04.19.

Please cite as:
Haas K, Ben Miled Z, Mahoui M
Medication Adherence Prediction Through Online Social Forums: A Case Study of Fibromyalgia
JMIR Med Inform 2019;7(2):e12561
URL: http://medinform.jmir.org/2019/2/e12561/
doi:10.2196/12561
PMID:30946020

Original Paper

A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study

Hervé Spechbach[1], MD; Johanna Gerlach[2], PhD; Sanae Mazouri Karker[1], MD; Nikos Tsourakis[2], PhD; Christophe Combescure[1], PhD; Pierrette Bouillon[2], Prof Dr

[1]Geneva University Hospitals, Genève, Switzerland

[2]Faculté de traduction et d'interprétation, Université de Genève, Genève, Switzerland

Corresponding Author:
Pierrette Bouillon, Prof Dr
Faculté de traduction et d'interprétation
Université de Genève
40, bd du Pont d'Arve
Genève, 1211
Switzerland
Phone: 41 223798701
Email: Pierrette.bouillon@unige.ch

Abstract

Background: In the context of the current refugee crisis, emergency services often have to deal with patients who have no language in common with the staff. As interpreters are not always available, especially in emergency settings, medical personnel rely on alternative solutions such as machine translation, which raises reliability and data confidentiality issues, or medical fixed-phrase translators, which sometimes lack usability. A collaboration between Geneva University Hospitals and Geneva University led to the development of BabelDr, a new type of speech-enabled fixed-phrase translator. Similar to other fixed-phrase translators (such as Medibabble or UniversalDoctor), it relies on a predefined list of pretranslated sentences, but instead of searching for sentences in this list, doctors can freely ask questions.

Objective: This study aimed to assess if a translation tool, such as BabelDr, can be used by doctors to perform diagnostic interviews under emergency conditions and to reach a correct diagnosis. In addition, we aimed to observe how doctors interact with the system using text and speech and to investigate if speech is a useful modality in this context.

Methods: We conducted a crossover study in December 2017 at Geneva University Hospitals with 12 French-speaking doctors (6 doctors working at the outpatient emergency service and 6 general practitioners who also regularly work in this service). They were asked to use the BabelDr tool to diagnose two standardized Arabic-speaking patients (one male and one female). The patients received a priori list of symptoms for the condition they presented with and were instructed to provide a negative or noncommittal answer for all other symptoms during the diagnostic interview. The male patient was standardized for nephritic colic and the female, for cystitis. Doctors used BabelDr as the only means of communication with the patient and were asked to make their diagnosis at the end of the dialogue. The doctors also completed a satisfaction questionnaire.

Results: All doctors were able to reach the correct diagnosis based on the information collected using BabelDr. They all agreed that the system helped them reach a conclusion, even if one-half felt constrained by the tool and some considered that they could not ask enough questions to reach a diagnosis. Overall, participants used more speech than text, thus confirming that speech is an important functionality in this type of tool. There was a negative association (P=.02) between the percentage of successful speech interactions (spoken sentences sent for translation) and the number of translated text items, showing that the doctors used more text when they had no success with speech.

Conclusions: In emergency settings, when no interpreter is available, speech-enabled fixed-phrase translators can be a good alternative to reliably collect information from the patient.

(JMIR Med Inform 2019;7(2):e13167) doi:10.2196/13167

KEYWORDS

anamnesis; emergencies; tools for translation and interpreting; fixed-phrase translator; speech modality

Introduction

Background

In the context of the current refugee crisis, emergency services are increasingly confronted with patients who have no language in common with staff and may not share the same culture. For example, at Geneva University Hospitals (HUG), 52% of patients are foreigners and 10% speak no French at all. In 2017, the 10 languages for which interpretation services were the most solicited were Tigrinya, Tamil, Albanian, Farsi, Spanish, Somalian, Syrian, Dari, Portuguese, and Arabic (North Africa). Taken together, these languages represent 75% of the interpreting hours at HUG (Geneva University Hospitals, personal communication, 2017).

This language barrier situation is known to pose many safety and ethical problems: It is responsible for increased risks for patients [1] and is very expensive. For example, as reported by Rechel et al in 2003 [2], the United States Institute for Healthcare Advancement estimated that US $73 billion was wasted annually in the United States as a result of communication problems in health care, many of which originate from language differences. Both ethically and legally, hospitals have a duty to offer all patients the same quality of care, including the right to have a dialogue with health professionals.

Different solutions are available for use in emergency settings to address these language barriers, but they all have their drawbacks. Phone-based interpreter services, which are the most common solution, are generally considered adequate, but they are expensive (3 Swiss francs/minute with AOZ Medios, a national interpreting service mandated by the Swiss Federal Office of Public Health), not always available for some languages, and less satisfactory than face-to-face interaction with a physically present interpreter [3]. Asking patients' relatives to translate speech is known to create substantial risks [1]. Machine translation, such as Google Translate, another low-cost solution more commonly used in emergency contexts, is also extremely problematic, as this type of tool has not been developed for medical use. Some recent studies have estimated that nearly 40% of sentences of medical speech translated by Google Translate are mistranslated [4,5]. However, such systems also pose ethical problems and are not currently compatible with the Swiss Data Protection Law. A plethora of specialized systems have also been developed for medical communication, both in the academic and industry settings (including fixed-phrase translation or machine translation systems [6]), but it is not always clear how they were built or evaluated and if they are extensible. As emphasized in the recent review by Dew et al [6], there is a lack of criteria for the development and evaluation of these systems, which impedes the adoption of these systems in emergency settings.

For these reasons, we have developed a new type of speech-enabled fixed-phrase translation tool for medical dialogue (BabelDr [7]), based on our previous experience in the field [8] in a collaborative venture between HUG and the University of Geneva Faculty of Translation and Interpreting. This tool is a compromise between speech-to-speech machine translation and fixed-phrase translation systems and directly addresses specific needs in emergency settings (ie, high accuracy, extensibility, portability to low-resource languages and domains, and data security). It was also designed as a way to collect doctor-patient dialogues and thereby improve our understanding of the criteria for the development of this type of system.

This study is the first step in this direction. It aims to determine whether this type of restricted translation tool can be used by doctors to perform a diagnostic interview and reach a correct diagnosis and to quantify if speech adds value to fixed-phrase translators. Although different evaluations of medical devices have been conducted [6], to the best of our knowledge, this is the first study that attempts to show the impact of "phraselators" on the diagnosis and to define a methodology to achieve this.

The BabelDr App

The BabelDr app can be characterized as a "phraselator" [9,10]. Similar to well-known medical fixed-phrase translation apps such as Medibabble [11] or UniversalDoctor [12], the system relies on a set of predefined sentences (mostly yes/no medical questions or instructions) translated by human translators to ensure translation reliability. However, in contrast to traditional fixed-phrase translators, the doctor can also freely ask his/her question and the system will match the recognition result to the closest predefined sentence in the list. The app was designed from the beginning to meet the hospital's needs. In particular, it is easy to extend it to new target languages and situations in order to follow demographic changes and allow its integration in different services. The content is described efficiently with rules (synchronized grammar [13]) that map multiple synonymous patterns ("variations") to a sentence expressing the core meaning ("core sentence"). For example, "Do you have a fever?" "Is your temperature high?" and "Have you observed a high temperature?" will all be mapped to the core sentence "Do you have fever?" In addition, patterns with variables (eg, "Is it a QUALITATIVE pain?" "Do you have a QUALITATIVE pain?" etc, where "QUALITATIVE" is a variable that can take multiple values such as "severe" and "dull") allow the description of content in a productive way. The system currently contains around 2500 patterns and 600 variables, linking more than one billion variations to approximately 25,000 core sentences. Translation follows the usual standards and is performed online with translation memory in two steps—translation of patterns followed by revision of complete sentences [14]. Target languages focus on the languages important for HUG (Spanish, Arabic, Swiss French sign language, Tigrinya, Farsi, Dari, and Albanian). To ensure data confidentiality, both speech recognition and translation are carried out on secure local servers and all interactions are saved locally.

For speech recognition and matching, the system combines rule-based and robust methods, derived from the rules. When the doctor speaks, the system first recognizes what is said using both a grammar-based version of "Nuance" and a specialized statistical version (Nuance Communications Inc, Burlington, MA). It then maps the recognition results to the closest core sentence using both rules and robust matching techniques borrowed from information retrieval, described in detail

elsewhere [15]. This closest core sentence is then translated orally for the patient who will answer nonverbally. As it is not an exact translation of the doctor's question, but a translation of the corresponding matched sentence, the core sentence is always echoed back to the doctor, so that he can verify what the system understood. The translation is thus only produced for the patient if the core sentence is approved by the doctor. Therefore, core sentences play a crucial role in the process by not only providing feedback to doctors concerning recognition accuracy, but also making the meaning of the sentence explicit for both translators and patients [16,17]. These core sentences were designed very carefully with doctors and translators, so that they are as accessible and explicit as possible in order to avoid communication problems. In addition to using core sentences for the verification of translations, users can also access them directly by browsing and searching via keywords. The associated translations can then be submitted without the need for further checking, similar to other phraselators [16].

Figure 1 illustrates the BabelDr interface and how an interaction is carried out. The doctor first selects the diagnostic domain based on the main patient complaint (headache, abdominal pain, dermatological problem, etc) and the language and gender of the patient (male or female). He/she can then speak a sentence ("speech interaction"). If the echoed core sentence corresponds to what the doctor wants to ask, he/she can click on it to produce the translation for the patient. In addition to speech input, doctors can search the list of core sentences using keywords (only with exact matching, as in traditional phraselators) and click on sentences to translate them for the patient ("text translated").

After translating a sentence to the patient (Figure 2), the translation is produced both in text and spoken form. The coverage list is automatically scrolled to the latest core sentence translated, giving quick access to related questions. The translated sentence is also added to a history list that can be downloaded as a PDF at the end of the dialogue.

Figure 1. Screenshot of the BabelDr app.

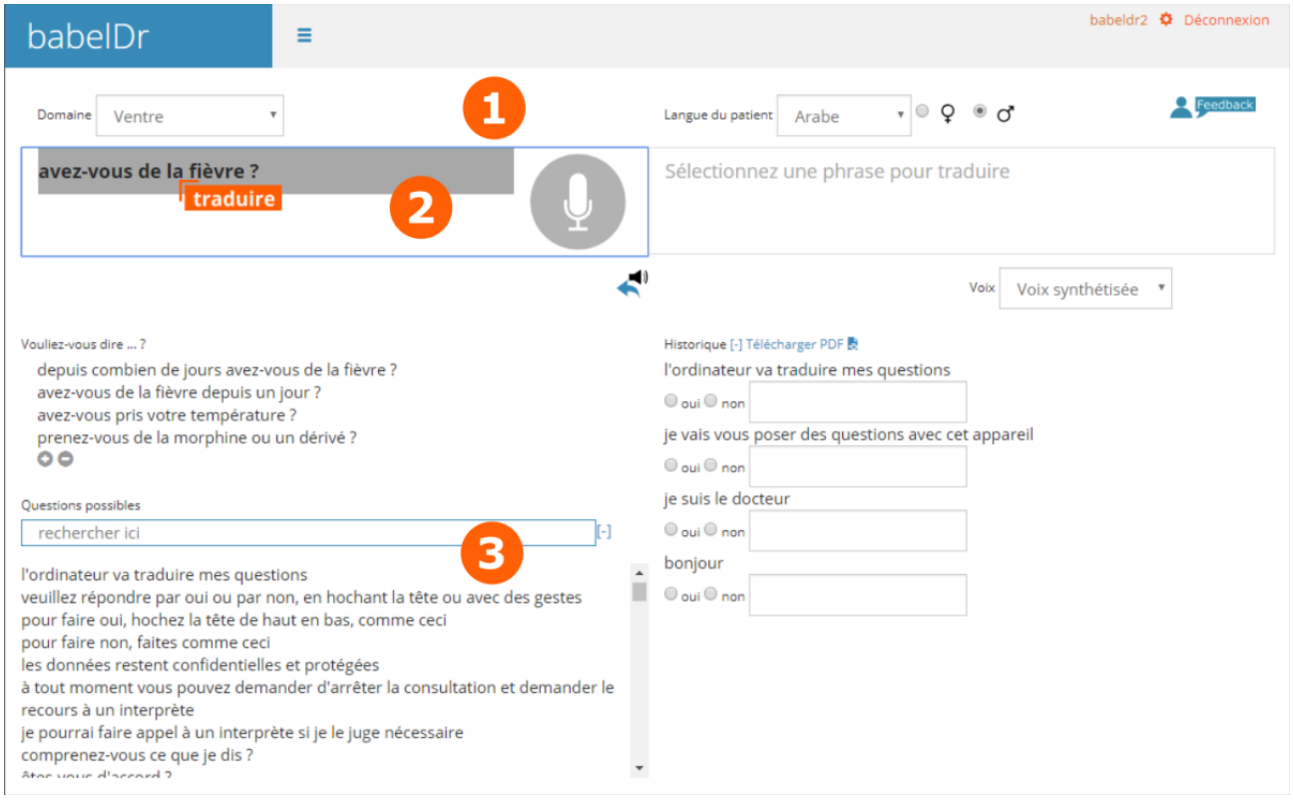
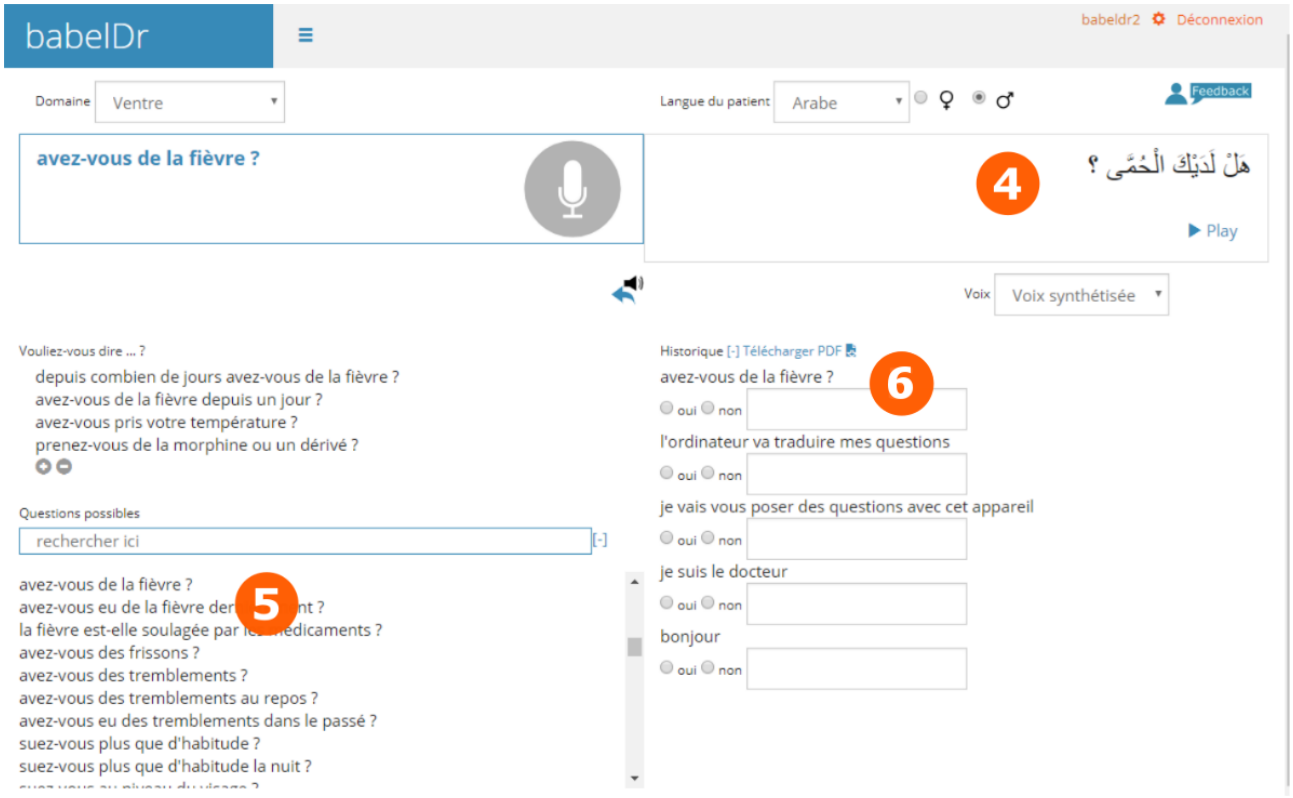


Figure 2. Example of the interface after the doctor translated “Avez-vous de la fièvre?” (Do you have fever?).



Methods

Identifying the Research Questions

This study aims (1) to determine whether a restricted translation tool like BabelDr can be used by doctors to perform a diagnostic interview and reach a correct diagnosis and (2) to quantify how doctors use text versus speech interactions in order to investigate if speech adds value to fixed-phrase translators. Our hypotheses were that this type of tool would demonstrate good functional suitability (doctors can collect all the information necessary to reach a diagnosis in an efficient way) and usability (doctors will use more speech to interact than text, as speaking should allow them to communicate more naturally, like when working with interpreters).

Design

The study was conducted at the HUG research laboratory in December 2017. In this crossover trial, 12 French-speaking doctors were asked to use BabelDr to diagnose two standardized Arabic-speaking patients (one male and one female) whose main complaint was lower back pain. The male patient was standardized for nephritic colic and the female patient, for cystitis. These two diagnoses are among the 10 most frequent at HUG (Geneva University Hospitals, personal communication, 2018). Each of the 12 doctors carried out a diagnostic interview with both patients, where half of the doctors began with the male patient and the other half began with the female patient.

Before the diagnostic interviews, doctors were informed about the main patient complaint (pain in the lower back). At the beginning of the session, they received a short introduction to BabelDr and tested a few interactions. It was strongly suggested that they use complete sentences and ask yes/no questions, so that the patients could answer nonverbally.

Tool and Interface

Doctors only had access to the BabelDr tool. The diagnostic domain was set to “lower back pain” to match the patient complaint. In the context of this study, the other domains were not made available in order to simplify system usage. It was ascertained beforehand that all available questions potentially relevant to the patient complaint were included in this domain. The language pair was French to Arabic; the male or female patient was chosen depending on the case.

Data Collection and Analysis

Diagnoses

During the sessions, the doctors wrote down the information they were able to collect based on the patient’s responses. At the end of each session, the doctors wrote down their diagnoses. These data allowed us to answer the first question on whether the system enables doctors to reach a correct diagnosis.

System Usage

All interactions with the system were logged. For each session, we collected audio recordings of each spoken interaction with the system as well as the corresponding recognition results. We also logged which recognition results or text examples the doctors chose to translate for the patients. Finally, the duration of each session was measured. These data were analyzed to provide a quantitative answer to our second research question, namely, whether speech interaction is useful in this type of tool.

User Satisfaction

At the end of each session, participants completed a satisfaction questionnaire that included a total of 23 questions. The questions were derived from the System Usability Scale questionnaire by Brooke [18] and adapted to the functionalities of BabelDr, especially the speech and core sentence mapping aspects. Questions covered usability and learnability aspects of the BabelDr system during the study (7 items), appropriateness of the system to confidently reach a diagnosis (6 items), the speech component of the system (3 items), and the user's opinion regarding the usefulness of such a system in their daily medical practice (7 items). A 5-point Likert scale ("strongly disagree," "disagree," "neutral," "agree," and "strongly agree") was used to rate agreement with question items. These data contribute to a qualitative answer to our second research question.

Participants

Doctors

Study participants were 12 French-speaking doctors: 6 from the emergency service at HUG and 6 general practitioners who also regularly work in this service. All work in French, but three were not native speakers (#6, #11, #12). Only one doctor (#6) had previously used a former version of BabelDr in another study [5].

Standardized Patients

Of the two Arabic standardized patients, one was a man from Syria and one was a woman from Jordan. Both were refugees and recruited from among master's degree students at the Faculty of Translation and Interpreting. They had a high level of literacy, but no specific medical knowledge. Neither of the patients spoke French. One week before the experiment, both patients received an a priori list of symptoms for the condition they were to present, expressed in layman's terms. They were instructed to provide a negative or noncommittal answer to questions relating to other symptoms during the diagnostic interview.

All participants received remuneration for their participation in the study.

Ethical Considerations

The institutional ethics committee approved the study protocol (Req-2017-00996). Participation in the study was voluntary, with written agreement obtained from all doctors and patients. All data were anonymous and stored on a secure University of Geneva server.

Results

Diagnoses

Doctors were able to reach a correct diagnosis in all 24 sessions based on the information collected using BabelDr. For the renal colic scenario, four doctors proposed multiple related diagnoses (Table 1). These results showed that BabelDr was suitable for the task and allowed doctors to collect information reliably.

Textbox 1 gives examples of the most frequently asked questions for each scenario. In total, more questions were translated for the renal colic scenario than for the cystitis one (170 vs 126 unique interactions, respectively), probably reflecting the fact that the first scenario was more complex due to a larger number of possible related diagnoses and thus required more different questions.

Table 1. Diagnoses made by the 12 doctors.

Doctor no.	Female patient (with cystitis)		Male patient (with renal colic)	
	Diagnosis	Other diagnoses	Diagnosis	Other diagnoses
1	Cystitis	No	Renal colic	Pyelonephritis
2	Cystitis	No	Renal colic	No
3	Cystitis	No	Renal colic	No
4	Cystitis	No	Renal colic	No
5	Cystitis	No	Renal colic	No
6	Cystitis	No	Renal colic	Lumbosciatica
7	Cystitis	No	Renal colic	No
8	Cystitis	No	Renal colic	No
9	Cystitis	No	Renal colic	Pyelonephritis, lumbosciatica
10	Cystitis	No	Renal colic	No
11	Cystitis	No	Renal colic	No
12	Cystitis	No	Renal colic	Pyelonephritis, lumbosciatica, appendicitis

Textbox 1. Most frequently translated core sentences for each scenario, sorted by frequency.

Female patient with cystitis:

- *Pouvez-vous me montrer avec le doigt où est la douleur?* [Could you point with your finger to where it hurts?]
- *Avez-vous déjà eu ce type de douleur?* [Have you already had this type of pain?]
- *Bonjour* [Hello]
- *Je suis le docteur* [I'm the doctor]
- *Quand vous urinez, est-ce que ça brûle?* [Do you feel a burning sensation when you urinate?]
- *Avez-vous eu de la fièvre dernièrement?* [Have you had fever recently?]
- *Je vais m'occuper de vous aujourd'hui* [I will take care of you today]
- *Avez-vous mal au niveau des reins?* [Do you have pain in the kidney area?]
- *Je vais vous poser des questions avec cet appareil* [I will use this machine to ask you some questions]
- *Vos urines sont-elles rouges?* [Is your urine red?]
- *Êtes-vous d'accord?* [Do you agree?]
- *Il y a combien de semaines que vous avez eu vos dernières règles?* [How many weeks ago did you have your last period?]
- *Avez-vous été traité par antibiotique pour l'infection urinaire?* [Have you had antibiotic treatment for a urinary tract infection?]
- *Avez-vous eu une infection urinaire dernièrement?* [Have you recently had a urinary tract infection?]
- *Avez-vous des allergies connues?* [Do you have any known allergies?]

Male patient with renal colic:

- *Bonjour* [Hello]
- *Je suis le docteur* [I'm the doctor]
- *Vos urines sont-elles rouges* [Is your urine red?]
- *Avez-vous déjà eu ce type de douleur* [Have you had this kind of pain before?]
- *Pouvez-vous me montrer avec le doigt où est la douleur* [Could you point with your finger to where it hurts?]
- *Avez-vous eu de la fièvre dernièrement* [Have you had fever recently?]
- *Avez-vous mal au niveau des reins* [Do you have pain in the kidney area?]
- *La douleur aux reins irradie-t-elle vers un autre endroit* [Does the pain in the kidney area spread to any other place?]
- *Quand vous urinez, est-ce que ça brûle* [Do you feel a burning sensation when you urinate?]
- *Je vais vous poser des questions avec cet appareil* [I will use this machine to ask you some questions]
- *La douleur aux reins est-elle continue* [Is the pain in the kidney area continuous?]
- *Êtes-vous d'accord* [Do you agree?]
- *Je vais m'occuper de vous aujourd'hui* [I will take care of you today]
- *Depuis combien de jours avez-vous mal aux reins* [For how many days have you had pain in the kidney area?]
- *Avez-vous de la fièvre* [Do you have fever?]

Analysis of Interactions

For each doctor, we measured the time to complete the dialogue, the number of speech interactions, the number of speech interactions resulting in a translation for the patient, and the number of text items directly translated from the list of sentences. Table 2 shows that both the median time and the median number of translated speech interactions were higher for the renal colic scenario (16 min for 26 interactions) than for the cystitis scenario (13 min for 19 interactions), confirming the fact that the renal colic scenario was more complex.

Table 2 shows that doctors translated both speech and text, but used more speech interactions, suggesting that speech was generally preferred to text. The median number of speech interactions per dialogue that led to translations was 28.5 for the cystitis scenario and 36 for the renal colic scenario, whereas the median numbers for text interactions were 4.5 and 10, respectively.

Table 2. Time and number of interactions for both scenarios.

Variable	Female patient with cystitis, median (range)	Male patient with renal colic, median (range)
Time to diagnosis (min:seconds)	13:37 (4:09-35:37)	16:37 (4:35-23:35)
Speech interactions (n)	28.5 (17-46)	36 (20-66)
Speech translated (n)	19.5 (8-23)	26.5 (13-51)
Text translated (n)	4.5 (0-36)	10 (0-23)

Figures 3 and 4 present the interactions by participant and show that some used the text mode more often than others and that the number of speech sentences sent to translation differed from one participant to another. For different doctors, the proportions of recognition results leading to translations varied from 40% (8/20) to 94% (16/17) for the cystitis scenario and 37% (13/35) to 100% (20/20) for the renal colic scenario.

The association between the percentage of translated speech and the number of translated texts was investigated using a linear regression model. Since each medical practitioner assessed two patients, data were clustered. Therefore, a regression model with mixed effects was used: A random effect was set on the intercept to account for between-practitioner variability. In addition, a multivariable analysis was conducted to adjust for the session and the scenario.

Figure 3. Interactions by participant for the scenario with the female patient.

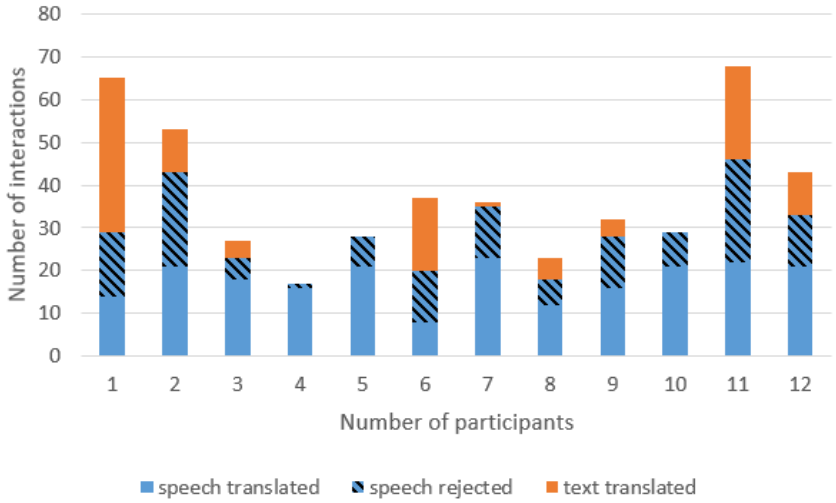
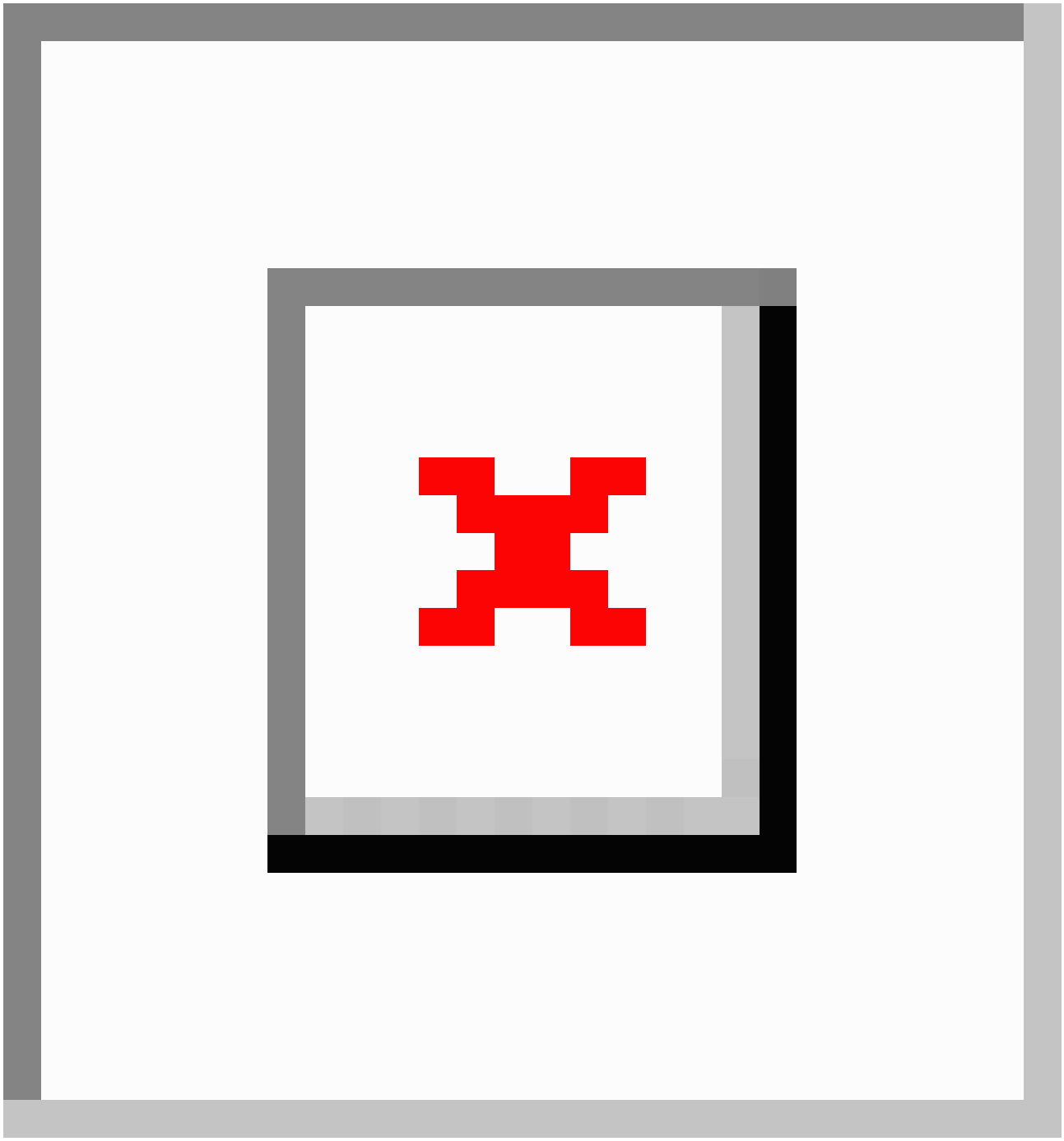


Figure 4. Interactions by participant for the scenario with the male patient.



The percentage of translated speech was negatively associated with the number of translated texts (*P*=.02): When the percentage of translated speech increased by 10%, the number of translated texts decreased by 2.6 (95% CI 0.7-4.4). After adjustment for the session and scenario, the decrease in the number of translated texts was similar (2.4; 95% CI 0.7-4.2; *P*=.02). This association is illustrated in Figure 5A. These results show that users who are not well recognized tend to use the text interface more often, thereby confirming the usefulness of including both modalities in such a tool.

The percentage of translated speech was higher in the second session than in the first session (difference=4.3%; 95% CI 1.1-7.4; *P*=.03). One possible interpretation may be that users familiarized themselves with system coverage in the first session and therefore used more coverage utterances in the second session, leading to better recognition of results and thus more translations.

Analyses by scenario showed that the proportion of translated speech was lower in the renal colic scenario than in the cystitis scenario (difference=4.3%; 95% CI –7.6 to –1.1; *P*=.03). This may be due to different factors such as concepts not covered by the system at the time of the study or errors in speech recognition or mapping to the core sentences (eg, cases where a sentence is badly recognized and therefore mapped to a different sentence). In some cases, the core sentence could also be too general or specific or considered inappropriate in the context. Table 3 presents some examples of these cases.

Figure 5. Association between the percentage of translated speech and the number of translated texts (A) and between French native speakers and the percentage of translated speech (B), system confidence score (C), and speech interaction (D). Circles represent each individual doctor's data; the black line represents the unadjusted regression line and black squares represent the mean values.

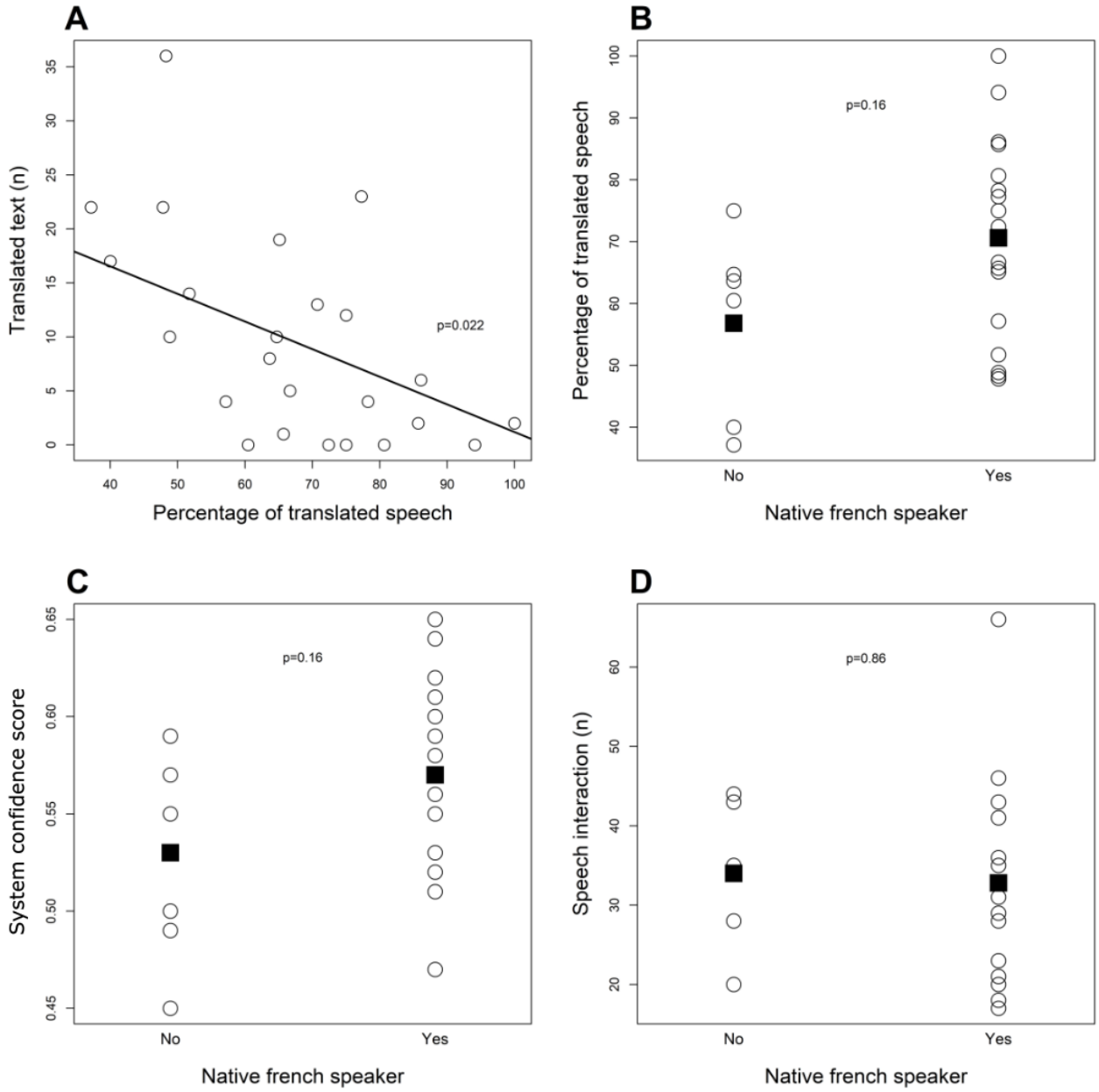


Table 3. Examples of transcriptions and mapped core sentences.

Speech utterances	Core sentences
Sent to translation	
Est-ce que vous avez fait du sport (Did you practice any sports?)	*Faites-vous de l'exercice* physique (Do you exercise?)
Est-ce qu'il y a du sang dans les urines (Is there any blood in your urine?)	*Les urines sont-elles rouges* (Is your urine red?)
Est-ce que vous avez beaucoup transpiré (Did you sweat a lot?)	*Suez-vous plus que d'habitude* (Do you sweat more than usual?)
Est-ce que c'est aujourd'hui (Is it today?)	*Avez-vous mal depuis aujourd'hui* (Do you have the pain since today?)
Est-ce que vous avez des pertes vaginales particulières (Have you observed any particular vaginal discharges?)	*Avez-vous des pertes blanches en dehors des règles* (Have you observed any white discharges outside normal menstruation?)
Not sent for translation by at least one doctor	
Avez-vous bu (Have you had anything to drink?)	*Avez-vous bu de* l'alcool (Have you consumed any alcohol?)
Est-ce que vous pourriez être enceinte (Could you be pregnant?)	*Êtes-vous enceinte* (Are you pregnant?)
Avez-vous du prurit (Do you have pruritus?)	*Avez-vous des* démangeaisons (Do you have itchiness?)
La douleur est-elle constante (Is the pain constant?)	*La douleur est-elle continue* (Is the pain continuous?)

Associations between French native speakers and the percentage of translated speech, system confidence, and speech interaction were also investigated using a linear regression model with fixed effects. No association was found between French native speakers and the percentage of translated speech (P=.16), system confidence (P=.16), and speech interaction (P=.86). Figure 5B-D illustrates these numbers. These results suggest that system performance is not significantly impaired by different accents.

User Satisfaction

Figures 6 and 7 show the results for seven questions related to the usefulness of the system for the diagnostic task and speech recognition included in the satisfaction questionnaires completed by the doctors after each dialogue (24 completed questionnaires). Overall, the doctors were satisfied with the speech interaction function and the usefulness of the system in the test context (19 negative, 54 neutral, and 116 positive judgments). All doctors considered that the system helped them reach a conclusion (Q3). They also liked the way the recognition result was presented (only one participant disagreed), which showed that they found the translation to the core sentence useful. All doctors thought that the system recognized their voice easily (Q4), and most believed that the system helped them to pose the question in a different way when the question could not be recognized (Q6: only 3 "disagree"). The most frequent criticism was that some doctors felt constrained by the tool (n=9/24) and were unable to ask all the questions they wanted to (5/24). In this respect, we observed differences between the two scenarios, suggesting that this issue is related to the system coverage or mapping of sentences to core sentences. Finally, all doctors believed that they could integrate such a system in their daily practice (Q7: no "disagree" or "strongly disagree").

Figure 6. Results of the satisfaction questionnaire completed after the dialogue with the female patient. The numbers in circles represent the number of doctors.

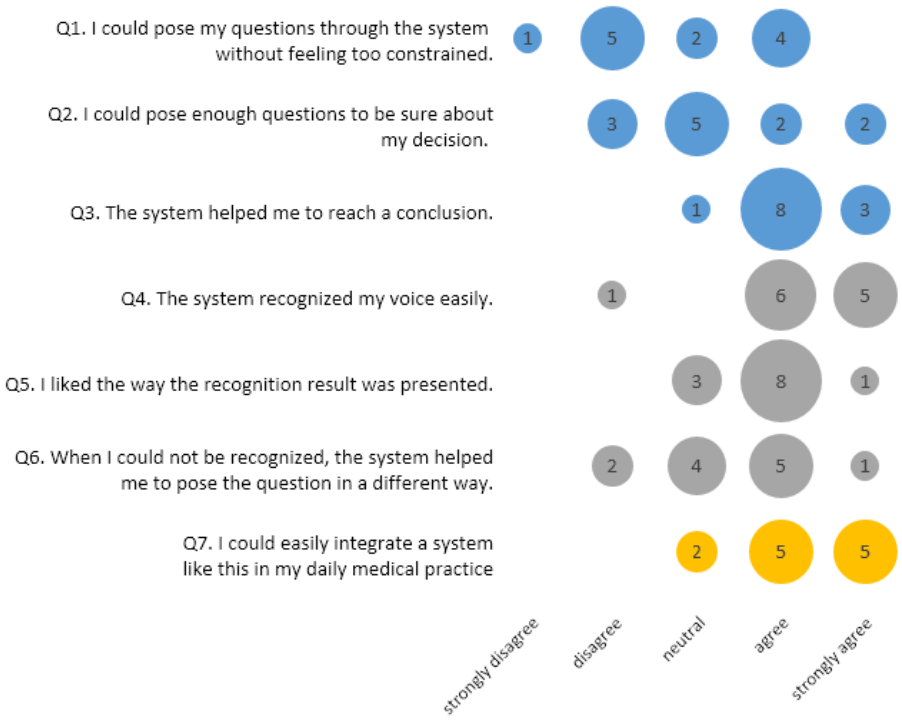
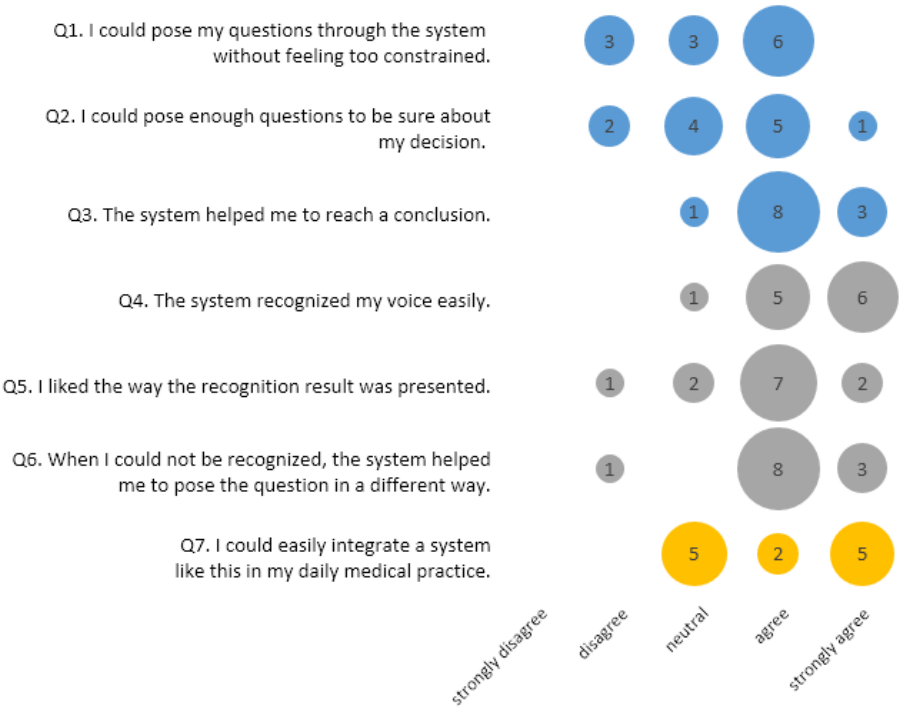


Figure 7. Results of the satisfaction questionnaire completed after the dialogue with the male patient. The numbers in circles represent the number of doctors.



Discussion

Principal Results

All participants were able to pose their questions to the patients and reach the correct diagnosis based on the information collected using BabelDr. However, although they believed that the system helped them to reach a conclusion, some felt constrained by the tool, as they could not ask enough questions to reach a diagnosis. Speech was the preferred modality, even if all doctors translated items from the text list, thus showing that both modalities are useful. The use of text was statistically

influenced by the percentage of successful speech interactions and by the session (first use vs second use). Therefore, speech seems to help in using the system, as participants can express themselves freely and see the most related core sentences.

Comparison With Previous Research

Other studies have analyzed user satisfaction (of both patients and medical staff) [19,20] or the quality of translation with translation systems [4]. However, to our knowledge, this study is the first to measure the impact of the medium on diagnosis. This study confirmed the results of two previous evaluations of BabelDr. A comparison with a traditional fixed-phrase translator (Medibabble) in artificial settings (doctors had to find answers to specific questions) [21] showed that speech improves both usability (reduces time and number of clicks required to ask a question) and satisfaction. Another study [5,22] compared an earlier version of BabelDr with Google Translate at the level of diagnosis, satisfaction, and translation quality in a setting similar to this study. The main result was that BabelDr produced a better translation quality, improved precision (odds ratio: 0.04, 95% CI 0.02-0.12; $P<.001$ in favor of BabelDr) and fluidity (odds ratio: 0.04, 95% CI 0.02-0.10; $P<.001$ in favor of BabelDr) and led to more correct diagnoses than Google Translate.

Limitations

A preliminary version of the tool was used in the study. The system coverage, that is, the questions available to the doctors, is being continually improved based on the collected data. It is possible that the perception of constraint reported by the users was at least partially caused by insufficient coverage for the scenarios selected for this study, rather than by the system itself.

For the cystitis scenario, doctors would have benefited to have been able to change to another domain (abdominal pain), which was not accessible for this study. In addition, the doctors were informed beforehand of the patient's chief complaint. This matches the usual practice at HUG where this information is collected from patients during admission, but another study without prior information would ascertain whether the subdivision into domains, as done in BabelDr, meets the doctor's requirements.

The two standardized patients had a higher education level and no difficulty understanding the Arabic translations provided by the system. In the case of less literate patients, misunderstandings might cause incorrect patient responses and thus lead to incorrect diagnoses. Although the BabelDr translations are aimed at simplicity, a study of the translation quality and accessibility is currently in progress to ascertain whether the translations are suited to patients of different ages, education levels, and cultural and geographic origins.

Due to the rehearsed nature of the patient narratives, based on the given lists of symptoms rather than the potentially vague or contradicting observations by a real patient, it can be argued that the system performance in terms of diagnostic success would be lower with real patients. However, we suspect that the system's restriction to yes/no questions might actually improve clarity by enforcing precise questions and unambiguous patient responses.

During this experiment, we observed very few user errors, such as doctors forgetting to shut off the microphone or using questions that could not be answered nonverbally. Anecdotally, we have observed more such errors in real-use cases with real patients. However, it is possible that in the artificial setting of this study, doctors were more attentive to the system than when using it with a real patient, where the focus would be more on the patient, and thus, the proportion of successful interactions might be lower.

The number of dialogues per doctor (n=2) in this study was insufficient to measure a quantifiable learning effect, but a study is currently in progress at HUG, where BabelDr is used in real settings and the collected data will allow us to study its learnability.

Future Research

Our results show that speech and text interaction are complementary in a tool such as BabelDr. Future developments of the system include an improved text-search module providing more flexibility than the current keyword search.

Development of a bidirectional version of the system is ongoing. In this new version, patients will have an interface where they are presented with a range of responses (eg, numeric values, colors, and pictograms). This will allow us to extend the questions available to the doctors by including open questions and will possibly reduce doctors' feelings of being constrained by the system.

Conclusions

This study showed that a phraselator can be an alternative to machine translation and traditional fixed-phrase translators to reliably collect information from the patient in situations where no interpreter is available. Although doctors felt constrained by the system, they were able to confidently reach a diagnosis, and all believed they could use this type of system in everyday medical practice. The relevance of task-based evaluation to assess the usefulness and usability of translation tools for the diagnosis task was also demonstrated and confirms the importance of reliability in this type of oral context. Doctors clearly appreciated the way in which speech recognition results were presented in the form of a back translation to French, which provided the exact meaning of the translation produced for the patient. Future studies with BabelDr have to confirm these conclusions in real-life settings and investigate the proportion of cases that can be reliably diagnosed with such a tool.

Acknowledgments

The study was funded by the Fondation privée des Hôpitaux Universitaires de Genève. We thank Nuance Communication, Inc, for making their software available free of charge for research purposes. We also thank Jessica Rochat, Rosemary Sudan, and Emmanuel Rayner for their contributions to this study.

Conflicts of Interest

None declared.

References

1. Flores G, Laws MB, Mayo SJ, Zuckerman B, Abreu M, Medina L, et al. Errors in medical interpretation and their potential clinical consequences in pediatric encounters. Pediatrics 2003 Jan;111(1):6-14. [Medline: 12509547]
2. Rechel B, Mladovsky P, Devillé W, Rijks B, Petrova-Benedict R, McKee M. Migration And Health In The European Union (european Observatory On Health Systems And Policies). Maidenhead, Berkshire, England: Open University Press; 2019.
3. Wu AC, Leventhal JM, Ortiz J, Gonzalez EE, Forsyth B. The interpreter as cultural educator of residents: improving communication for Latino parents. Arch Pediatr Adolesc Med 2006 Nov;160(11):1145-1150. [doi: 10.1001/archpedi.160.11.1145] [Medline: 17088518]
4. Patil S, Davies P. Use of Google Translate in medical communication: evaluation of accuracy. BMJ 2014 Dec 15;349:g7392 [FREE Full text] [doi: 10.1136/bmj.g7392] [Medline: 25512386]
5. Bouillon P, Gerlach J, Spechbach H, Tsourakis N, Halimi S. BabelDr vs Google Translate: a user study at Geneva University Hospitals (HUG). In: Proceedings of the 20th Annual Conference of the European Association for Machine Translation. 2017 Presented at: 20th Annual Conference of the European Association for Machine Translation; 29-31 May 2017; Prague, Czech Republic p. 747-752.
6. Dew K, Turner A, Choi Y, Bosold A, Kirchhoff K. Development of machine translation technology for assisting health communication: A systematic review. J Biomed Inform 2018 Sep;85:56-67. [doi: 10.1016/j.jbi.2018.07.018] [Medline: 30031857]
7. BabelDr: Spoken Language Translation of Dialogues in the Medical Domain. URL: https://babeldr.unige.ch/ [accessed 2019-04-23] [WebCite Cache ID 74jFPa3QF]
8. Bouillon P, Flores G, Starlander M, Chatzihrisafis N, Santaholma M, Tsourakis N, et al. A bidirectional grammar-based medical speech translator. In: Proceedings of the Workshop on Grammar-Based Approaches to Spoken Language Processing. 2007 Presented at: ACL 2007; June 2007; Prague, Czech Republic p. 41-48.
9. Wikipedia. Phraselator URL: https://en.wikipedia.org/wiki/Phraselator [accessed 2019-04-23] [WebCite Cache ID 76RzhshCe]
10. Ehsani F, Kimzey J, Zuber E, Master D, Sudre K. Speech to speech translation for nurse patient interaction. In: Proceedings of the workshop on Speech Processing for Safety Critical Translation and Pervasive Applications. 2008 Presented at: 22nd International Conference on Computational Linguistics (Coling 2008); August 2008; Manchester, UK p. 54-59.
11. Medibabble. URL: http://medibabble.com/ [accessed 2019-04-23] [WebCite Cache ID 74jFaAVTG]
12. Universal Doctor. URL: http://www.universaldoctor.com/ [accessed 2019-04-23] [WebCite Cache ID 74jFhoKAS]
13. Rayner M, Armando A, Bouillon P, Ebling S, Gerlach J, Halimi S, et al. Helping domain experts build phrasal speech translation systems. In: Quesada JF, Martín Mateos FJ, Lopez-Soto T, editors. Future and Emergent Trends in Language Technology. First International Workshop, FETLT 2015, Seville, Spain, November 19-20, 2015, Revised Selected Papers. Cham, Switzerland: Springer International Publishing; 2016.
14. Gerlach J, Spechbach H, Bouillon P. Creating an online translation platform to build target language resources for a medical phraselator. In: Proceedings of the 40th Edition of the Translating the Computer Conference. 2018 Presented at: 40th Edition of the Translatingthe Computer Conference (TC 40); November 2018; London, UK p. 60-65.
15. Rayner M, Tsourakis N, Gerlach J. Lightweight spoken utterance classification with CFG, tf-idf and dynamic programming. In: Statistical Language and Speech Processing. Cham, Switzerland: Springer International Publishing; 2017:143-154.
16. Seligman M, Dillinger M. Automatic speech translation for healthcare: some internet and interface aspects. In: Proceedings of the 10th International Conference on Terminology and Artificial Intelligence (TIA-13). 2013 Presented at: 10th International Conference on Terminology and Artificial Intelligence (TIA-13); October 28-30; Paris, France.
17. Gao Y, Gu L, Zhou B, Sarikaya R, Afify M, Kuo K, et al. IBM MASTOR SYSTEM: Multilingual automatic speech-to-speech translator. In: Proceedings of the First International Workshop on Medical Speech Translation. 2006 Presented at: NAACL/HLT; June 4-9; New York, NY, USA.
18. Brooke J. SUS: a quickdirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability evaluation in industry. London: Taylor and Francis; 1996:189-194.
19. Albrecht U, Behrends M, Schmeer R, Matthies H, von Jan U. Usage of multilingual mobile translation applications in clinical settings. JMIR Mhealth Uhealth 2013 Apr 23;1(1):e4 [FREE Full text] [doi: 10.2196/mhealth.2268] [Medline: 25100677]
20. Taicher B, Alam R, Berman J, Epstein R. Design, implementation, and evaluation of a computerized system to communicate with patients with limited native language proficiency in the perioperative period. Anesth Analg 2011 Jan;112(1):106-112. [doi: 10.1213/ANE.0b013e3182009cd1] [Medline: 21081767]
21. Boujon V, Bouillon P, Spechbach H, Gerlach J, Strasly I. Can speech-enabled phraselators improve healthcare accessibility? A case study comparing BabelDr with MediBabble for anamnesis in emergency settings. In: Proceedings of the 1st Swiss Conference on Barrier-free Communication. 2017 Presented at: 1st Swiss Conference on Barrier-free Communication; September 2017; Winterthur, Switzerland p. 32-38. [doi: 10.21256/zhaw-3000]

22. Spechbach H, Halimi S, Gerlach J, Tsourakis N, Bouillon P. Comparison of the quality of two speech translators in emergency settings: A case study with standardized Arabic speaking patients with abdominal pain. In: Proceedings of the 11th European Congress of Emergency Medicine. 2017 Presented at: 11th European Congress of Emergency Medicine; September; Athens, Greece.

Abbreviations

HUG: Geneva University Hospitals

Edited by C Lovis; submitted 17.12.18; peer-reviewed by M Seligman, P Boisrond, K Day, A Cox, J Liang; comments to author 12.01.19; revised version received 29.03.19; accepted 15.04.19; published 07.05.19.

Please cite as:
Spechbach H, Gerlach J, Mazouri Karker S, Tsourakis N, Combescure C, Bouillon P
A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study
JMIR Med Inform 2019;7(2):e13167
URL: http://medinform.jmir.org/2019/2/e13167/
doi:10.2196/13167
PMID:31066702

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org