JMIR Medical Informatics

Clinical informatics, decision support for health professionals, electronic health records, and eHealth infrastructures
Volume 6 (2018), Issue 4 ISSN: 2291-9694

Contents

Original Papers

Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study (e11949)
David Maslove, Jacob Klein, Kathryn Brohman, Patrick Martin. . . . 4

Most Influential Qualities in Creating Satisfaction Among the Users of Health Information Systems: Study in Seven European Union Countries (e11252)
Shahryar Eivazzadeh, Johan Berglund, Tobias Larsson, Markus Fiedler, Peter Anderberg. . . . 11

Novel Approach to Cluster Patient-Generated Data Into Actionable Topics: Case Study of a Web-Based Breast Cancer Forum (e45)
Josette Jones, Meeta Pradhan, Masoud Hosseini, Anand Kulanthaivel, Mahmood Hosseini. . . . 30

Extraction of Information Related to Adverse Drug Events from Electronic Health Record Notes: Design of an End-to-End Model Based on Deep Learning (e12159)
Fei Li, Weisong Liu, Hong Yu. . . . 40

Predictive Modeling of 30-Day Emergency Hospital Transport of Patients Using a Personal Emergency Response System: Prognostic Retrospective Study (e49)
Jorn op den Buijs, Mariana Simons, Sara Golas, Nils Fischer, Jennifer Felsted, Linda Schertzer, Stephen Agboola, Joseph Kvedar, Kamal Jethwani. . . . 54

Utilization of Electronic Medical Records and Biomedical Literature to Support the Diagnosis of Rare Diseases Using Data Fusion and Collaborative Filtering Approaches (e11301)
Feichen Shen, Sijia Liu, Yanshan Wang, Andrew Wen, Liwei Wang, Hongfang Liu. . . . 68

Predicting Current Glycated Hemoglobin Values in Adults: Development of an Algorithm From the Electronic Health Record (e10780)
Brian Wells, Kristin Lenoir, Jose-Franck Diaz-Garelli, Wendell Futrell, Elizabeth Lockerman, Kevin Pantalone, Michael Kattan. . . . 84

Assessing the Impacts of Integrated Decision Support Software on Sexual Orientation Recording, Comprehensive Sexual Health Testing, and Detection of Infections Among Gay and Bisexual Men Attending General Practice: Observational Study (e10808)
Denton Callander, Christopher Bourne, Handan Wand, Mark Stoové, Jane Hocking, John de Wit, John Kaldor, Basil Donovan, Catherine Pell, Robert Finlayson, David Baker, Bradley Forssman, BK Tee, Bill Kefalas, Tim Duck, Rebecca Guy. . . . 95

Impact of Electronic Versus Paper Vital Sign Observations on Length of Stay in Trauma Patients: Stepped-Wedge, Cluster Randomized Controlled Trial (e10221)
David Wong, Julia Knight, Jacqueline Birks, Lionel Tarassenko, Peter Watkinson. . . . 105

XSL•FO
RenderX

Processing of Electronic Medical Records for Health Services Research in an Academic Medical Center: Methods and Validation (e10933)
Nabilah Rahman, Debby Wang, Sheryl Ng, Sravan Ramachandran, Srinath Sridharan, Astrid Khoo, Chuen Tan, Wei-Ping Goh, Xin Tan. . . . 1 1 6

Development of an eHealth System to Capture and Analyze Patient Sensor and Self-Report Data: Mixed-Methods Assessment of Potential Applications to Improve Cancer Care Delivery (e46)
Alexander Lucas, Michael Bass, Nan Rothrock, Mary O'Connor, Mia Sorkin, Jason Nawyn, Fahd Albinali, Lynne Wagner. 136

Identifying Principles for the Construction of an Ontology-Based Knowledge Base: A Case Study Approach (e52)
Xia Jing, Nicholas Hardiker, Stephen Kay, Yongsheng Gao. 149

Identifying Patients Who Are Likely to Receive Most of Their Care From a Specific Health Care System: Demonstration via Secondary Analysis (e12241)
Gang Luo, Peter Tarczy-Hornoch, Adam Wilcox, E Lee. 164

Benefits and Costs of Digital Consulting in Clinics Serving Young People With Long-Term Conditions: Mixed-Methods Approach (e48)
Sung Kim, Jason Madan, Melina Dritsaki, Carol Bryce, Vera Forjaz, Joe Fraser, Frances Griffiths, Kathryn Hamilton, Caroline Huxley, Jackie Sturt. 176

Using an Internet-Based Hospital to Address Maldistribution of Health Care Resources in Rural Areas of Guangdong Province, China: Retrospective and Descriptive Study (e51)
Cui He, Qiru Zhou, Wei Chen, Junzhang Tian, Lihua Zhou, Hong Peng, Shuo Luan, Shengfei Wang. 187

A Computerized Method for Measuring Computed Tomography Pulmonary Angiography Yield in the Emergency Department: Validation Study (e44)
Safiya Richardson, Philip Solomon, Alexander O'Connell, Sundas Khan, Jonathan Gong, Alex Makhnevich, Guang Qiu, Meng Zhang, Thomas McGinn. 202

Appropriateness of Hospital Admission for Emergency Department Patients with Bronchiolitis: Secondary Analysis (e10498)
Gang Luo, Michael Johnson, Flory Nkoy, Shan He, Bryan Stone. 208

Health Data for Research Through a Nationwide Privacy-Proof System in Belgium: Design and Implementation (e11428)
Nicolas Delvaux, Bert Aertgeerts, Johan van Bussel, Geert Goderis, Bert Vaes, Mieke Vermandere. 218

Clinical Named Entity Recognition From Chinese Electronic Health Records via Machine Learning Methods (e50)
Yu Zhang, Xuwen Wang, Zhen Hou, Jiao Li. 241

Effect of a Multimedia Patient Decision Aid to Supplement the Informed Consent Process of a Peripherally Inserted Central Venous Catheter Procedure: Pre-Post Quasi-Experimental Study (e11056)
Azizeh Sowan, Arlienita Beraya, Adrian Carrola, Charles Reed. 255

Validation and Testing of Fast Healthcare Interoperability Resources Standards Compliance: Data Analysis (e10870)
Jason Walonoski, Robert Scanlon, Conor Dowling, Mario Hyland, Richard Ettema, Steven Posnack. 264

XSL•FO
RenderX

Proposal

Contextual Anonymization for Secondary Use of Big Data in Biomedical Research: Proposal for an Anonymization Matrix (e47)
John Rumbold, Barbara Pierscionek. 228

XSL•FO
RenderX

Original Paper

Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study

David M Maslove[1], MD, MS; Jacob Klein[2]; Kathryn Brohman[3], PhD; Patrick Martin[2], PhD

[1]Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada

[2]School of Computing, Queen's University, Kingston, ON, Canada

[3]Smith School of Business, Queen's University, Kingston, ON, Canada

Corresponding Author:
David M Maslove, MD, MS
Department of Critical Care Medicine
Queen's University
Kingston General Hospital
76 Stuart St.
Kingston, ON, K7L 2V7
Canada
Phone: 1 613 650 7311
Email: david.maslove@queensu.ca

Abstract

Background: Blockchain technology is emerging as an innovative tool in data and software security.

Objective: This study aims to explore the role of blockchain in supporting clinical trials data management and develop a proof-of-concept implementation of a patient-facing and researcher-facing system.

Methods: Blockchain-based Smart Contracts were built using the Ethereum platform.

Results: We described BlockTrial, a system that uses a Web-based interface to allow users to run trials-related Smart Contracts on an Ethereum network. Functions allow patients to grant researchers access to their data and allow researchers to submit queries for data that are stored off chain. As a type of distributed ledger, the system generates a durable and transparent log of these and other transactions. BlockTrial could be used to increase the trustworthiness of data collected during clinical research with benefits to researchers, regulators, and drug companies alike. In addition, the system could empower patients to become more active and fully informed partners in research.

Conclusions: Blockchain technology presents an opportunity to address some of the common threats to the integrity of data collected in clinical trials and ensure that the analysis of these data comply with prespecified plans. Further technical work is needed to add additional functions. Policies must be developed to determine the optimal models for participation in the system by its various stakeholders.

(JMIR Med Inform 2018;6(4):e11949) doi:10.2196/11949

KEYWORDS

blockchain; clinical trial; informatics; data accuracy; data collection

Introduction

Clinical trials generate data used in the preparation of peer-reviewed journal papers and applications to regulatory bodies for approval of new treatments. In both cases, the integrity of these data is important to numerous stakeholders, including academic researchers, journal editors and publishers, drug and device companies, government regulators, and most importantly, prospective patients and the general public. For researchers, published papers serve as a key external validator of the rigor of their work and represent important academic achievements. For companies, considerable investment in research and development may be at stake along with the prospect of future earnings, both of which typically count in the billions of dollars. In brief, both publishers and regulators need to trust the data presented to them.

Several threats to the validity of clinical trials data stand to undermine this trust [1]. First, data can be altered or lost, either accidentally or by nefarious intent. Though redundancy exists in many database systems, these are often opaque to outside

XSL•FO
RenderX

observers. Second, there is a risk that the published analysis is not a true representation of the analysis that was initially planned [2]. Reasons for variation include failing to report all outcomes measured, selective reporting of positive results (publication bias), stopping data collection after achieving the desired result, and excluding some data after assessing the impact of doing so, among many others. Third, data may be fabricated, manipulated, or duplicated by researchers committing outright fraud. Because of these risks, journals, regulators, and other oversight groups are expected to trust the data generated by clinical trials in the absence of a fully trustworthy relationship with those who generated them.

Blockchain technology has recently emerged as an alternative means for transferring data between participating parties based on a "distributed ledger" model that affords a fully transparent and immutable record of data transactions [3]. A blockchain consists of consecutive chained blocks that are replicated on the nodes of a peer-to-peer network. Blocks consist of records, and each record contains details of a transaction between the users of the system. Blockchain technology was originally designed to serve as the basis for electronic cash systems such as bitcoin [4]; it eliminates the need for trusted third parties in financial transactions by providing a secure and verifiable history for every transaction in the system. Depending on the application, transactions may involve a cryptocurrency, such as bitcoin, or other kinds of assets.

Smart contracts—code that is stored, executed, and verified on a blockchain—are a central component of the next-generation blockchain platforms [5]. Smart contracts can play several roles, including encoding the business logic for an application, ensuring that preconditions for action are met before it is executed, and enforcing permissions for an action. Because smart contracts run on a blockchain, they have unique characteristics compared with other types of software. First, the program itself is recorded on the blockchain, which imparts the blockchain's characteristic permanence and resistance to censorship. Second, the program can control blockchain assets. Third, the program is executed by the blockchain, meaning it will always execute as written and no one can interfere with its operation.

As blockchain technology matures, applications outside of finance are increasingly being explored, including in the health care sector [6-10]. Blockchain-based models for electronic medical records have been proposed that would empower patients to exercise greater ownership of their medical data and enhance data sharing between platforms [11-13]. Blockchain technology might also prove useful in supporting or even supplanting the traditional data infrastructure used in clinical trials [14-16]. Because the blockchain can be used to establish a permanent record agreed upon by all participating parties, it has the potential to mitigate some of the threats to data validity as outlined above. Immutable clinical trials data recorded using a blockchain may inspire greater confidence in its veracity, resulting in better science, safer medicines, and enhanced public trust in biomedical research.

This study aims to describe BlockTrial, a blockchain system for clinical trials management based on the Ethereum platform. We discuss some design considerations and describe a proof-of-concept system for patient enrollment and data retrieval.

Methods

The clinical trials process (Figure 1) consists of trial protocol setup and registration, patient enrollment, data collection, data analysis, report generation, and publication of results [14]. Blockchain solutions may be useful in managing study metadata, including protocol registration, descriptions of prespecified analyses, screening and enrollment logs, and data upload and query logs. The first steps involve the creation of metadata, whereas the later reporting stages refer to these metadata to ensure adherence to protocols and prespecified analyses. Blockchain-based solutions stand to address questions of data integrity and reproducibility at these various stages through the creation of a detailed, time-stamped ledger of data upload and query events.

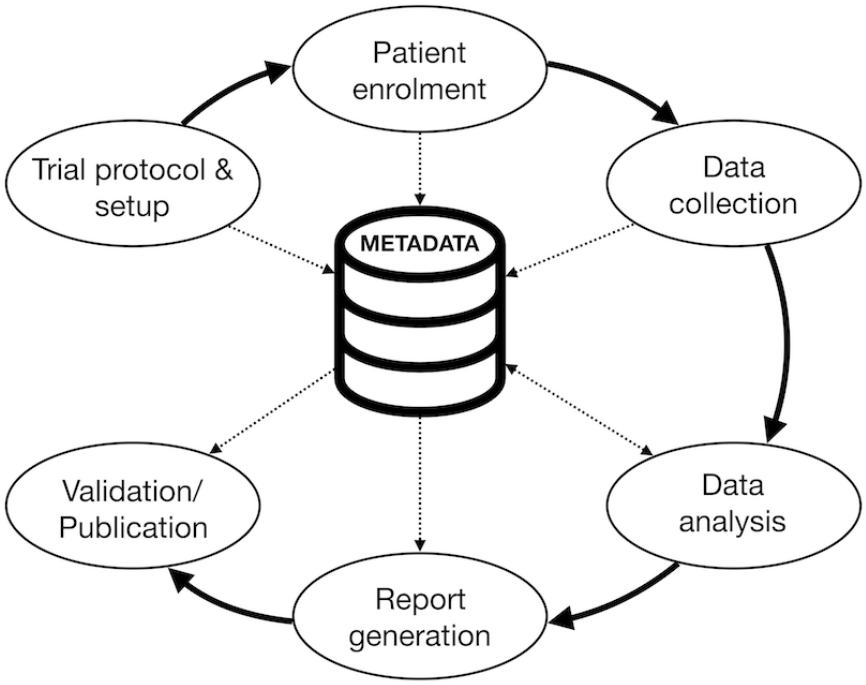
Our development process addressed 3 main design considerations. The first relates to the type of blockchain that would best suit clinical trials applications. Blockchains can be public, consortium-based, or private. Public blockchains are those that anyone in the world can read and send transactions to, conferring maximal transparency. In this configuration, the nodes extend beyond the research community and could include anyone in the world participating in the public network. Blockchains controlled by a single organization (private) or group of organizations (consortium-based) can have restricted permissions for reading and writing and are therefore only partly decentralized. A consortium-based blockchain might be run by a group of oversight bodies (eg, academic journals and licensing authorities) with nodes distributed across members of the consortium, which in this model, might include investigators active in clinical research; this could, in theory, come at the cost of diminished visibility by outside organizations. Second is the question of who would participate in a blockchain-based clinical trials platform. Several different groups have an important stake in the trials process, including patients, researchers, ethics boards, funders, pharma and device companies, regulatory agencies, publishers, and others. Each of these stakeholders has various interests in the process with differing assets at stake and resources to commit. One advantage of using a blockchain to manage clinical trials is that it stands to directly empower these various actors, who may have substantial investment in the study, be it financial, academic, or personal. The potential to involve patients in the management of their own data is a particularly useful prospect aligned with the ethical principle of autonomy that is central to clinical research. The third is the question of whether the blockchain should store the research data or whether transaction records alone should be stored. For most studies, transaction records would be much smaller in size and could be used to certify the veracity of the data, provide means of determining provenance, and point to clinical and research data stored elsewhere. Moreover, storing transaction records alone on the blockchain allows the platform to exploit existing approaches to managing clinical research data, ensuring compatibility with current clinical trials data management tools such as OpenClinica or REDCap [17]. Although the study data

XSL•FO
RenderX

could, in theory, be stored directly on the blockchain, this approach may be impractical because of the limitations of storage space and difficulties in querying the data in such a configuration, especially when considering large datasets such as those generated by genome sequencing.

We developed BlockTrial using a private Ethereum blockchain, allowing us to have full control of the design and development process and leverage the smart contract functionality that is central to our model. Given the abovementioned trade-offs regarding storage space and ease of querying, we developed a model in which the blockchain stores transaction records rather than the clinical trials data itself. Our proof-of-concept platform models the actions of patients who may give or withhold consent to participate in the study, as well as researchers, who collect, store, and analyze the data.

Figure 1. Clinical trials workflow: various phases of the clinical trials process generate and make use of study metadata.



Results

Ethereum is an open-source programmable blockchain that supports smart contracts and allows developers to create applications at any scale [5]; it includes a peer-to-peer network protocol with each node in the network running the Ethereum Virtual Machine. Operations can be distributed across the network, thereby leveraging the security, accountability, and efficiency of distributed computing. The Ethereum Foundation makes available several Ethereum client implementations, including C++, Python, Go, Java, JavaScript, and Ruby.

Figure 2 shows the structure of BlockTrial. The patient Smart Contract (Patient SC) governs patient enrollment and granting of permissions, whereas the research Smart Contract (Research SC) allows researchers to send and receive queries to the trial database. A blockchain "oracle" is used to interface with "off-chain" resources, including the trial database.

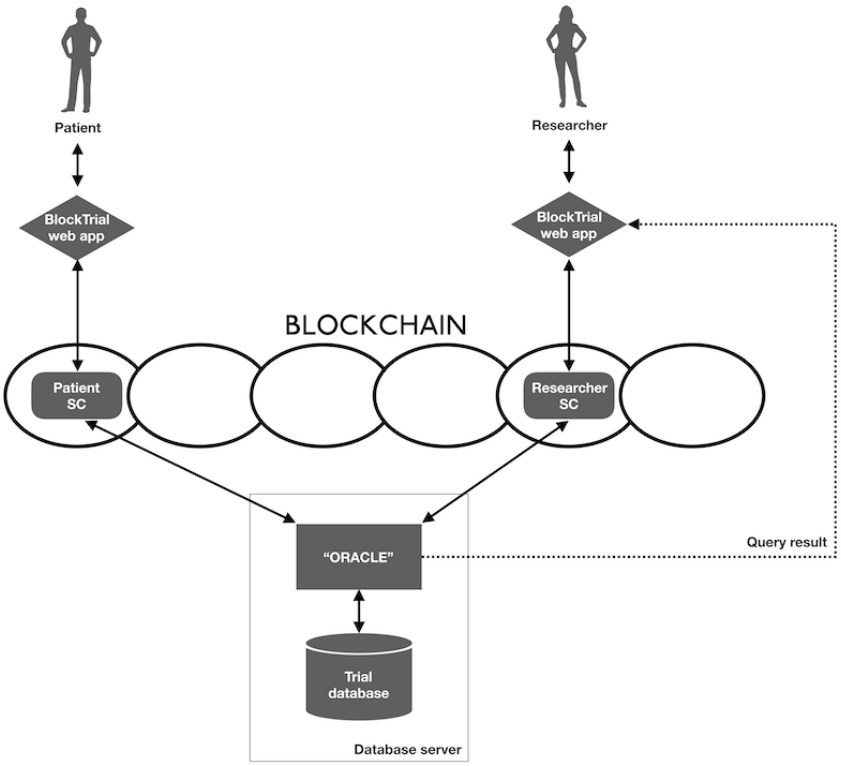
The implementation, which is intended to demonstrate the feasibility of our blockchain-based approach, supports a simplified clinical trial scenario in which patients and researchers interact with one another as well as with the system through a Web app (BlockTrial App). Patient data have been collected and stored in a database that is managed by a *Database Server* outside of the blockchain. Following appropriate discussions with research personnel and upon providing informed consent to participate in a study, patients use the BlockTrial App to register in the study and set permissions on their data for the study. Their registration and permission settings are recorded in the blockchain. Researchers use the BlockTrial App to retrieve study data from the database. Data requests are filtered by patients' permissions and registered in the blockchain. The requests are in turn retrieved from the blockchain and executed by the database server which sends the results of the requests to the researcher. The data analysis can then be performed outside the system at the researcher's site.

XSL•FO
RenderX

Smart contracts are placed on the Ethereum blockchain to implement the actions for the patients and researchers. The contracts create transactions to record the actions in the blockchain, thus ensuring the integrity of the study and facilitate requests to the external database server managing the trial database. Furthermore, actions initiated by users through the BlockTrial App are implemented by calling methods made available by smart contracts.

Although smart contracts can run algorithmic calculations, store data, and retrieve data, it is not practical to make arbitrary requests to a service off the blockchain, such as the database server, from a smart contract. Smart contracts interact with the off-chain world through "oracles," which are agents that watch the blockchain for events and respond to them by invoking a service or performing some action [16]. The database server includes a blockchain oracle that looks for new queries on the blockchain, executes the queries on the database, and publishes the results of the query back to the contract. Textbox 1 shows the methods provided by BlockTrial's smart contracts.

Figure 2. The BlockTrial structure—patients and researchers interact with BlockTrial through the Web-based BlockTrial App. SC: Smart Contract.



Textbox 1. Smart contract methods.

Patient Smart Contract

- addPatient: Registers a patient for the study and sets access permissions
- editPermissions: Changes a patient's access permissions
- getPeople: Gets the set of patients registered for the study

Researcher Smart Contract

- addQuery: Submits a new query to the blockchain
- addQueryResult: Places a hash of a query result on the blockchain
- getQueries: Retrieves waiting queries from the blockchain
- getUnsolvedCount: Retrieves the number of waiting queries from the blockchain

XSL•FO
RenderX

The patient smart contract (*Patient SC*) maintains an array in the blockchain that holds patients registered for the study and the permissions they have set for their data. Patients register their permission to access data for the study with the *addPatient* method and change their permissions with the *editPermissions* method.

The researcher smart contract (*Researcher SC*) supports researchers issuing queries to the study database and receiving their results. To retrieve study data, a researcher calls the *addQuery* method, which places the SQL text of the query and the address of the researcher issuing the query in a structure on the blockchain maintained by the *Researcher SC*. The researcher then waits for a notification from the database server that the result is available. When the oracle in the database server recognizes that a new query is available in the *Researcher SC*, it uses the *getQueries* method to fetch the query text and researcher address and the *getPeople* method of the *Patient SC* to get the current permissions for all patients registered in the study. The oracle applies the permissions to the query and submits the query to the database. Because query results can be arbitrarily large, it is not practical to put the results on the blockchain. Therefore, we apply a secure hash function, such as SHA-1, to obtain a unique signature that can be used for later validation and append a new transaction to the blockchain containing the query and the result signature. The oracle then makes the result of the query available to the researcher through a separate channel.

Discussion

Principal Findings

Clinical trials are one of the cornerstones of biomedical research, providing a mechanism for the objective evaluation of the benefits and risks of new therapies. Trust in clinical trials reporting is synonymous with that in the medical profession itself. Much is at stake, given the potential consequences for patients and their doctors as well as major stakeholders, including researchers, pharmaceutical and device manufacturers, regulators, publishers, and funding agencies. The trustworthiness of clinical trials data is central to modern medical research and practice but for a variety of reasons, it cannot be assumed.

A blockchain platform for clinical trials could be useful in ensuring the trustworthiness of clinical trials data collection and reporting. The need for such a platform reflects the convergence of several trends in clinical research. First, patient autonomy is increasingly recognized as necessary to the conduct of ethical studies and to enhance recruitment into trials [18]. Blockchain technology has already been explored as a means of enhancing agency among patients in the form of electronic medical records that empower patients to share their data with their care providers. A natural extension of this use case is in ongoing consent to participate in clinical trials and to share data with researchers.

Second, regulators such as the Food and Drug Administration often require that data collected in the course of a clinical trial investigating a new drug or device be posted in a registry, such as clinicaltrials.gov, upon study completion [19]. However, adherence to this practice remains poor [20,21]. Blockchain technology could improve adherence to reporting through various forcing functions, particularly at the time that data are queried by investigators for further analysis. Querying the dataset could trigger automatic posting of the data to a registry, perhaps with an embargo period to allow the researchers to complete their initial analysis. In addition, blockchain technology might be useful in ensuring that analysis is done in accordance with the statistical analysis plans that were specified at the time of trial registration. This need reflects a third trend, highlighting the potential negative consequences of multiple post hoc analyses (also referred to as “data dredging”) [22,23].

Finally, blockchain technology suggests a mechanism through which researchers can interact with research oversight bodies, including local research ethics boards. By uploading ethics applications to a blockchain, permanence of the protocol and approval process can be established. Furthermore, a blockchain might facilitate approval by a single research ethics board across multiple sites participating in a multicenter study.

This study described BlockTrial, a proof-of-concept implementation of a blockchain for clinical trials research. BlockTrial operates as a private blockchain, where patients and researchers act as nodes. Patients can assign permissions for which of their data can be viewed, and by whom, whereas researcher requests for data are written into an immutable record. Previous work has explored the use of blockchain technology to enroll patients in clinical trials and record their consent to participate [15,16]. This study extends this model to include research participation in a blockchain with data query and retrieval capabilities. This model empowers patients to take a more active role in their participation in research studies. The Web-based framework stands to improve the efficiency of study enrollment and ensure that consent is as informed as possible. Records of assigning and revoking consent and permission would be visible to all parties and could be used for auditing purposes to ensure proper adherence to trial recruitment guidelines. Immutable records of researcher requests for data and of delivery of said data could be tied to required reporting tasks to ensure that trial registries, such as clinicaltrials.gov, are populated timely with up-to-date and complete data.

This study has several strengths. It uses the existing Ethereum blockchain, which is widely used and which offers several application programming interfaces to facilitate further software development. We model 2 different types of participants, including both patients and researchers as nodes in the network. In addition, we developed a Web-based interface for ease of use. However, this study also has several limitations that must be considered. Chief among these is the cost of participating in an Ethereum-based blockchain; this would require that some participants either purchase ether or mine ether directly (and therefore incur computing and power costs). The optimal distribution of costs remains to be worked out; however, we believe that significant incentives exist for researchers and companies to contribute to these costs, especially if regulations around reporting requirements continue to expand. In addition, our system requires integration with an existing clinical database such as REDCap. Although this does increase the overhead in terms of computing infrastructure and middleware, it is likely

XSL•FO
RenderX

to improve the ease and speed of querying data and allows for the use of large data stores, which are unlikely to be easily accommodated on the blockchain itself. The patient-facing functionalities of our model are designed to increase patient autonomy through the ability to assign and revoke permissions for researchers to access data. However, this does introduce the potential that more patients will opt out of a study in progress, potentially introducing considerable selection bias. Finally, the blockchain implementation we describe does not guarantee absolute data integrity; data entry itself remains a point of vulnerability. Although it remains unknown how this model of clinical trial data management will be received by various stakeholders, we believe that substantial benefits may accrue to users of all types and that these would offset the upfront costs of further developing and subscribing to a new system.

Future research and development should address the sociotechnical barriers to implementing blockchain solutions for clinical trials management. Additional participant roles should be modeled, including funders, ethics boards, research institutions, academic journals, trial registries, pharmaceutical corporations, clinical trial organizations, and regulatory agencies. In addition, payment models should be explored, which may include consideration of a specific "BlockTrial coin" cryptocurrency that could be used to run the system and create the necessary incentives to induce the desired participant engagement. Integration with clinical trials registries and journal publishers should be explored to increase data sharing and transparency and facilitate the process of publishing trial results without bias, including selection bias, confirmation bias, and various types of confounding, all of which could be introduced in the course of post hoc analyses. Finally, qualitative research is needed to better describe the attitudes and preferences of patients, researchers, and other stakeholders toward blockchain-enabled clinical trials data infrastructure.

Conclusions

This study presents a proof-of-concept blockchain-enabled clinical trials data management solution that enables the interaction of patients and researchers engaged in clinical research. BlockTrial affords immediate benefits to patients by empowering them to better control access to their data and to researchers by affording them useful tools to maintain adherence to reporting requirements. Further developed with more stakeholder roles, BlockTrial stands to enhance the integrity of clinical trials data and promote trust throughout the clinical research community and beyond in the output of medical research.

Acknowledgments

DMM received funding as part of the Southeastern Ontario Academic Medical Organization New Clinician Scientist program.

Conflicts of Interest

None declared.

References

1. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci 2012 May 01;23(5):524-532. [doi: 10.1177/0956797611430953] [Medline: 22508865]
2. Hartung DM, Zarin DA, Guise J, McDonagh M, Paynter R, Helfand M. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. Ann Intern Med 2014 Apr 01;160(7):477-483 [FREE Full text] [doi: 10.7326/M13-0480] [Medline: 24687070]
3. Lemieux V. Trusting records: is Blockchain technology the answer? Records Management Journal 2016 Jul 18;26(2):110-139. [doi: 10.1108/RMJ-12-2015-0042]
4. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System URL: https://bitcoin.org/bitcoin.pdf [accessed 2018-09-24] [WebCite Cache ID 72fh0AX6C]
5. Etherum Foundation. A Next-Generation Smart Contract and Decentralized Application Platform URL: http://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf [accessed 2018-09-24] [WebCite Cache ID 72fi46muG]
6. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220. [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
7. Roman-Belmonte JM, De LCH, Rodriguez-Merchan EC. How blockchain technology can change medicine. Postgrad Med 2018 May;130(4):420-427. [doi: 10.1080/00325481.2018.1472996] [Medline: 29727247]
8. Gammon K. Experimenting with blockchain: Can one technology boost both data integrity and patients' pocketbooks? Nat Med 2018 Apr 10;24(4):378-381. [doi: 10.1038/nm0418-378] [Medline: 29634691]
9. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain Technology for Detecting Falsified and Substandard Drugs in Distribution: Pharmaceutical Supply Chain Intervention. JMIR Res Protoc 2018 Sep 13;7(9):e10163 [FREE Full text] [doi: 10.2196/10163] [Medline: 30213780]
10. Ichikawa D, Kashiyama M, Ueno T. Tamper-Resistant Mobile Health Using Blockchain Technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
11. Mertz L. (Block) Chain Reaction: A Blockchain Revolution Sweeps into Health Care, Offering the Possibility for a Much-Needed Data Solution. IEEE Pulse 2018;9(3):4-7. [doi: 10.1109/MPUL.2018.2814879] [Medline: 29757744]

XSL•FO
RenderX

12. Angraal S, Krumholz HM, Schulz WL. Blockchain Technology: Applications in Health Care. Circ Cardiovasc Qual Outcomes 2017 Sep;10(9). [doi: 10.1161/CIRCOUTCOMES.117.003800] [Medline: 28912202]
13. Roehrs A, da Costa CA, da Rosa Righi R. OmniPHR: A distributed architecture model to integrate personal health records. J Biomed Inform 2017 Dec;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]
14. Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. Trials 2017 Dec 19;18(1):335 [FREE Full text] [doi: 10.1186/s13063-017-2035-z] [Medline: 28724395]
15. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: Transparency and traceability of consent. F1000Res 2017;6:66 [FREE Full text] [doi: 10.12688/f1000research.10531.5] [Medline: 29167732]
16. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. F1000Res 2016;5:2541 [FREE Full text] [doi: 10.12688/f1000research.9756.1] [Medline: 28357041]
17. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
18. Committee on Strategies for Responsible Sharing of Clinical Trial Data; Board on Health Sciences Policy; Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. National Academies Press (US) 2015 Apr. [Medline: 25590113]
19. Kuehn BM. Few studies reporting results at US government clinical trials site. JAMA 2012 Feb 15;307(7):651-653. [doi: 10.1001/jama.2012.127] [Medline: 22337663]
20. Saito H, Gill CJ. How frequently do the results from completed US clinical trials enter the public domain?--A statistical analysis of the ClinicalTrials.gov database. PLoS One 2014;9(7):e101826 [FREE Full text] [doi: 10.1371/journal.pone.0101826] [Medline: 25025477]
21. Anderson ML, Chiswell K, Peterson ED, Tasneem A, Topping J, Califf RM. Compliance with results reporting at ClinicalTrials.gov. N Engl J Med 2015 Mar 12;372(11):1031-1039 [FREE Full text] [doi: 10.1056/NEJMsa1409364] [Medline: 25760355]
22. Kerr N. HARKing: hypothesizing after the results are known. In: Pers Soc Psychol Rev. HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review 6 ed; 1998:196-217.
23. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLoS Biol 2015 Mar;13(3):e1002106 [FREE Full text] [doi: 10.1371/journal.pbio.1002106] [Medline: 25768323]

Abbreviations

SC: Smart Contract

Edited by M Focsa; submitted 15.08.18; peer-reviewed by JT te Gussinklo, C Kruse; comments to author 21.09.18; revised version received 25.09.18; accepted 26.09.18; published 21.12.18.

Please cite as:
Maslove DM, Klein J, Brohman K, Martin P
Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study
JMIR Med Inform 2018;6(4):e11949
URL: http://medinform.jmir.org/2018/4/e11949/
doi:10.2196/11949
PMID:30578196

©David M Maslove, Jacob Klein, Kathryn Brohman, Patrick Martin. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 21.12.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Most Influential Qualities in Creating Satisfaction Among the Users of Health Information Systems: Study in Seven European Union Countries

Shahryar Eivazzadeh[1], MSc, Lic; Johan S Berglund[1], MD, PhD; Tobias C Larsson[2], PhD; Markus Fiedler[3], PhD; Peter Anderberg[1], PhD

[1]Department of Health Science, Blekinge Institute of Technology, Karlskrona, Sweden

[2]Department of Mechanical Engineering, Blekinge Institute of Technology, Karlskrona, Sweden

[3]Department of Technology and Aesthetics, Blekinge Institute of Technology, Karlskrona, Sweden

Corresponding Author:
Shahryar Eivazzadeh, MSc, Lic
Department of Health Science
Blekinge Institute of Technology
Blekinge Tekniska Högskola
Karlskrona, 371 79
Sweden
Phone: 46 455 38 50 00
Email: shahryar.eivazzadeh@bth.se

Abstract

Background: Several models suggest how the qualities of a product or service influence user satisfaction. Models such as the Customer Satisfaction Index (CSI), Technology Acceptance Model (TAM), and Delone and McLean Information Systems Success demonstrate those relations and have been used in the context of health information systems.

Objective: This study aimed to investigate which qualities foster greater satisfaction among patient and professional users. In addition, we are interested in knowing to what extent improvement in those qualities can explain user satisfaction and whether this makes user satisfaction a proxy indicator of those qualities.

Methods: The Unified eValuation using ONtology (UVON) method was used to construct an ontology of the required qualities for 7 electronic health (eHealth) apps being developed in the Future Internet Social and Technological Alignment Research (FI-STAR) project, a European Union (EU) project in electronic health (eHealth). The eHealth apps were deployed across 7 EU countries. The ontology included and unified the required qualities of those systems together with the aspects suggested by the Model for ASsessment of Telemedicine apps (MAST) evaluation framework. Moreover, 2 similar questionnaires for 87 patient users and 31 health professional users were elicited from the ontology. In the questionnaires, the user was asked if the system has improved the specified qualities and if the user was satisfied with the system. The results were analyzed using Kendall correlation coefficients matrices, incorporating the quality and satisfaction aspects. For the next step, 2 partial least squares structural equation modeling (PLS-SEM) path models were developed using the quality and satisfaction measure variables and the latent construct variables that were suggested by the UVON method.

Results: Most of the quality aspects grouped by the UVON method are highly correlated. Strong correlations in each group suggest that the grouped qualities can be measures that reflect a latent quality construct. The PLS-SEM path analysis for the patients reveals that the effectiveness, safety, and efficiency of treatment provided by the system are the most influential qualities in achieving and predicting user satisfaction. For the professional users, effectiveness and affordability are the most influential. The parameters of the PLS-SEM that are calculated allow for the measurement of a user satisfaction index similar to CSI for similar health information systems.

Conclusions: For both patients and professionals, the effectiveness of systems highly contributes to their satisfaction. Patients care about improvements in safety and efficiency, whereas professionals care about improvements in the affordability of treatments with health information systems. User satisfaction is reflected more in the users' evaluation of system output and fulfillment of expectations but slightly less in how far the system is from ideal. Investigating satisfaction scores can be a simple and fast way to infer if the system has improved the abovementioned qualities in treatment and care.

XSL•FO
RenderX

(JMIR Med Inform 2018;6(4):e11252) doi:10.2196/11252

KEYWORDS

health information systems; telemedicine; evaluation studies as topic; consumer behavior; treatment outcome; safety; efficiency; health care costs; ontology engineering; equation models

Introduction

Background

The normative evaluation of health information systems is articulated through a frequently used set of keywords such as *acceptance* or *adoption* [1,2], *success* [3], and *satisfaction* [4,5]. Each of these keywords reminds us how a health information system inherits traits from its conceptual ancestors, that is, the information system, technology, and product. For an overall evaluation of these systems, one might measure how well these information systems succeed [6,7], how these technologies are accepted by users [8,9], or how the customers of these systems, patients, or professionals are satisfied with these products [10]. Below this layer of top indicators, there exist sets of constructs and relationships that cause success, acceptance, or satisfaction. Researchers have tried to capture and demonstrate through the models how success [6,10], acceptance [9], or satisfaction [11-13] are created by constructs such as perceived quality, perceived expectation, ease of use, and other variables. Some of these models have largely been employed in diverse contexts [14]. There are also models, whether novel or customized from the mainstream, that are specific to a smaller context such as health information systems [15-17].

The Customer Satisfaction Index (CSI) model family places the *satisfaction* construct at the core of their path structures. There, the satisfaction construct is affected by leading indicators such as perceived quality. At the same time, it impacts lagging indicators such as user loyalty. There are at least three versions of CSI widely being used. The original CSI model was introduced in Sweden [11]. The American Customer Satisfaction Index improved the Swedish version [12], and then the European Customer Satisfaction Index (ECSI) enhanced the American version [18]. The ECSI model consists of 9 latent construct variables [18], which are measured by a series of measure or manifest variables. Historically, CSI models have been used at macro levels where the satisfaction of customers at the national level or the level of an enterprise was the matter of concern. The wording of CSI models, such as the *customer* term, and the inclusion of some constructs, such as *loyalty*, suggest that measuring user satisfaction at the micro level, that is, the product level, was not their main target. However, the manifestation of the satisfaction concept in the CSI models through its 3 manifest variables [18] is versatile enough to measure satisfaction at both micro or macro levels with the same wording. CSI models introduce a way of measuring satisfaction scores through adding the weighted scores of those 3 variables, which inspires similar approaches in various disciplines. In comparison with using the CSI models for health information systems, one might consider the patient satisfaction models [19] that share a set of common constructs and relations with the CSI models but do not necessarily embed the same structures or components.

The Delone and McLean Information Systems Success (D&M IS) model, a prevalent model for analyzing the success of information systems, sets out the relationship between user satisfaction and quality dimensions [6,20]. In this model, 3 categories of qualities, information, system, and service contribute to user satisfaction. There are 2 other constructs, *net benefits* and *intention to use*, that are in a bidirectional impact relationship with *user satisfaction* [20]. Several studies have validated the causal relationships between the constructs in the D&M IS model [21]. In addition, there exists a long list of validated measures for each of the constructs [22]. The D&M IS model has broadly been used in the health information system context [23]. Furthermore, it has also been extended and customized to be more specific for this context, such as in the Human, Organization, and Technology Fit model [24], but the extension has been directed more toward a wider perspective of organization and technology. Although some studies have incorporated [25] or prioritized [26] more specific qualities, further investigation is needed to be more specific about the impacting qualities and their degree of importance.

Technology acceptance models, such as Technology Acceptance Model (TAM) as well as the Unified Theory of Acceptance and Use of Technology, which are supported by a great many of studies [8,9], placed the *acceptance* of a system or technology at the core of relationships. These models have been applied in health information system studies, although they have reported the significance of some of relationships differently [8]. The user acceptance of a health information system can be a prelude to or reflection of their satisfaction in using that system, but acceptance is not the same as satisfaction. Some researchers have considered *acceptance*, that is, the behavioral intention to use in TAM, an equivalent for *satisfaction* [27], but the *intention to use* is a different construct from *satisfaction* in a well-studied model such as D&M IS.

Contextualizing TAM by adding variables has been a common practice [8,28,29]. Indeed, contextualizing TAM for health information technology has led to the introduction of some frequently employed variables, such as *fit* [30]. However, there is a shortage of studies applying a systematic approach, such as *belief elicitation* [31], when introducing a new variable [8].

Objectives

For the CSI, D&M IS, and TAM, the set of relationships between their proposed constructs have been already examined in various contexts [8,14,13]. Nevertheless, for a specific context such as health information systems, one might seek to develop new models, probably inspired by those that are well established, to expand a construct into more detailed constructs or find manifest variables more relevant to a case. For example, the constructs that represent the qualities of a system are generalized in those models or their variations as perceived quality [13]; system, service, or information quality [20]; and output quality

XSL•FO
RenderX

[8]. However, none of those models represents a specific quality, such as safety, as a stand-alone construct.

In addition, finding a systematic approach to define manifests for construct variables, as mentioned before, is another direction for extending a model [8]. Many of the evaluation frameworks for health information systems suggest the qualities to evaluate [32], arranged as categories or domains. These frameworks implicitly suggest constructs and the manifests to each construct. Simultaneously, the end users of health information systems are another source for eliciting the qualities and their groupings [33].

In the forthcoming sections, we put forth a list of qualities that create and predict user satisfaction with health information systems. The qualities are embedded within a path model that demonstrates their relationships with user satisfaction. This study's methods and materials are discussed in the Methods section. The qualities elicited from the Future Internet Social and Technological Alignment Research (FI-STAR) project by applying the Unified eValuation using ONtology (UVON) method are reported in the Results section, an exploratory result is demonstrated in the Correlation Patterns section, and an estimated model is presented in the Partial Least Squares Structural Equation Modeling Path Models section. The European Union (EU)–wide empirical data collected through the FI-STAR project, detailed in the Methods section and Multimedia Appendix 1, is used to calculate and validate the model in the Partial Least Squares Structural Equation Modeling Path Models section. The results of exploration and model calculations are discussed in the Correlations section and the Partial Least Squares Structural Equation Modeling Path Models section. Subsequently, based on the model, the relative importance of qualities in creating and predicting user satisfaction is discussed in the Most Influential Qualities section. In the section Satisfaction Index, weightings are suggested for the calculation of a satisfaction index for health information systems. Finally, we examine the limits and extensions to our approach and suggested model in the Extensions and Limitations section.

Methods

Data Collection

The empirical data for this study have been collected from the FI-STAR project, an EU electronic health (eHealth) project with 7 subprojects across the EU [34]. A convenience sampling approach was used for the recruitment of the participants. Each eHealth app was deployed in a hospital or health facility site and the users on the site—patients or health professionals—were asked about their assessment of the impact of that specific solution on treatment. Participation in the trials, and therefore the survey, was voluntary with no mentioned preconditions. There was no constraint on the type of eHealth project being developed provided that they follow the FI-STAR requirements, especially using the FIWARE infrastructure. However, most of the subprojects could be categorized as telehealth apps. A summary of all subprojects can be found in Multimedia Appendix 1.

Data Analysis

We applied the UVON method [33] to the FI-STAR requirement documents, together with the evaluation aspects from the Model for ASsessment of Telemedicine framework [35]. The quality aspects appearing in the result of the UVON method are supposed to be provided by the eHealth apps developed in the FI-STAR project. For each quality appearing in the UVON's output, a question was formulated according to that specific quality in the treatment. The questions were categorized according to the resulting domains in the UVON's output. The answer alternatives to the questions were formed as a 5-point Likert rating scale with unweighted scores. There were 20 qualities alongside the 3 user satisfaction questions from ECSI [18] that were converted to 2 questionnaires. One questionnaire was customized for the patients and the other one for the health professionals. The content of the questionnaires can be found in Multimedia Appendices 2 and 3. Responses to the questionnaires by 87 patients and 31 health professionals, physicians or nurses, were used for the models suggested in this study.

In 2 steps, we arrived at a model based on the empirical data from the answers to the questionnaires. The calculations were done using the R language, version 3.4.0 [36]. The bootstrapped significance calculation was performed in SmartPLS software version 3.2.7 (SmartPLS GmbH) [37].

In the first step, a matrix of Kendall correlation coefficients tau (τ) for each of the patient and professional questionnaires was formed. The results are presented in Figures 1 and 2. Moreover, we used Cronbach alpha (α) test to measure the consistency of the results in the UVON-suggested families of qualities as hints for finding constructs in the later steps.

In the second step, we created a PLS-SEM path model. For each set of the qualities that have already been grouped by the UVON method and show a high degree of correlation, a latent construct variable was considered. These latent variables are not directly measurable but manifest themselves through quality and satisfaction variables. If a quality corresponds only with one question in the questionnaire, one latent proxy variable was considered. Consequently, it would be possible to add more measure variables to the same latent variable in future studies. The PLS-SEM analysis was performed using the *matrixpls* library in R, version 1.0 [38]. The sample size adequacy calculations were performed using G*Power version 3.1 (Heinrich Heine University Düsseldorf), a program for statistical power analysis for a variety of statistical tests [39].

The result of PLS-SEM should be interpreted in the context of the questionnaire. Accordingly, as discussed in the Partial Least Squares Structural Equation Modeling Path Models section, negative coefficients were considered noninformative and were excluded from the final results. The validity of the result was demonstrated through a toolbox of significance, discrimination analysis, internal consistency reliability, and convergence validity. The calculation of significance indicators was performed applying the bootstrapping approach using SmartPLS software [37]. Whenever applicable, the noninformative nature of negative coefficients was considered during the validity and fitness calculations [40].

XSL•FO
RenderX

Figure 1. Correlation matrix for the patient questionnaire results across all cases. For the details of each variable, refer to the corresponding question in Multimedia Appendix 2. Insignificant (*P*>.05) results are left blank. Negative results are marked with leftward slanting lines. Note that the qualities grouped by the Unified eValuation using ONtology (UVON) method usually show higher correlations together.

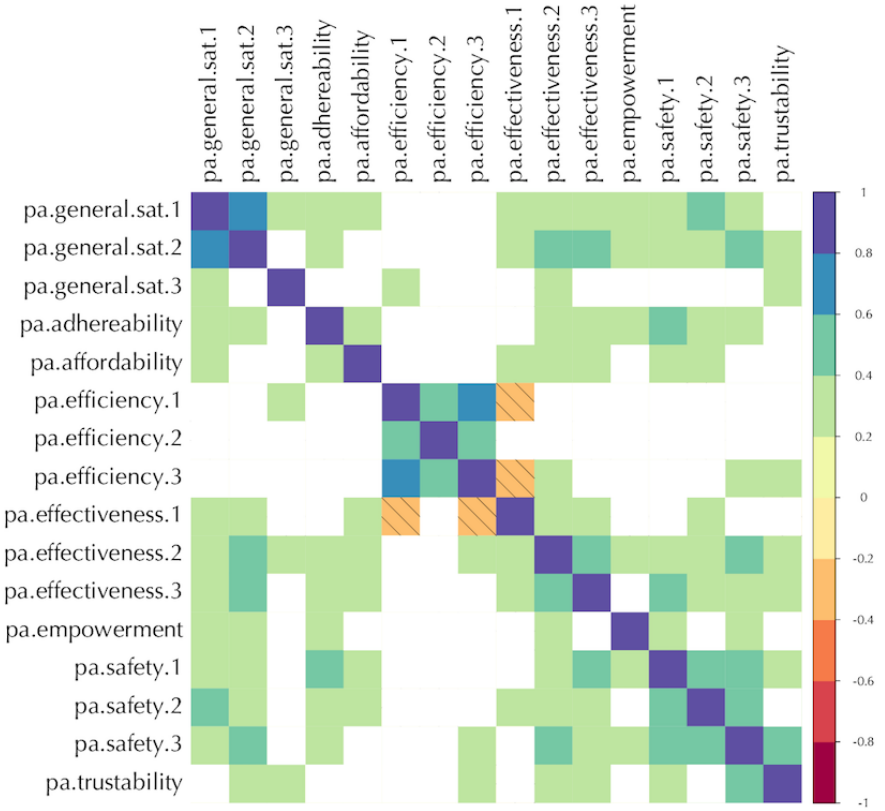
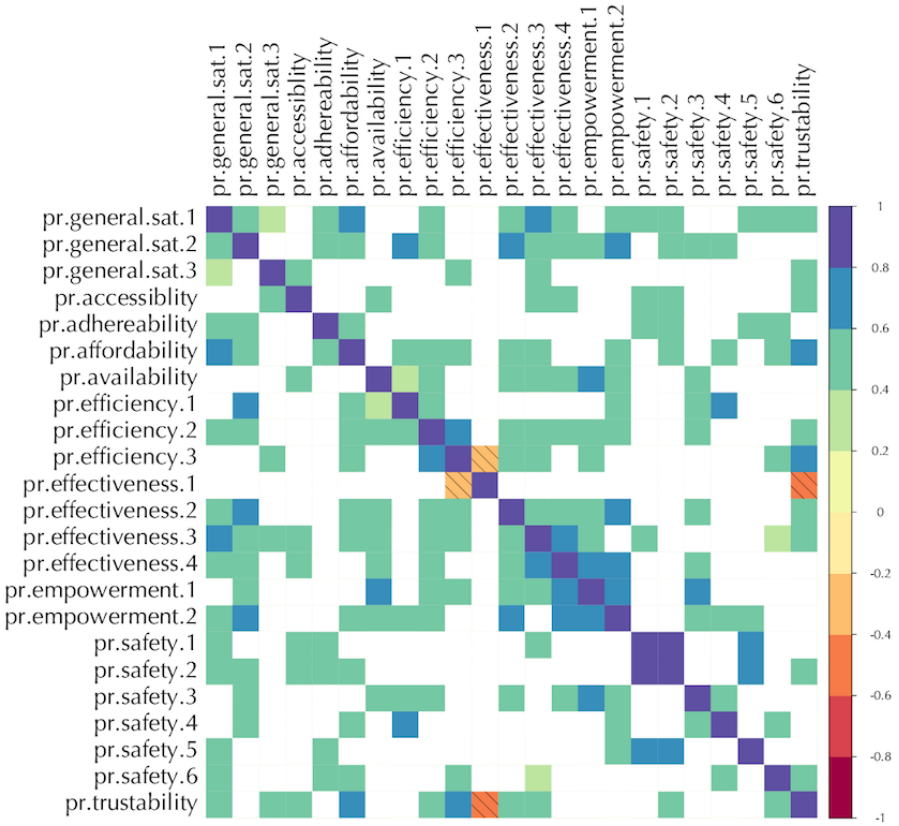


Figure 2. Correlation matrix for the professional questionnaire results across all cases. For the details of each variable, refer to the corresponding question in Multimedia Appendix 3. Insignificant (*P*>.05) results are left blank. Note that the qualities grouped by the Unified eValuation using ONtology (UVON) method usually show higher correlations together.



XSL•FO
RenderX

Results

Unified Evaluation Using Ontology Method Outcome

The result of applying the UVON method on the FI-STAR project was a tree-style ontology of qualities [33], of which the top-level qualities are listed in Table 1. The questionnaires articulate those qualities and their more specific subqualities. Table 1 is just an overview of the qualities; details of the questions that were created for each quality can be found in Multimedia Appendices 2 and 3.

Descriptive statistics of the variables in the patient and professional questionnaires, including mean, SD, and median, are shown in Tables 2 and 3. Details of the corresponding question for each quality in Tables 2 and 3 can be found in Multimedia Appendices 2 and 3.

Correlation Pattern

As can be seen in Figures 1 and 2, a spectrum of weak to strong correlation coefficients appeared in the *Corrgram* diagrams [41] for the patient and professional questionnaire. The blank cells are the results that were not statistically important (P>.05). The results of the Cronbach alpha (α) test can be found in Table 4.

Table 1. The quality attributes resulting from applying the Unified eValuation using ONtology (UVON) method to Future Internet Social and Technological Alignment Research (FI-STAR) requirement documents.

Quality name	Description
Accessibility	If the app is accessible to different users
Adherence	If the patients adhere more to treatment because of the app
Affordability	If the treatment became more affordable for the patient or health care system because of the app
Authenticity	If the information provided by the app is authentic and correct (combined with safety)
Availability	If the service provided by the app is available on demand
Efficiency	If the treatment is more efficient because the app was used
Effectiveness	If the treatment process is more effective because the app was used (except for clinical effectiveness)
Empowerment	If the app empowers the patient or health professional to know more about their conditions or perform their tasks better
Safety	If the app itself is safe or makes the treatment process safer
Trustability	If the app improves the trust of the patients in treatment

Table 2. Descriptive statistics of the variables in the patient questionnaire.

Quality[a]	Mean (SD)	Median
pa.adhereability	4.35 (0.73)	4
pa.affordability	4.1 (1.05)	4
pa.effectiveness.1	3.76 (0.94)	4
pa.effectiveness.2	4.28 (0.94)	5
pa.effectiveness.3	4.57 (0.77)	5
pa.efficiency.1	3.82 (0.88)	4
pa.efficiency.2	3.47 (0.9)	3
pa.efficiency.3	3.78 (0.97)	4
pa.empowerment	4.33 (0.9)	5
pa.general.sat.1	4.51 (0.79)	5
pa.general.sat.2	4.46 (0.73)	5
pa.general.sat.3	4.01 (1)	4
pa.safety.1	4.76 (0.46)	5
pa.safety.2	4.58 (0.64)	5
pa.safety.3	4.5 (0.86)	5
pa.trustability	4.62 (0.72)	5

[a]Details of the corresponding question for the items in the Quality column can be found in Multimedia Appendix 2.

XSL•FO
RenderX

Table 3. Descriptive statistics of the variables in the professional questionnaire.

Qualitya	Mean (SD)	Median
pr.accessiblity	3.96 (0.88)	4
pr.adhereability	3.91 (0.9)	4
pr.affordability	4.22 (0.8)	4
pr.availability	3.61 (0.99)	4
pr.effectiveness.1	3.39 (0.78)	4
pr.effectiveness.2	4.04 (0.77)	4
pr.effectiveness.3	4.26 (0.81)	4
pr.effectiveness.4	4.26 (0.81)	4
pr.efficiency.1	3.04 (1.02)	3
pr.efficiency.2	3.65 (0.93)	4
pr.efficiency.3	3.91 (1.12)	4
pr.empowerment.1	4.39 (0.58)	4
pr.empowerment.2	4.3 (0.47)	4
pr.general.sat.1	3.87 (1.1)	4
pr.general.sat.2	3.87 (0.97)	4
pr.general.sat.3	3.52 (0.9)	4
pr.safety.1	4.61 (0.58)	5
pr.safety.2	4.57 (0.59)	5
pr.safety.3	4 (0.67)	4
pr.safety.4	3.78 (0.8)	4
pr.safety.5	4.26 (0.69)	4
pr.safety.6	3.96 (0.82)	4
pr.trustability	4.22 (0.74)	4

[a]Details of the corresponding question for the items in the Quality column can be found in Multimedia Appendix 3.

Table 4. Cronbach alpha (α) test results for the quality groups.

Quality group	Cronbach alpha (α)[a]
Patients	
pa.general.sat.X	.63
pa.efficiency.X	.8
pa.effectiveness.X	.63
pa.safety.X	.67
Professionals	
pr.general.sat.X	.7
pr.efficiency.X	.77
pr.effectiveness.X	.75
pr.empowerment.X	.82
pr.safety.X	.79

[a]Although a score over 0.7 is usually considered the desired cut-off criterion, the composite reliability (CR) values in the table numbered 9 can still better determine reliability.

XSL•FO
RenderX

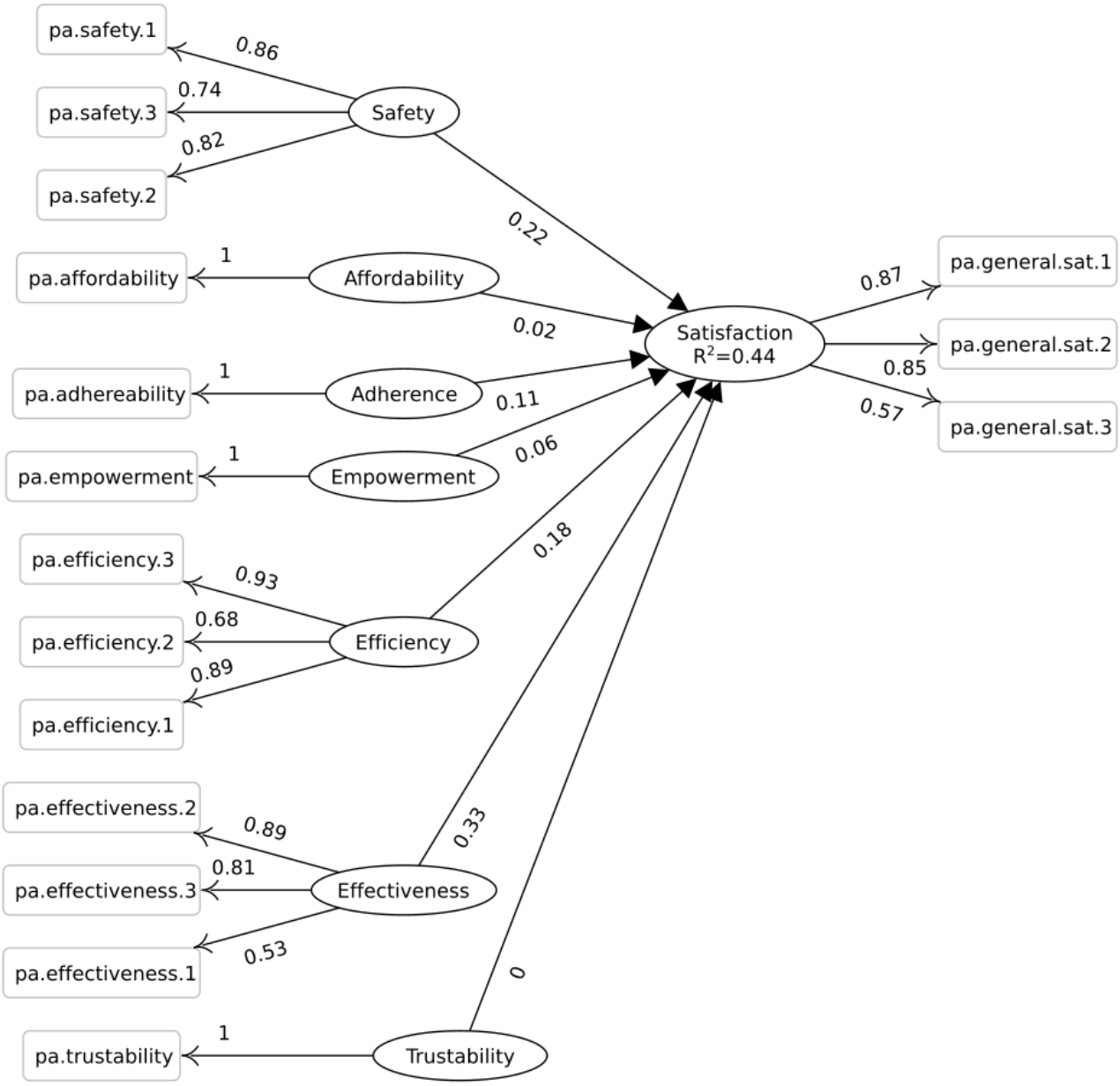
Partial Least Squares Structural Equation Modeling Path Models

The 2 PLS-SEM models and their loadings and coefficient values are depicted in Figures 3 and 4. As is common with path models, latent variables are depicted as ovals, whereas manifests are shown as boxes. We considered all the measure variables as reflective, that is, they do not construct their respected latent variables, but they measure or manifest them. The number label on the edge between a manifest and a latent variable is the loading, and the number label on the edge between two latent variables is the coefficient.

The most contributing and predictive qualities regarding satisfaction are reported in Table 5 by specifying the coefficients of relationships between qualities and the satisfaction construct. For the patients, the coefficients of effectiveness, safety, and efficiency qualities were higher than the average (.13) of all coefficients. Similar to the professional, coefficients of the affordability and effectiveness qualities were higher than the average (.51) of all coefficients.

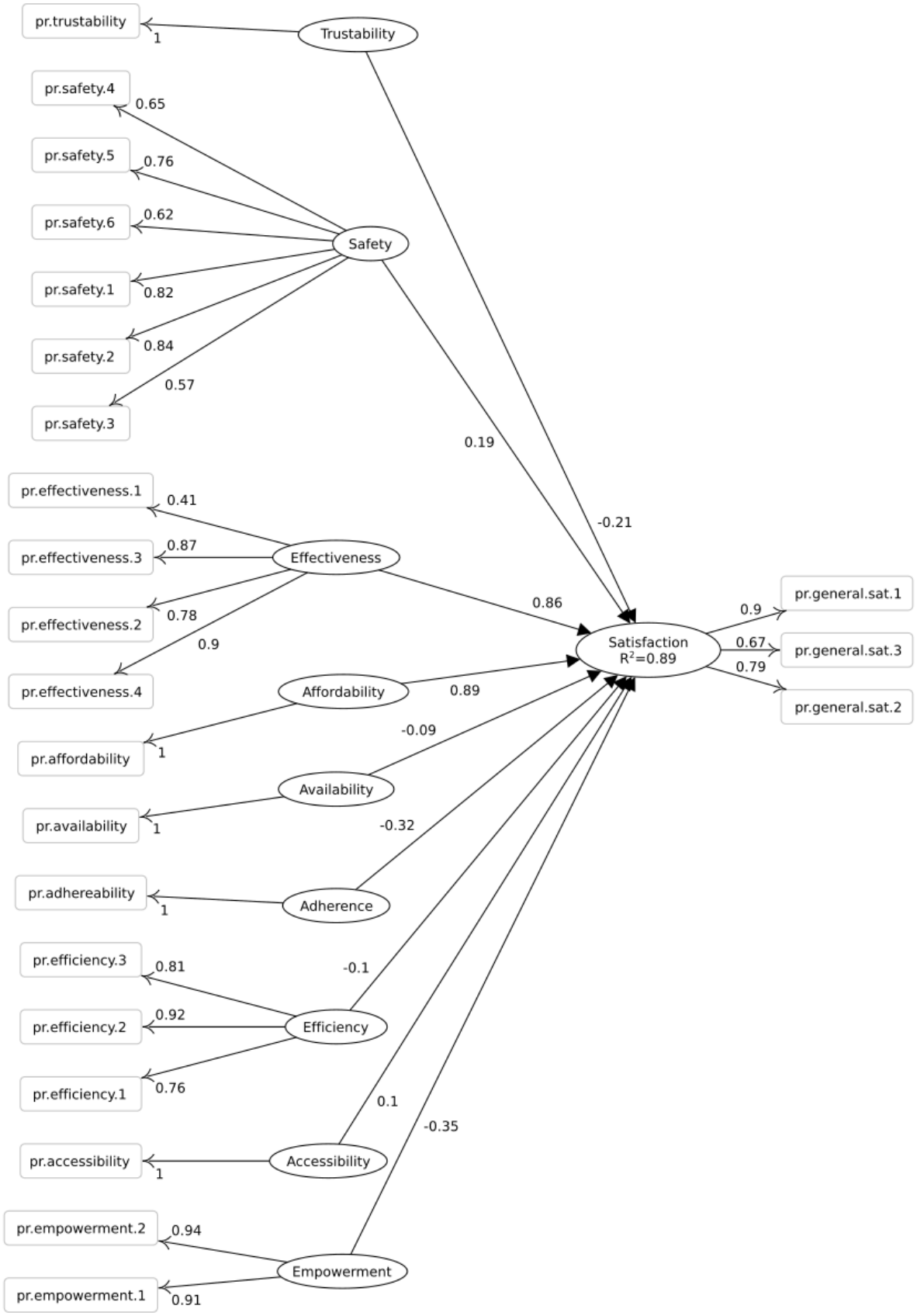
The relationship of each measure variable in the path model with its construct is associated with weights. The standardized weights for the satisfaction construct measure variables are required for calculating the *user satisfaction index* and can be found in Table 6.

Figure 3. Partial least squares path model for the patient questionnaire. The constructs are shown as ovals, and the number between constructs is the coefficient value. Manifests are shown as rectangles, and the number between a manifest and a construct is the loading value of that manifest.



XSL•FO
RenderX

Figure 4. Partial least squares path model for the professional questionnaire. The constructs are shown as ovals, and the number between constructs is the coefficient value. Manifests are shown as rectangles, and the number between a manifest and a construct is the loading value of that manifest.



XSL•FO
RenderX

Table 5. The coefficients of the qualities to satisfaction relationships in the partial least squares structural equation modeling path model show which qualities contribute more to satisfaction.

Quality construct[a]	Patient coefficient	Professional coefficient
Effectiveness	.33	.86
Safety	.22	.19
Affordability	.02	.89
Efficiency	.18	—[a]
Adherence	.11	—
Empowerment	.06	—
Trustability	0	—
Accessibility	—	.1

[a]— represents negative values as being noninformative (see section Partial Least Squares Structural Equation Modeling Models). For the conclusion, one should consider the significance, as shown in table numbered 7.

Table 6. Standard weights for calculating the satisfaction index, based on the manifest variable loadings for the Satisfaction constructs in the patient and professional path models.

Manifest variable	Standardized weight
Patients	
pa.general.sat.1	0.36
pa.general.sat.2	0.38
pa.general.sat.3	0.25
Professionals	
pr.general.sat.1	0.4
pr.general.sat.2	0.32
pr.general.sat.3	0.28

Table 7. Significance of the quality to satisfaction relationships by calculating the *P* values of the relationships between the qualities and the Satisfaction construct. If a relation does not exist in model, the corresponding cell in the table is left blank.

Antecedent to Satisfaction	Patient *P* value	Professional *P* value
Adherance	.16	.13
Affordability	.38	.04[a]
Effectiveness	.01[a]	.01[a]
Efficiency	.03[a]	.37
Empowerment	.25	.20
Safety	.04[a]	.38
Trustability	.48	.30
Accessibility		.41
Availability		.42

[a]*P* values<.05.

The *P* value results of calculating the significance of quality to success relationships using the bootstrapping approach are shown in Table 7. The *P* values below .05 are marked with footnotes. Regarding discriminant analysis, the results of the Heterotrait-Monotrait (HTMT) ratio are demonstrated in Table 8. HTMT being below 1.0, preferably 0.9, satisfies the discriminatory criterion [42]. In addition, Multimedia Appendix 4 depicts cross-loadings in the path models. The effect sizes of the samples were enough to show significant results for the highest loading constructs, as shown in Table 5. Details of the effect sizes and their associated power, by recalculating the PLS-SEM and focusing on significant relations, are shown in Multimedia Appendix 5. The list of acronyms is provided in Multimedia Appendix 6.

XSL•FO
RenderX

Table 8. The discriminant validity analysis shows if the manifests of a construct in the patient or professional Partial Least Squares Structural Equation Modeling models have the strongest relationship with that construct compared with another construct. If a relation does not exist in model, the corresponding cell in the table is left blank.

Construct pairs (A vs B)[a]	Patient HTMT[b]	Professional HTMT
Efficiency→Satisfaction	0.21	0.84
Effectiveness→Satisfaction	0.79	1.07
Safety→Satisfaction	0.69	0.91
Effectiveness→Efficiency	0.04	0.68
Safety→Efficiency	0.09	0.62
Safety→Effectiveness	0.75	0.68
Empowerment→Satisfaction		0.69
Empowerment→Efficiency		0.76
Empowerment→Effectiveness		0.97
Safety→Empowerment		0.63

[a]Heterotrait-Monotrait ratio results below 1.0, preferably 0.9, satisfy the discriminatory criterion.

[b]HTMT: Heterotrait-Monotrait.

Table 9. The result of internal consistency reliability of the manifest variables by calculating composite reliability and their convergence by measuring average variance extracted, grouped by constructs.

Construct[a]	CR[b] patient	AVE[c] patient	CR professional	AVE professional
Satisfaction	0.82	0.6	0.83	0.63
Adherence	1	1	1	1
Affordability	1	1	1	1
Efficiency	0.88	0.71	0.87	0.7
Effectiveness	0.8	0.57	0.84	0.58
Empowerment	1	1	0.92	0.85
Safety	0.85	0.66	0.86	0.51
Trustability	1	1	1	1
Accessibility			1	1
Availability			1	1

[a]A composite reliability value above 0.7 and an average variance extracted value above 0.5 are preferred.

[b]CR: composite reliability.

[c]AVE: average variance extracted.

For internal consistency reliability, composite reliability (CR) values, and for convergence validity, average variance extracted (AVE) values were calculated for each construct, as depicted in Table 9. The minimum CR should preferably be above 0.7 [43]. The minimum AVE should preferably be above 0.5 [44].

Discussion

Overview

This study advances and prioritizes the qualities in health information systems that determine and predict user satisfaction, both for patients and health professionals. As a secondary outcome, it also suggests weightings for calculating the *satisfaction index*. The outcomes of the study exhibit the effectiveness of the UVON method in proposing quality constructs that can be applied to a path analysis. Conclusions from the results are achieved in 2 steps. First, the correlations give better insight about the groupings of the qualities as manifest variables of the latent constructs. Second, the path model justifies and quantifies the relationship between those grouped qualities, their latent construct variables, and the satisfaction construct.

Correlations

In the exploratory step, as shown in Figures 1 and 2, strong correlations appear between the qualities that have been already grouped into a family by the UVON method. These high correlations result from the semantic unification of qualities across branches of an ontological tree by the UVON method [33]. For example, a set of above-medium correlations exist within the efficiency, effectiveness, and safety family of qualities in the patient questionnaire results, as well as within

XSL•FO
RenderX

the empowerment, effectiveness, safety, and efficiency families in the professional questionnaire results. The same is true for the satisfaction questions in both groups of questionnaires.

The above correlations suggest that the members of a quality or satisfaction group can be combined. Alternatively, in other words, they manifest a common latent variable. It is worth mentioning that the Kendall correlation tau (τ) is less generous than Spearman rank correlation rho (ρ) in confirming the correlations [45]. Hence, there would be more confidence in positively interpreting the correlation results and suggesting a common latent origin. The Cronbach alpha results in Table 4 confirm the same explanation in the quality groups.

Besides the possible existence of latent variables, there are 2 other corollaries to the correlations. First, the high degree of correlation between a family of qualities results in the multicollinearity problem. Multicollinearity makes ordinary regression techniques inefficient and the interpretation of the regression coefficients challenging [46]. Overall, 2 solutions can be taken here: choose one of the variables that show high correlation with each other by using variable selection methods, such as Least Absolute Shrinkage and Selection Operator, or apply a method that is tolerant to the problem. The PLS-SEM approach used in this study is tolerant to multicollinearity; meanwhile, it can investigate the causality relations between some correlated quality groups.

Second, the correlation between qualities and satisfaction aspects suggests a causality relationship between quality and user satisfaction. Similarly, there are models, such as the CSI family of models [11-13] as well as the D&M IS model [20], that demonstrate a causality relationship between qualities and satisfaction in parts of their structure. We can draw on corroborations from the extensive amount of literature about those models, both to enrich our model and verify the results.

A summary of the above discussion is that we can group the qualities within a family as manifests of a latent variable, consider a causal relationship from those quality latent variables to a satisfaction latent variable, and present these groupings and relationships through a PLS-SEM model.

Partial Least Squares Structural Equation Modeling Path Models

The PLS-SEM path model has traditionally been used to represent causalities for the CSI series [11,12,47], the models related to D&M IS [48,49], and similar intentions [50]. This prevalence of usage gives the opportunity to reuse some of those models' parts, compare their structures, and collate their results. Other advantages of the PLS-SEM approach are the need for small sample size and the ability of handling non-normal data [50].

The 2 PLS-SEM path models in this study comply with the general pattern in CSI, D&M IS, and TAM series models in which a central construct—be it called system success, user satisfaction, customer satisfaction, or user acceptance—is influenced by system qualities. Besides the use of different constructs, each model captures a distinct level of detail for the same or similar concepts. The CSI and TAM models are more concerned about the general perception of quality, whereas D&M IS examines further details about the qualities by considering 3 separate constructs: *system quality*, *information quality*, and *service quality*. The model presented in this study is inclined to be more domain-specific by focusing on the health information system domain. The model is also more concerned with the qualities improved in the whole treatment setting by using the health information system rather than the qualities of the system. Finally, in comparison with the previously mentioned mainstream models, the model in this study is more specific about the type of qualities and how they compare in determining and predicting user satisfaction.

Before discussing the qualities with the most influence on satisfaction, model validity and the right way of interpreting the results need to be investigated.

Regarding internal consistency reliability, CR indicators need to be higher than 0.7 [43], which is well satisfied (see Table 9). The CR shows if the manifest variables of each construct measure the same thing. The convergent indicator AVE needs to be more than 0.5 to indicate that more than half of the variance in the measures is because of the variance in the construct [44]. All the constructs in our PLS-SEM model satisfy this criterion (refer to Table 9).

From the discriminatory validation perspective, both patient and professional models show indications of correctly assigning the measure to construct variables. The HTMT ratio demonstrates if the assignment of measuring variables to a specific construct is better, that is, more relevant, than other alternatives. All the HTMT results in both models, except one as shown in Table 8, satisfy the specified criterion of being less than 1. Furthermore, most of the HTMT values are less than 0.9, which confirms discriminant validity [42]. The only pair of constructs that have a ratio value above 1 is the *effectiveness* and *satisfaction* pair for professionals. However, this unfulfilled criterion might be justified considering that, in the domain, one should segregate satisfaction and effectiveness, while effectiveness highly contributes to satisfaction. The relatively high HTMT ratio for the pair of *empowerment* and *effectiveness* can indicate that the users' empowerment to reach effectiveness is not very distinctive from the improved effectiveness.

The interpretation of the negative coefficients that appeared in the models must be discussed. The wording in the FI-STAR questionnaires captures only the user perspective on positive relationships but not the negative ones. The information gathered from the questions is unidirectional. Therefore, one cannot interpret the negative coefficients as an indication that some qualities are inversely related to satisfaction.

For example, if it is asked whether the app has increased the effectiveness of a treatment by decreasing the number of mistakes, a responder might answer "disagree." This answer can mean whether the user does not believe that the app has decreased the number of mistakes in the treatment or the user might think that the app has decreased the number of mistakes but does not contribute to the system's effectiveness. On the other hand, it is still a possible interpretation that the app has caused more mistakes; however, we cannot separate this interpretation from the other previously mentioned valid

XSL•FO
RenderX

interpretations. Therefore, we can only confirm the positive part of relationships, where quality contributes to user satisfaction.

Most Influential Qualities

As shown in Table 7, there are constructs whose relationships to the satisfaction construct are statistically significant. Those constructs also have a considerable impact on satisfaction, as depicted in Table 5. Regarding satisfaction as the major contributor to success [51] and the indicator of voluntary acceptance [14], we extend our discussion to cover similar studies that report on these 2 indicators.

Within the list of qualities in Table 7 with a significant relationship to the satisfaction construct, the degree of *effectiveness* is considerably predictive in creating satisfaction. Both patients and professionals care considerably whether an app has increased the effectiveness of treatment and care. This result highlights the nonintuitive contrast between the effectiveness of other qualities, such as the efficiency for patients, in affecting their satisfaction. Nevertheless, there can be alternative interpretations. For example, if the apps in the FI-STAR project could significantly improve efficiency for patients, efficiency might have shown an impact as the effectiveness on satisfaction.

This result confirms the studies that consider effectiveness the major contributor or even equivalent to user satisfaction, generally, in information systems [14,51,52]. More specifically, this study parallels the studies that reported effectiveness (sometimes expressed as usefulness) as the most, or one of the most, influential qualities for the satisfaction of patients [53] or health professional users [54-56] in a variety of health information systems. Nevertheless, there exist studies that reached a different conclusion in prioritizing the most influential qualities [57].

It should be noted that improvements in the *effectiveness* of the treatment are not articulated identically in all those studies. For comparison, one needs to consider this discrepancy [8]. Some studies have used *performance* [54], a term presumably borrowed from the D&M IS and TAM families. In addition, some studies reported similar manifest variables to effectiveness in our study—see Multimedia Appendices 2 and 3 —such as making fewer mistakes [58].

According to our results, *affordability* has a high degree of impact on satisfaction for professionals, similar to effectiveness. The affordability to satisfaction relationship for the patients was not statistically significant in this study, whereas its magnitude was also negligible compared with other relations. An explanation might be that, in the FI-STAR setting, patients did not have to be concerned about the costs and affordability of solutions, whereas professionals might have a more holistic perspective. In a different setting, where patients are more concerned about treatment costs, their satisfaction might be more influenced by the improvements in affordability, showing higher significance and magnitude in the coefficient that relates affordability to satisfaction constructs.

Although studies report increased affordability and cost reduction in treatment can improve the satisfaction and acceptance of health information systems [2,59], only some of these studies quantify or compare the degree that affordability and cost reduction affect satisfaction. In addition, many studies such as ours could not report definitively how patients perceived cost-reduction quality considering patients usually do not pay for treatment in the context of a study [5,60]. Some studies that rely on TAM models have considered affordability, alongside other factors, as a manifest to *perceived usefulness* [8,61,62]. These studies show a relatively high or above-average impact on professional users' satisfaction, acceptance, or intention to use [61,62]. Our results in Table 5 comply with these studies. It is worth noting that some studies declared the same idea in a negative form, where being costly is considered a barrier to acceptance [2,60,63] or success [7].

Patients showed some degree of improved satisfaction when there was an improvement in *efficiency* or *safety*. It is important to note that health information systems, as a side effect, can degrade the status of efficiency or safety in a health care setting [64]. Hence, their contribution to overall satisfaction can even be negative. However, as highlighted before, our questionnaires were not designed to differentiate between the states of negative impact and no impact.

Considering most of the apps in the FI-STAR project could be categorized as telemedicine apps (see Multimedia Appendix 1), it could be predicted that efficiency, achieved by eliminating the hassle of distant travels, contributes to user satisfaction to some extent. Despite our initial expectations, the degree of impact on satisfaction, although existing, was less than the previously mentioned factors. Our expectation was based on similar studies that investigated the impact of efficiency improvements on satisfaction and acceptance: the degree of impact was recorded relatively more than our study [1]. It is also important to note that we considered a separate category for affordability and cost saving quality, whereas some studies considered cost saving a form of improving efficiency [65] or a manifest to *perceived usefulness* [62]; therefore, other results should be compared with more attention to this detail. From the other side, some studies considered other forms of efficiency, rather than affordability, as manifests to *perceived usefulness* [55,66,67]. Regarding the coefficients and loadings in models of the aforementioned studies, in comparison with our results, they have recorded a higher impact from efficiency on the satisfaction or acceptance of users.

Looking at the safety questions for the patients in Multimedia Appendix 2, it seems that being informed about the situation and capable of keeping the situation in check is the source of the safety to satisfaction causality. Similar to other qualities, safety has been categorized in various constructs in the studies, whether as a manifest to *provider performance* [68], *perceived usefulness* [62], *outcome* [62], or *information satisfaction* [62].

To the best of our knowledge, there are few studies that investigate the impact of safety brought by health information systems on the satisfaction or acceptance of the users of those systems. However, one should pay attention to this caveat that the *safety* concept might have divergent embodiments in various studies. Two of the manifest variables in our models, *pa.safety.2* and *pr.safety.3*, refer to providing correct information, which is also mentioned in distinct studies, mostly as a manifest

XSL•FO
RenderX

variable for the *information quality* construct [26,58,69]. In these studies, providing the correct information influences satisfaction or acceptance relatively high or above average. Moreover, there are some studies on the systems in which their primary function is to improve safety aspects. As can be anticipated, they report a high impact from safety on the intention to use [70,71].

Satisfaction Index

Each of the satisfaction constructs in the PLS-SEM models is operationalized by 3 measure variables, embodied as 3 questions. The relationship of each of these measure variables with the latent satisfaction construct is characterized by loading and weighting values. The weighting values make it possible to calculate a weighted *satisfaction index*, both for patients and professionals. Using this index can be a makeshift way of assessing qualities inside health information systems.

The standardized weightings for the scores of the 3 measure questions of satisfaction are determined by the overall balance in the model as it is implemented by the PLS-SEM algorithm. The weightings, along with the scores of those 3 questions, make it possible to calculate the user satisfaction that is engendered by the improvement of qualities. Without the weightings, the satisfaction score represents the evaluation of a kind of unidentified trait that is relevant but not necessarily the same as a quality engendered satisfaction. A larger and more diverse sample population of respondents and apps might be needed to stabilize the weightings for a larger scope.

Collecting satisfaction scores is a common practice in health-related studies. However, to the best of our knowledge, the studies on health information systems that suggest a kind of satisfaction index with adjusted weights regarding other qualities are limited. Conversely, the studies that have calculated the parameters of TAM or D&M IS models in their contexts or have used a path model that includes user satisfaction alongside other impacted qualities, such as studies by Jo et al, Schaper et al, and Pai et al [53,56,69], implicitly suggest a kind of indexing for adoption and satisfaction aspects. Nevertheless, there are barriers to utilizing the measures suggested for the satisfaction construct. First, the extent and diversity of subjects in these studies are important factors to reuse their suggested measures and their associated weights. Second, using arbitrary measure questions or the number of items for measurement can create a burden beyond a study's resources. A sample of this case is the End User Computing Satisfaction measures that range from 12 to 39 items [52,72]. Although some studies have the required resources to apply them [25], using those instruments is not feasible in many other cases. Our focus on the 3 standard satisfaction measures from CSI makes it easier to implement the study and simultaneously facilitates the running of interdisciplinary comparisons and knowledge about the satisfaction of users (customers) of health information systems and other services and products.

The qualities investigated in this study can explain different amounts of satisfaction variations with R^2=.43 for patient satisfaction and R^2=.88 for professional satisfaction (as depicted in Figures 3 and 4). The satisfaction index can facilitate an informed guess about the qualities when the user perspectives on those qualities are missing or hard to elicit. This approach is a makeshift way to evaluate the qualities improved by a system. A practical application might be to use the satisfaction index when comparing a pair of similar systems in similar contexts. Another application is to compare the past and present state of the same system that has undergone quality improvement, but no other system or environmental aspect has been changed. Generally, if there are similarities in context and functionalities, and there has been no drastic change or difference in qualities improved by systems, the satisfaction index can serve as a good indicator for an informed guess about those qualities.

Extensions and Limitations

The list of most influential qualities should be read with the precaution of how the similar or even the same qualities have been articulated differently in studies [73]. Studies that recruit highly cited frameworks also tend to recruit similar wordings for the qualities. However, other studies practice their freedom to use the wording that best matches their case, resulting in divergent wordings for similar concepts. In our case, notably, the improvement of a treatment's *effectiveness* is largely similar to *performance* or *performance expectancy* in the studies that are based on the TAM family, and consider the *performance expectancy* definition as "the degree to which an individual believes that using the system will help him or her to attain gains in job performance" [74]. It is similar to what we asked about effectiveness (refer to Multimedia Appendices 2 and 3). However, one can still find research based on TAM studies that used *performance expectancy* as a form of efficiency [29]. In addition, *efficiency* in our study is more or less similar to *effort expectancy* or *perceived usefulness* in the TAM series. Some other studies have used *productivity* or even *performance* instead of what we called efficiency [68], and others considered efficiency as a manifest to *perceived usefulness* [8,55,66,67].

Another caveat when comparing the results is that the qualities specified in our study are the qualities of the treatment or care that using FI-STAR systems might have improved, which are different from the intrinsic qualities of those systems. Accordingly, for example, a higher speed system might increase the affordability of treatment and care in some way. What we have focused on was the affordability of care but not the speed of the system. Therefore, reports on the intrinsic system qualities that have increased user satisfaction or acceptance cannot be compared directly with our study, unless those qualities get translated to their final impact on treatment and care.

All the eHealth apps, being developed in the FI-STAR project, were supposed to support the required general technical specifications, which includes using the FIWARE infrastructure and being based on *software to data* paradigm [75]. These requirements were not constraints on the diversity of apps as indicated in Multimedia Appendix 1. Nevertheless, the diversity of the apps should be considered when generalizing the results of this study. For example, if the main outcome of an app is improving the safety of treatment, users might consider both safety and effectiveness almost the same; however, in other apps, they can consider them as distinctive qualities. As another

XSL•FO
RenderX

example, the effect of some apps on affordability of treatment can be varied in a different context, when the patients pay for the treatment. This study relies on the output of the UVON method, which extracts common qualities between a set of apps; however, as it is shown in abovementioned examples, user perspectives on those common qualities could be diverse. Therefore, generalization of the results of this study should be done bearing this caveat in mind.

The predictive power of qualities in projecting user satisfaction can support design decision making for health information systems. When trade-offs are necessary, designers can prioritize features if they can compare their user satisfaction yield. Knowing the quality profile of each feature, one can combine that with the table of quality to satisfaction magnitudes, such as in Table 5, to arrive at more informed feature selection decisions [76].

Another extension to the model of this study is to consider the qualities or system usage ramifications that impact satisfaction negatively. This needs articulation of questions to capture negative attitudes—not noninformative ones—about the impacts of systems. That kind of wording permits elicitation of constructs that are negatively related to other constructs or new manifest variables for the current constructs that reflect the construct negatively.

A possible future extension to our PLS-SEM model is to consider relationships between quality latent variables. In the model presented in this study, no relationship has been proposed between the quality constructs, but one might try, for example, to investigate if a system that improves adherence also changes user attitudes about its contribution to effectiveness. However, considering that only recursive relationships can be used in the PLS-SEM models, we cannot investigate the circular impact between qualities and satisfaction with this technique [77,78].

Conclusions

The satisfaction of health information system users is highly influenced by certain qualities that are improved by those systems. Both patient and professional users consider improvements to the effectiveness of health care a highly important quality that makes them satisfied with the system. For patient users, safety and efficiency qualities come after effectiveness in creating satisfaction. For health professionals, better health care affordability brought by health information systems is important, much like effectiveness, in creating their satisfaction.

The PLS-SEM model presented in this study can demonstrate the above ranking of qualities in the creation of user satisfaction. Furthermore, the model suggests weightings to calculate the *satisfaction index* for health information systems. The satisfaction index can be used to compare and monitor health information systems from user satisfaction and quality improvement perspectives.

Acknowledgments

The authors would like to acknowledge the contribution of project partners, especially Dr Samuel Fricker from the FI-STAR project, for providing the context of this study. The FI-STAR project was funded by the European Commission under the Seventh Framework Programme under grant agreement number 604691.

Authors' Contributions

SE drafted the manuscript and incorporated the contributions from other authors, contributed to the design of the study, developed the proposed models and methods, and analyzed results. JSB contributed to the design of the study, supervised the research process, and critically reviewed the manuscript. TCL contributed to the design of the study and critically reviewed the manuscript. MF facilitated the collection of data and critically reviewed the manuscript. PA contributed to the design of the study, contributed to the method development, supervised the research process, and critically reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of the FI-STAR trial cases.

[PDF File, 14KB - medinform_v6i4e11252_app1.pdf]

Multimedia Appendix 2

Questionnaire for the patients.

[DOCX File, 15KB - medinform_v6i4e11252_app2.docx]

Multimedia Appendix 3

Questionnaire for the health professionals.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 31KB - medinform_v6i4e11252_app3.docx]

Multimedia Appendix 4

Crossloadings in the path models.

[PDF File (Adobe PDF File), 61KB - medinform_v6i4e11252_app4.docx]

Multimedia Appendix 5

Effect size and power analysis.

[PDF File (Adobe PDF File), 38KB - medinform_v6i4e11252_app5.docx]

Multimedia Appendix 6

Acronyms.

[PDF File (Adobe PDF File), 32KB - medinform_v6i4e11252_app6.pdf]

References

1. Fitterer R, Rohner P, Mettler T, Winter R. A Taxonomy for Multi-Perspective Ex Ante Evaluation of the Value of Complementary Health Information Systems - Applying the Unified Theory of Acceptance and Use of Technology. 2010 Presented at: 43rd Hawaii International Conference on System Sciences; January 5-8, 2010; Hawaii p. 1-10. [doi: 10.1109/HICSS.2010.37]
2. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of ehealth: systematic literature review. Interact J Med Res 2013;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
3. Fritz F, Tilahun B, Dugas M. Success criteria for electronic medical record implementations in low-resource settings: a systematic review. J Am Med Inform Assoc 2015 Mar;22(2):479-488. [doi: 10.1093/jamia/ocu038] [Medline: 25769683]
4. Ammenwerth E, Gräber S, Herrmann G, Bürkle T, König J. Evaluation of health information systems-problems and challenges. Int J Med Inform 2003 Sep;71(2-3):125-135. [Medline: 14519405]
5. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. Br Med J 2000 Jun 3;320(7248):1517-1520 [FREE Full text] [Medline: 10834899]
6. DeLone WH, McLean ER. Information systems success: the quest for the dependent variable. Inf Syst Res 1992 Mar;3(1):60-95. [doi: 10.1287/isre.3.1.60]
7. Heeks R. Health information systems: failure, success and improvisation. Int J Med Inform 2006 Feb;75(2):125-137. [doi: 10.1016/j.ijmedinf.2005.07.024] [Medline: 16112893]
8. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
9. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
10. Cronin JJ, Brady MK, Hult GM. Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. J Retailing 2000 Jun;76(2):193-218. [doi: 10.1016/S0022-4359(00)00028-2]
11. Fornell C. A national customer satisfaction barometer: the Swedish experience. J Mark 1992 Jan;56(1):6-21. [doi: 10.2307/1252129]
12. Fornell C, Johnson MD, Anderson EW, Cha J, Bryant BE. The American Customer Satisfaction Index: nature, purpose, and findings. J Mark 1996 Oct;60(4):7-18. [doi: 10.2307/1251898]
13. Eklöf JA, Westlund AH. The Pan-European Customer Satisfaction Index Programme—current work and the way ahead. Total Qual Manag 2006 Jul 28;13(8):1099-1106. [doi: 10.1080/09544120200000005]
14. Urbach N, Smolnik S, Riempp G. The state of research on information systems success. Bus Inf Syst Eng 2009 Jul 18;1(4):315-325. [doi: 10.1007/s12599-009-0059-y]
15. Aragon SJ, Gesell SB. A patient satisfaction theory and its robustness across gender in emergency departments: a multigroup structural equation modeling investigation. Am J Med Qual 2003;18(6):229-241. [doi: 10.1177/106286060301800603] [Medline: 14717381]
16. Karimi F, Poo DC, Tan YM. Clinical information systems end user satisfaction: the expectations and needs congruencies effects. J Biomed Inform 2015 Feb;53:342-354 [FREE Full text] [doi: 10.1016/j.jbi.2014.12.008] [Medline: 25542853]
17. Hadji B, Degoulet P. Information system end-user satisfaction and continuance intention: a unified modeling approach. J Biomed Inform 2016 Jun;61:185-193 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.021] [Medline: 27033175]
18. Bayol MP, de la Foye A, Tellier C, Tenenhaus M. Use of PLS path modeling to estimate the European Consumer Satisfaction Index (ECSI) model. Statistica Applicata 2000;12(3):361-375 [FREE Full text]

XSL•FO
RenderX

19. Jackson JL, Chamberlin J, Kroenke K. Predictors of patient satisfaction. Soc Sci Med 2001 Feb;52(4):609-620. [Medline: 11206657]
20. DeLone W, McLean E. 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]
21. Urbach N, Müller B. The Updated DeLone and McLean Model of Information Systems Success. In: Information Systems Theory: Explaining and Predicting Our Digital Society. New York: Springer; 2012:1-18.
22. Petter S, DeLone W, McLean E. Measuring information systems success: models, dimensions, measures, and interrelationships. Eur J Inf Syst 2008 Jun;17(3):236-263. [doi: 10.1057/ejis.2008.15]
23. Yu P, Li H, Gagnon M. Health IT acceptance factors in long-term care facilities: a cross-sectional survey. Int J Med Inform 2009 Apr;78(4):219-229. [doi: 10.1016/j.ijmedinf.2008.07.006] [Medline: 18768345]
24. Yusof MM, Paul RJ, Stergioulas LK. Towards a Framework for Health Information Systems Evaluation. 2006 Jan Presented at: 39th Annual Hawaii International Conference on System Sciences; January 4-7, 2006; Hawaii p. 95a-95a. [doi: 10.1109/HICSS.2006.491]
25. Aggelidis VP, Chatzoglou PD. Hospital information systems: measuring end user computing satisfaction (EUCS). J Biomed Inform 2012 Jun;45(3):566-579 [FREE Full text] [doi: 10.1016/j.jbi.2012.02.009] [Medline: 22426283]
26. Kimiafar K, Sadoughi F, Sheikhtaheri A, Sarbaz M. Prioritizing factors influencing nurses' satisfaction with hospital information systems: a fuzzy analytic hierarchy process approach. Comput Inform Nurs 2014 Apr;32(4):174-181. [doi: 10.1097/CIN.0000000000000031] [Medline: 24469556]
27. Ives B, Olson MH, Baroudi JJ. The measurement of user information satisfaction. Commun ACM 1983 Oct 10;26(10):785-793. [doi: 10.1145/358413.358430]
28. Venkatesh V. Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inf Syst Res 2000 Dec 1;11(4):342-365. [doi: 10.1287/isre.11.4.342.11872]
29. Phichitchaisopa N, Naenna T. Factors affecting the adoption of healthcare information technology. EXCLI J 2013;12:413-436 [FREE Full text] [Medline: 26417235]
30. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006;6:3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]
31. Francis J, Johnston M, Eccles M, Walker A, Grimshaw JM, Foy R, et al. Constructing questionnaires based on the theory of planned behaviour: A manual for Health Services Researchers. In: Quality of Life and Management of Living Resources. Newcastle: Centre for Health Services Research; 2004.
32. Eslami Andargoli VP, Scheepers H, Rajendran D, Sohal A. Health information systems evaluation frameworks: a systematic review. Int J Med Inform 2017 Dec;97:195-209. [doi: 10.1016/j.ijmedinf.2016.10.008] [Medline: 27919378]
33. Eivazzadeh S, Anderberg P, Larsson TC, Fricker SA, Berglund J. Evaluating health information systems using ontologies. JMIR Med Inform 2016 Jun 16;4(2):e20 [FREE Full text] [doi: 10.2196/medinform.5185] [Medline: 27311735]
34. Cordis.europa.eu. European Commision: CORDIS; 2016. Future Internet Social and Technological Alignment Research URL: https://cordis.europa.eu/project/rcn/109408_en.html [accessed 2018-06-05] [WebCite Cache ID 6zwXnr6e2]
35. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: MAST. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
36. R-project. Vienna, Austria: R Foundation for Statistical Computing; 2017. R: A Language and Environment for Statistical Computing URL: https://www.r-project.org/ [accessed 2018-10-17] [WebCite Cache ID 73Ep9bbGy]
37. Ringle CM, Wende S, Becker JM. SmartPLS.: SmartPLS GmbH; 2015. SmartPLS 3 URL: https://www.smartpls.com/ [accessed 2018-10-17] [WebCite Cache ID 73EpIScTe]
38. Rönkkö M. Github. 2017. Matrixpls: Matrix-Based Partial Least Squares Estimation URL: https://github.com/mronkko/matrixpls [accessed 2018-10-17] [WebCite Cache ID 73EpTdROc]
39. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
40. Kock N. One-Tailed or Two-Tailed P Values in PLS-SEM? Int J e-Collab 2015 Apr;11(2):1-7. [doi: 10.4018/ijec.2015040101]
41. Friendly M. Corrgrams. Exploratory displays for correlation matrices. Am Stat 2002 Nov;56(4):316-324. [doi: 10.1198/000313002533]
42. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
43. Henseler J, Ringle CM, Sarstedt M. Using partial least squares path modeling in advertising research: basic concepts and recent issues. In: Okazaki S, editor. Handbook of Research on International Advertising. Massachusetts, USA: Edward Elgar Publishing; 2012:252-276.
44. Chin WW. The partial least squares approach to structural equation modeling. In: Marcoulides GA, editor. Methodology for Business and Management. Modern Methods for Business Research. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers; 1998:295-336.
45. Gilpin AR. Table for conversion of Kendall's Tau to Spearman's Rho within the context of measures of magnitude of effect for meta-analysis. Educ Psychol Meas 2016 Sep 7;53(1):87-92. [doi: 10.1177/0013164493053001007]

XSL•FO
RenderX

46. Farrar DE, Glauber RR. Multicollinearity in regression analysis: the problem revisited. Rev Econ Stat 1967 Feb;49(1):92-107. [doi: 10.2307/1937887]
47. Henseler J, Hubona G, Ray PA. Using PLS path modeling in new technology research: updated guidelines. Ind Manage Data Syst 2016 Feb;116(1):2-20. [doi: 10.1108/IMDS-09-2015-0382]
48. Hsu SH. Developing an index for online customer satisfaction: adaptation of American Customer Satisfaction Index. Expert Syst Appl 2008 May;34(4):3033-3042. [doi: 10.1016/j.eswa.2007.06.036]
49. Iivari J. An empirical test of the DeLone-McLean Model of Information System Success. ACM SIGMIS Database 2005 Jun 7;36(2):8-27. [doi: 10.1145/1066149.1066152]
50. Hair Jr JF, Sarstedt M, Hopkins L, Kuppelwieser V. Partial least squares structural equation modeling (PLS-SEM): an emerging tool in business research. Eur Bus Rev 2014 Mar 4;26(2):106-121. [doi: 10.1108/EBR-10-2013-0128]
51. Seddon P, Kiew MY. A partial test and development of Delone and Mclean's model of IS success. Australas J Inf Syst 1996 Nov 1;4(1). [doi: 10.3127/ajis.v4i1.379]
52. Bailey JE, Pearson SW. Development of a tool for measuring and analyzing computer user satisfaction. Manag Sci 1983 May;29(5):530-545. [doi: 10.1287/mnsc.29.5.530]
53. Jo HS, Song TM, Kim BG. Analysis of the factors affecting consumer acceptance of accredited online health information. J Korean Med Sci 2017 Nov;32(11):1757-1763 [FREE Full text] [doi: 10.3346/jkms.2017.32.11.1757] [Medline: 28960026]
54. Liu L, Miguel Cruz A, Rios Rincon A, Buttar V, Ranson Q, Goertzen D. What factors determine therapists' acceptance of new technologies for rehabilitation – a study using the Unified Theory of Acceptance and Use of Technology (UTAUT). Disabil Rehabil 2015;37(5):447-455. [doi: 10.3109/09638288.2014.923529] [Medline: 24901351]
55. Hu PJ, Chau PY, Sheng ORL, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. J Manag Inf Syst 2015 Dec 2;16(2):91-112. [doi: 10.1080/07421222.1999.11518247]
56. Schaper LK, Pervan GP. ICT and OTs: a model of information and communication technology acceptance and utilisation by occupational therapists. Int J Med Inform 2007 Jun;76 Suppl 1:S212-S221. [doi: 10.1016/j.ijmedinf.2006.05.028] [Medline: 16828335]
57. Wu JH, Wang SC, Lin LM. Mobile computing acceptance factors in the healthcare industry: a structural equation model. Int J Med Inform 2007 Jan;76(1):66-77. [doi: 10.1016/j.ijmedinf.2006.06.006] [Medline: 16901749]
58. Kim J, Chae YM, Kim S, Ho SH, Kim HH, Park CB. A study on user satisfaction regarding the clinical decision support system (CDSS) for medication. Healthc Inform Res 2012 Mar;18(1):35-43 [FREE Full text] [doi: 10.4258/hir.2012.18.1.35] [Medline: 22509472]
59. Akter S, D'Ambra J, Ray P. Trustworthiness in mHealth information services: an assessment of a hierarchical model with mediating and moderating effects using partial least squares (PLS). J Am Soc Inf Sci 2010 Oct 25;62(1):100-116. [doi: 10.1002/asi.21442]
60. Handayani PW, Hidayanto AN, Budi I. User acceptance factors of hospital information systems and related technologies: systematic review. Inform Health Soc Care 2017 Aug 22:1-26. [doi: 10.1080/17538157.2017.1353999] [Medline: 28829650]
61. Hsiao JL, Wu WC, Chen RF. Factors of accepting pain management decision support systems by nurse anesthetists. BMC Med Inform Decis Mak 2013 Jan 29;13:16 [FREE Full text] [doi: 10.1186/1472-6947-13-16] [Medline: 23360305]
62. Steininger K, Stiglbauer B, Baumgartner B, Engleder B. Factors Explaining Physicians? Acceptance of Electronic Health Records. 2014 Jan Presented at: 47th Hawaii International Conference on System Sciences; 2014; Hawaii p. 2768-2777. [doi: 10.1109/HICSS.2014.347]
63. Berner ES, Detmer DE, Simborg D. Will the wave finally break? A brief view of the adoption of electronic medical records in the United States. J Am Med Inform Assoc 2005;12(1):3-7 [FREE Full text] [doi: 10.1197/jamia.M1664] [Medline: 15492029]
64. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010;10:231 [FREE Full text] [doi: 10.1186/1472-6963-10-231] [Medline: 20691097]
65. Schweitzer J, Synowiec C. The economics of eHealth and mHealth. J Health Commun 2012;17 Suppl 1:73-81 [FREE Full text] [doi: 10.1080/10810730.2011.649158] [Medline: 22548602]
66. Tung FC, Chang SC, Chou CM. An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. Int J Med Inform 2008 May;77(5):324-335. [doi: 10.1016/j.ijmedinf.2007.06.006] [Medline: 17644029]
67. Zailani S, Gilani MS, Nikbin D, Iranmanesh M. Determinants of telemedicine acceptance in selected public hospitals in Malaysia: clinical perspective. J Med Syst 2014 Sep;38(9):111. [doi: 10.1007/s10916-014-0111-4] [Medline: 25038891]
68. Mohd Salleh MI, Zakaria N, Abdullah R. The influence of system quality characteristics on health care providers' performance: empirical evidence from Malaysia. J Infect Public Health 2016;9(6):698-707 [FREE Full text] [doi: 10.1016/j.jiph.2016.09.002] [Medline: 27659115]
69. Pai FY, Huang KI. Applying the Technology Acceptance model to the introduction of healthcare information systems. Technol Forecast Soc Change 2011 May;78(4):650-660. [doi: 10.1016/j.techfore.2010.11.007]

XSL•FO
RenderX

70. Wong AM, Chang WH, Ke PC, Huang CK, Tsai TH, Chang HT, et al. Technology acceptance for an Intelligent Comprehensive Interactive Care (ICIC) system for care of the elderly: a survey-questionnaire study. PLoS One 2012;7(8):e40591 [FREE Full text] [doi: 10.1371/journal.pone.0040591] [Medline: 22870200]
71. Peeters JM, de Veer AJ, van der Hoek L, Francke AL. Factors influencing the adoption of home telecare by elderly or chronically ill people: a national survey. J Clin Nurs 2012 Nov;21(21-22):3183-3193. [doi: 10.1111/j.1365-2702.2012.04173.x] [Medline: 22827253]
72. Doll WJ, Deng X, Raghunathan T, Torkzadeh G, Xia W. The meaning and measurement of user satisfaction: a multigroup invariance analysis of the end-user computing satisfaction instrument. J Manage Inf Syst 2014 Dec 8;21(1):227-262. [doi: 10.1080/07421222.2004.11045789]
73. DeFranco J, Kassab M, Laplante P, Laplante N. The nonfunctional requirement focus in medical device software: a systematic mapping study and taxonomy. Innovations Syst Softw Eng 2017 Aug 12;13(2-3):81-100. [doi: 10.1007/s11334-017-0301-6]
74. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
75. Fi-star.eu. 2014. About FI-STAR URL: https://www.fi-star.eu/about-fi-star.html [accessed 2015-09-29] [WebCite Cache ID 6buGZ5wU3]
76. Eivazzadeh S, Anderberg P, Berglund J, Larsson T. Designing with Priorities and Thresholds for Health Care Heterogeneity: The Approach of Constructing Parametric Ontology. In: Proceedings of the 20th International Conference on Engineering Design (iced 15) Vol 2: Design Theory and Research Methodology Design Processes. Glasgow: Design Soc; 2015 Presented at: 20th International Conference on Engineering Design; 2015; Milan.
77. Monecke A, Leisch F. semPLS: structural equation modeling using partial least squares. J Stat Softw 2012;48(3). [doi: 10.18637/jss.v048.i03]
78. Fricker S, Thuemmler C, Paulin A. Bscw.fi-ppp.eu. 2015. General, social and ethical-legal requirements URL: https://bscw.fi-ppp.eu/.../D1.5-General,%20social%20and%20ethical-legal%20requireme [accessed 2017-02-03] [WebCite Cache ID 6o0CsjBri]

Abbreviations

AVE: average variance extracted
CR: composite reliability
CSI: Customer Satisfaction Index
D&M IS: Delone and McLean Information Systems Success
ECSI: European Customer Satisfaction Index
eHealth: electronic health
EU: European Union
FI-STAR: Future Internet Social and Technological Alignment Research
HTMT: Heterotrait-Monotrait
PLS-SEM: partial least squares structural equation modeling
TAM: technology acceptance model
UVON: Unified eValuation using ONtology

Edited by G Eysenbach; submitted 08.06.18; peer-reviewed by A Sheikhtaheri, S Koch-Weser; comments to author 12.07.18; revised version received 05.09.18; accepted 08.09.18; published 30.11.18.

Please cite as:
Eivazzadeh S, Berglund JS, Larsson TC, Fiedler M, Anderberg P
Most Influential Qualities in Creating Satisfaction Among the Users of Health Information Systems: Study in Seven European Union Countries
JMIR Med Inform 2018;6(4):e11252
URL: http://medinform.jmir.org/2018/4/e11252/
doi:10.2196/11252
PMID:30504120

©Shahryar Eivazzadeh, Johan S Berglund, Tobias C Larsson, Markus Fiedler, Peter Anderberg. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 30.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Novel Approach to Cluster Patient-Generated Data Into Actionable Topics: Case Study of a Web-Based Breast Cancer Forum

Josette Jones[1], PhD, RN; Meeta Pradhan[2*], PhD; Masoud Hosseini[1], PhD; Anand Kulanthaivel[1*], MIS; Mahmood Hosseini[3], BS

[1]Health Informatics, BioHealth Informatics Department, Indiana University, Indianapolis, Indianapolis, IN, United States

[2]Indiana Biosciences Research Institute, Indianapolis, IN, United States

[3]Purdue School of Science and Technology, Purdue University, Indianapolis, Indianapolis, IN, United States

[*]these authors contributed equally

Corresponding Author:
Josette Jones, PhD, RN
Health Informatics
BioHealth Informatics Department
Indiana University, Indianapolis
535 West Michigan Street
Indianapolis, IN, 46202
United States
Phone: 1 3172748059
Email: jofjons@iupui.edu

Abstract

Background: The increasing use of social media and mHealth apps has generated new opportunities for health care consumers to share information about their health and well-being. Information shared through social media contains not only medical information but also valuable information about how the survivors manage disease and recovery in the context of daily life.

Objective: The objective of this study was to determine the feasibility of acquiring and modeling the topics of a major online breast cancer support forum. Breast cancer patient support forums were selected to discover the hidden, less obvious aspects of disease management and recovery.

Methods: First, manual topic categorization was performed using qualitative content analysis (QCA) of each individual forum board. Second, we requested permission from the Breastcancer.org Community for a more in-depth analysis of the postings. Topic modeling was then performed using open source software Machine Learning Language Toolkit, followed by multiple linear regression (MLR) analysis to detect highly correlated topics among the different website forums.

Results: QCA of the forums resulted in 20 categories of user discussion. The final topic model organized >4 million postings into 30 manageable topics. Using qualitative analysis of the topic models and statistical analysis, we grouped these 30 topics into 4 distinct clusters with similarity scores of ≥0.80; these clusters were labeled Symptoms & Diagnosis, Treatment, Financial, and Family & Friends. A clinician review confirmed the clinical significance of the topic clusters, allowing for future detection of actionable items within social media postings. To identify the most significant topics across individual forums, MLR demonstrated that 6 topics—based on the Akaike information criterion values ranging from −642.75 to −412.32—were statistically significant.

Conclusions: The developed method provides an insight into the areas of interest and concern, including those not ascertainable in the clinic. Such topics included support from lay and professional caregivers and late side effects of therapy that consumers discuss in social media and may be of interest to clinicians. The developed methods and results indicate the potential of social media to inform the clinical workflow with regards to the impact of recovery on daily life.

(JMIR Med Inform 2018;6(4):e45) doi:10.2196/medinform.9162

KEYWORDS

data interpretation; natural language processing; patient-generated information; social media; statistical analysis; infodemiology

XSL•FO
RenderX

Introduction

Health care is currently undergoing transformation by capitalizing on information technology and patient-consumer engagement and activation through health information technology such as patient portals and mHealth apps. Consumer engagement is assumed to strengthen providers' abilities to tailor their care to the consumers' needs, preferences, and abilities. The increasing use of smartphones, mobile apps, and remote monitoring devices, coupled with providers' deployments of electronic health records, patient portals, and secure messaging, offers innovative ways to connect patients and providers and to strengthen consumers' engagement in their health and well-being [1]. In addition, health consumers have embraced social media, enabling them to share and discuss how they manage their health and well-being with others with similar health issues. These social media and mHealth apps generate important data outside the health care settings and, when shared with providers, expand the depth, breadth, and continuity of information available to optimize health care and outcomes.

Despite the proliferation of social media use, such as blogs and forums, little is known about the scope and quality of information shared, or the purposes that social media sites serve for consumer decisional and support needs [2]. Social media retains large amounts of valuable information about consumers' contextual and environmental (day-to-day) factors while managing their health and well-being; such issues form a major foundation of human health. However, analyzing those free-text data to discover these *hidden* aspects of health consumers' lives and understand their health information needs beyond those routinely addressed by health care providers is challenging [3].

This study explores approaches for analyzing the social media data and extract potential valuable information on managing health and well-being beyond the context of health care. As it is known that breast cancer patients and survivors often join social media to fulfill their information needs and discuss their daily challenges and concerns, all or not related to health [4], using those venues is apparent. Some concerns might not be shared with health care providers for many reasons. For example, patients might think it is not necessary to discuss the topic, may feel embarrassed about the issue, or they do not even know there is a problem [5,6]. Hence, we explored ways to discover issues that are not commonly shared but are important for the overall health and well-being.

Methods

Overview

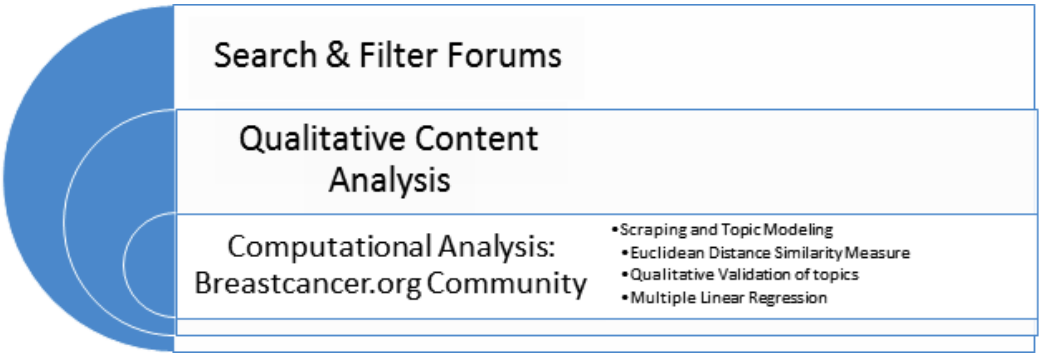
First, manual topic categorization was performed using qualitative content analysis (QCA) of each individual forum board. Second, we requested permission from the Breastcancer.org Community for a more in-depth analysis of the postings. In addition, natural language processing (NLP) and statistical modeling approach were used to cluster >4 million postings into manageable topics. Finally, topic modeling was performed using open source software, followed by multiple linear regression (MLR) analysis to detect highly correlated topics among the different website forums. The methodology is outlined in Figure 1.

Manual Categorization of Posts

A Google search was performed for breast cancer forum websites. Selection criteria were active websites (having posts in the week of search) and in the English language. Each website must have at least 5000 members or have a minimum of 50,000 posts in total, and the posts on the site must be organized into categories. Among the resulting 20 websites, 5 were included on the basis of the selection criteria. These 5 remaining websites contained 4,901,516 posts organized in 211 forums (Table 1). The forum posts were further manually analyzed for consensus among the team members.

Team members were assigned to review the titles of 211 forums across the 5 forum websites and organize them into 4 main top-level categories and 16 subcategories correlating to several domains from the report on the social determinants of health published by the Institute of Medicine [12]. The quantities of posts belonging to each category and subcategory were calculated by the forum that the posts belonged to. The 3 most dominant subcategories across all websites were as follows: Treatments (1.49/4.90 million, 30.5%, posts), Diagnosed–Psychosocial Support–Similar Patients (1.34/4.90 million, 27.3%, posts), and Diagnosed–Psychosocial Support–Life (0.83/4.90 million, 16.9%, posts). After the posts were categorized, the research team iteratively validated (with a random sample of 20 posts from each forum) and consolidated the initial categorization, assuring the quality and correctness of the method.

Figure 1. Overview of the methods used to analyze the study content.



XSL•FO
RenderX

Table 1. Breast cancer websites explored.

Website name	Site URL	Country	Forums	Threads	Posts	Members
Breastcancer.org Community [7]	community.breastcancer.org	US[a]	80	121,688	3,608,324	153,620
Breast Cancer Care [8]	forum.breastcancercare.co.uk	UK[b]	54	36,949	782,486	N/A[c]
Susan G Komen Foundation: Message Board [9]	apps.komen.org/Forums	US	24	44,175	354,592	26,883
Triple Negative Breast Cancer: Forums [10]	forum.tnbcfoundation.org	US	17	9641	100,706	123,427
No Surrender Breast Cancer Foundation [11]	nosurrenderbreastcancersupportforum.com	US	36	2443	55,498	5549

[a]US: United States.

[b]UK: United Kingdom.

[c]N/A: not applicable.

Data Extraction, Natural Language Processing, and Statistical Modeling

Data from a public breast cancer internet discussion forum were extracted, cleaned, and processed; multiple approaches merging NLP with statistical modeling were implemented for knowledge discovery. In addition, off-the-shelf products were used to develop and streamline the analytical approach to cluster most-occurring topics of discussions. The methodology developed revealed several topics that may be of importance for care planning and, thus, need to be incorporated in the electronic health record. In addition, advanced text mining will be a foundation for predictive modeling of consumers' health information needs and provide interactive solutions.

Extracting (Scraping) Forum Data

Postings from the Breastcancer.org Community website were selected for further analyses, as this website contained the highest number of total posts (3.61/4.90 million 73.6%, posts across all 5 websites selected in this study). Permission was obtained from the Web administrators to download and analyze all the data logged in the site.

The Breastcancer.org Community site includes 80 main forums organized by the site administrators into 9 sections. Users self-select in which forum, and thus in which section, a post that they make will go. To capture information within the forum posts, an in-house scraping tool in the *PHP Hypertext Preprocessor* language was developed by a team member. Forum metadata along with the actual posts were extracted; the text within the different posts was aggregated into 80 text files each corresponding to a forum. The files were named based on the forum ID number. The data were saved in the JavaScript Object Notation format. Multimedia Appendix 1 shows the forum names along with the number of threads and posts in each forum.

Applying Topic Modeling

Topic models provide a simple way to analyze large volumes of unlabeled text. A "topic" consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings. One of the leading approaches used for topic modeling is Latent Dirichlet allocation (LDA), which is one of the most popular methods in NLP [13]. LDA represents a document as a distribution of "topics," where a topic is itself a distribution over words (and may or may not be similar to a forum topic). Looping through each word in every document, the LDA algorithm assigns every word to a temporary topic in a semirandom manner and iteratively updates topic assignments. For each word, its topic assignment is subsequently updated based on 2 criteria as follows: (1) the prevalence of the word across all topics and (2) the prevalence of the topics within the documents.

The Machine Learning Language Toolkit (MALLET) open source tool (University of Massachusetts, Amherst, MA, USA) [14] was used to execute the LDA algorithm on the data to extract the main topics. MALLET is a Java-based tool developed at the University of Massachusetts Amherst, which is used for the analysis of data in a textual format such as document classification, clustering, topic modeling, information extraction, and other machine learning apps. After scraping the forum data and saving into 80 files representing each forum, the files were imported into the MALLET tool. MALLET generates two tab-delimited text files as a result of algorithm execution. One file contains the topic ID, and the words related to that topic (aka the topic keywords; Table 2).

MALLET was run iteratively, customized to generate 15 topics, 20 topics, and 30 topics, respectively. Topic labels were added by consensus of the research team based on the semantics of the word cluster. Some topic labels in different sets of topics were identical based on the semantic similarity, but the topic words and strength are different for each of the 3 sets generated. No new topics were generated at the third iteration; the MALLET categorization of 30 word baskets was used for further analysis.

For each iteration and each file, the topic composition and corresponding LDA strength were computed, providing us a way to infer the latent structure of the text file. The resulting output is a topic ID-by-text file matrix known as a file-feature set (Table 3). The first column shows the name of the file; the rest of the columns are best considered as (topic-ID, topic-strength) pairs. For example, it is noted that file F100 has a Topic 12 strength of 0.275 (27.5%). For each document, there are as many of these pairs as there are topics, although only the top 5 topics for each file and the first 4 files are shown for brevity.

XSL•FO
RenderX

Table 2. The partial table of topics generated by Machine Learning Language Toolkit in the 30-topic model, with interpretations (the list goes on up to the 30th topic; only 3 are shown for brevity).

Machine Learning Language Toolkit topic identifier	Topic label	Topic keywords
1	Diagnostic testing and waiting for results	breast biopsy cancer lump results ultrasound benign surgeon mammogram doctor mri waiting back mammo good radiologist feel pain left i'm
2	Side effects of inflammation and its treatment	breast ibc skin symptoms pain rash red cancer nipple biopsy infection diagnosed antibiotics swollen treatment left specialist redness swelling lymph
3	Positive results after recurrence	chemo stage years cancer treatment nodes onc tumor triple negative taxol positive rads year diagnosed node recurrence congratulations lymph radiation

Table 3. A portion of the file-feature set generated by Machine Learning Language Toolkit software (the list goes on up to the 80th file and 30th topic; values were truncated for brevity of display).

File identifier[a]	Topic ID[b]	Strength[c]	Topic ID	Strength	Topic ID	Strength	Topic ID	Strength	Topic ID	Strength
F100	12	0.275	18	0.269	2	0.251	5	0.06	7	0.053
F102	2	0.542	18	0.136	7	0.087	12	0.056	1	0.04
F104	2	0.315	14	0.118	1	0.104	7	0.09	20	0.043
F105	2	0.295	11	0.25	6	0.213	7	0.067	14	0.042

[a]Scraped forum file.

[b]Topic identifier: Machine Learning Language Toolkit-generated topics.

[c]Weight of topic in the file.

Statistical Analyses

The output from MALLET assigned weight scores (ie, topic-strength) to each topic-ID within each file. Statistical analysis was carried out (1) to understand the similarity across the feature sets and files and (2) to identify the topics that are most relevant to patients with breast cancer. Euclidean Distance Similarity Measures (EDSM) were computed to evaluate the similarity across the files based on their weight scores for each topic. Equation (1) is an example of how each file and its feature vectors were assigned, using file F100 as an example.

F100 = [Topic6(.33), Topic1(.28), Topic9(.20), Topic2(.08), Topic0(.05), Topic3(.04), Topic12(.01), Topic4(.01), Topic n(weight m)…] [1]

The EDSM between all potential file pairs were computed on the basis of Equation (2):

Where *i* and *j* are identifiers for each file pair, *k* is the total number of topics in the dataset (ie, 15, 20, or 30 topics), and *x* is the topic weight score in each file.

To identify the most relevant topics in the dataset, an MLR analysis was performed on all files. The MLR analysis was computed using the R Statistical Package [15]. The MLR analysis identified the topics that seem to be most relevant within and across the forums in the study. The equation model (Equation 3) for the MLR used is:

Where $topic_1$, $topic_2$ are the weight scores of the topics in the files; ε_i is the error in the model; β_0 is the intercept; β_1, β_2 are the coefficients for the $topic_1$, $topic_2$, respectively, computed by the model; and y_i is the outcome (dependent variable) for each file *i*.

Results

Manual Categorization of Posts

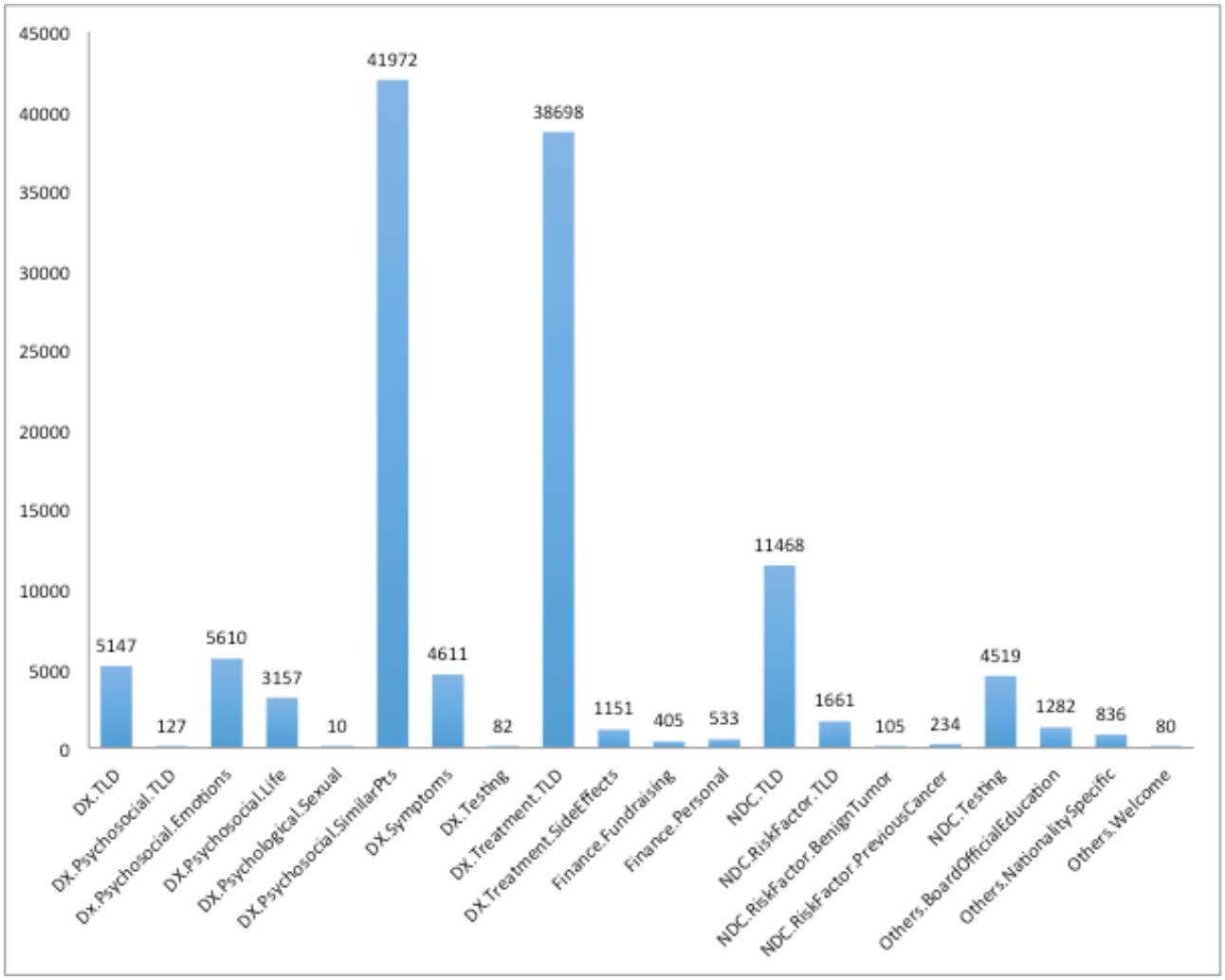
Among the 5 sites studied, Breastcancer.org Community presented the majority of the total post volume. From all websites analyzed for post counts, 73.6% (3.61/4.90 million) of posts were from Breastcancer.org Community; hence, we selected this site for further exploration.

Manual Categorization

Initially, the research team performed via QCA a manual categorization of topics discussed in the 5 selected public websites. The popularity distribution of the manually generated categories, as discussed in the Methods section, was assessed by the number of posts made in the forums. For example, the qualitatively generated categories *Diagnosed—Psychosocial—Similar Patients* have an overall popularity of 41,972 posts. The forum categorization was either general or granular depending on the forum structure. The distribution of qualitative categories across the threads on the Breastcancer.org Community website is visualized in the figures below by frequencies of category popularity (Figure 2); 20 of our QCA-generated categories mapped to forums on breastcancer.org. The *x*-axis shows the QCA-assigned category names.

XSL•FO
RenderX

Figure 2. Distribution of qualitative content analysis-generated categories according to the number of forum threads that each manual category possesses. DX: Diagnosed; TLD: Top-Level Domain; NDC: Not Diagnosed but Concerned.



Data Extraction, Natural Language Processing, and Statistical Modeling

Data Extraction

The data for all 80 forums on the Breastcancer.org Community website were successfully extracted into 80 files, each containing all communications posted over its respective forum.

Natural Language Processing: Topic Modeling

As mentioned earlier, exhaustion was reached at baskets of 30 cooccurring words. The remainder of the analyses will be only for these topics. All machine-generated topics were assigned topic labels based on the semantics of the word cluster and validated by a domain expert (clinical); the topic ID was equated with the term key. MALLET assigned an LDA strength to each topic indicating its overall dominance across all forum files that were analyzed. Two example topics, IDs #8 and #29, are listed in Table 4 (below), along with the authors' labels for these topics. Multimedia Appendix 2 provides a full list of generated topics from this model as well as the authors' semantic interpretations. Each file represents the text of one forum, and topic-strength pairs for the strongest five topics per MALLET LDA analysis of that file are found to the right of the file's ID. For any file, the strength across all 30 topics will always add up to 1.00.

MALLET also correlated the topic relationship strengths between all files based on their topics. Strengths assigned to document-topic pairs by MALLET ranged from almost zero (<0.000001) up to 0.796. The maximum theoretical possible strength for a single file-topic pair would be 1.00. Multimedia Appendix 3 provides a list of the top 5 correlated topics of each file.

Statistical Analysis: Euclidean Distance Similarity Measures

EDSM were calculated to find the similarity between files. Figure 3 shows the file-file similarity matrix and a subset of the similarity matrix. The files are mirrored across both axes and ordered by their alphabetical file name (with F100 being first and F99 being last). Darker cells indicate that the files were more similar. File 109 has a similarity measure of 0.89 with file 104; similarly, file 108 has a similarity measure of 0.78 with file 104.

Table 5 illustrates the similarity measures among the file pairs with a similarity score ≥0.8.

XSL•FO
RenderX

Table 4. Topics #8 and #29 with Latent Dirichlet allocation strengths author topic label interpretations.

Topic identifier	Latent Dirichlet allocation strength	Topic words	Authors' topic label
8	1.38724	cancer chemo years feel life family mom time support things breast people treatment don't husband care friends diagnosed talk mother	Hope, love, family, and friends
29	0.19954	hair book pink survivor happy deb health country president shirley obama congratulations cats article eye mammo fumi beth beautiful vote	Daily living and breast cancer

Figure 3. File-file similarity matrix.

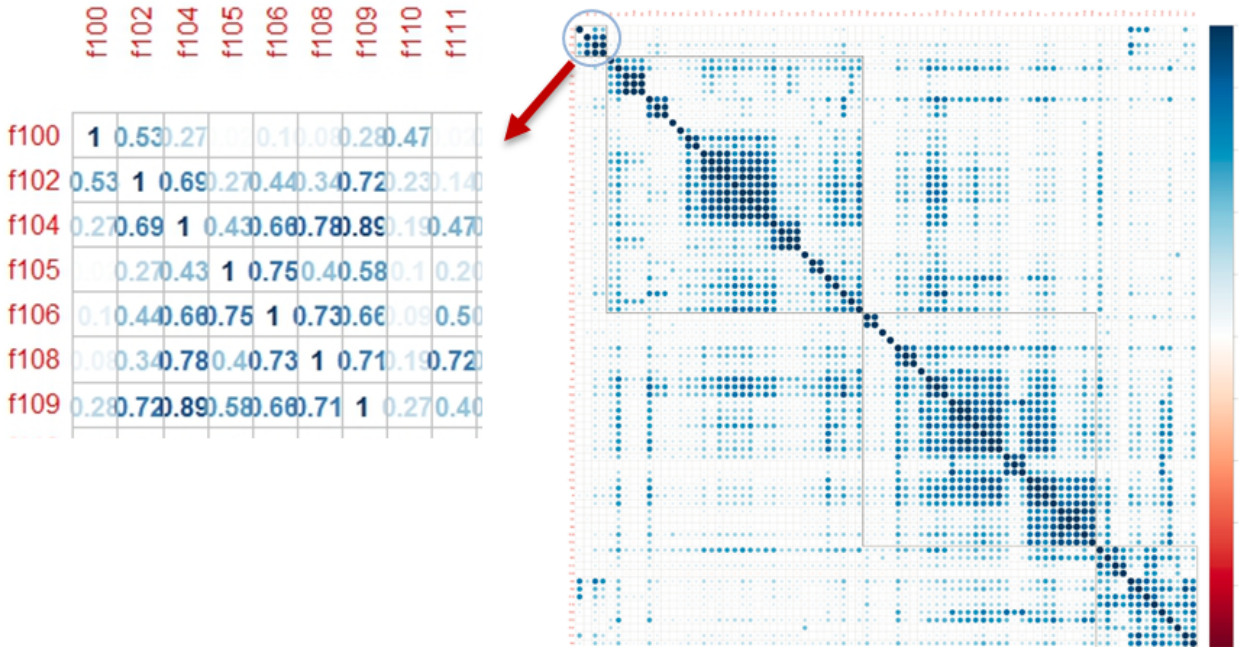


Table 5. Top scored file-file similarity measures.

File identifier	Associated files (similarity score)
F102	F133 (0.85), F144 (0.91), F152 (0.97), F116 (0.95)
F104	F109 (0.89), F142 (0.81), F150 (0.82), F27 (0.89)
F108	F132 (0.94), F137 (0.86), F145 (0.97), F5 (0.90), F71 (0.93), F88 (0.86), F96 (0.97)
F109	F104 (0.89), F142 (0.89), F127 (0.85)
F110	F26 (0.80)
F111	F132 (0.8), F68 (0.87)
F112	F47 (0.92), F93 (0.94)
F113	F139 (0.89), F55 (0.87)
F133	F102 (0.85), F135 (0.86), F145 (0.90), F5 (0.87), F71 (0.94), F96 (0.94), F88 (0.87)

A cursory review of the files with high similarity also revealed clinical relevance and connection. For example, files F108, 132, and 145, while in different discussion categories on the website, all discuss the diagnosis, treatment, and potential side effects from the treatments and also discuss living with different stages and types of breast cancer. In addition, F96 has a high similarity (97%) with F108, which is devoted to the breast cancer type known as invasive ductal carcinoma. F112 discusses more specific genetic risks of breast cancer (BRCA1 or BRCA2 positive), while F47 (similarity 92%) chats about more general risks. F93 at first glance seems not related (Comments, Suggestions, Feature Requests), but reading the postings revealed the need for more information and social support for users who find out that they are *at risk* for breast cancer.

XSL•FO
RenderX

Figure 4. Topic-topic similarity matrix.

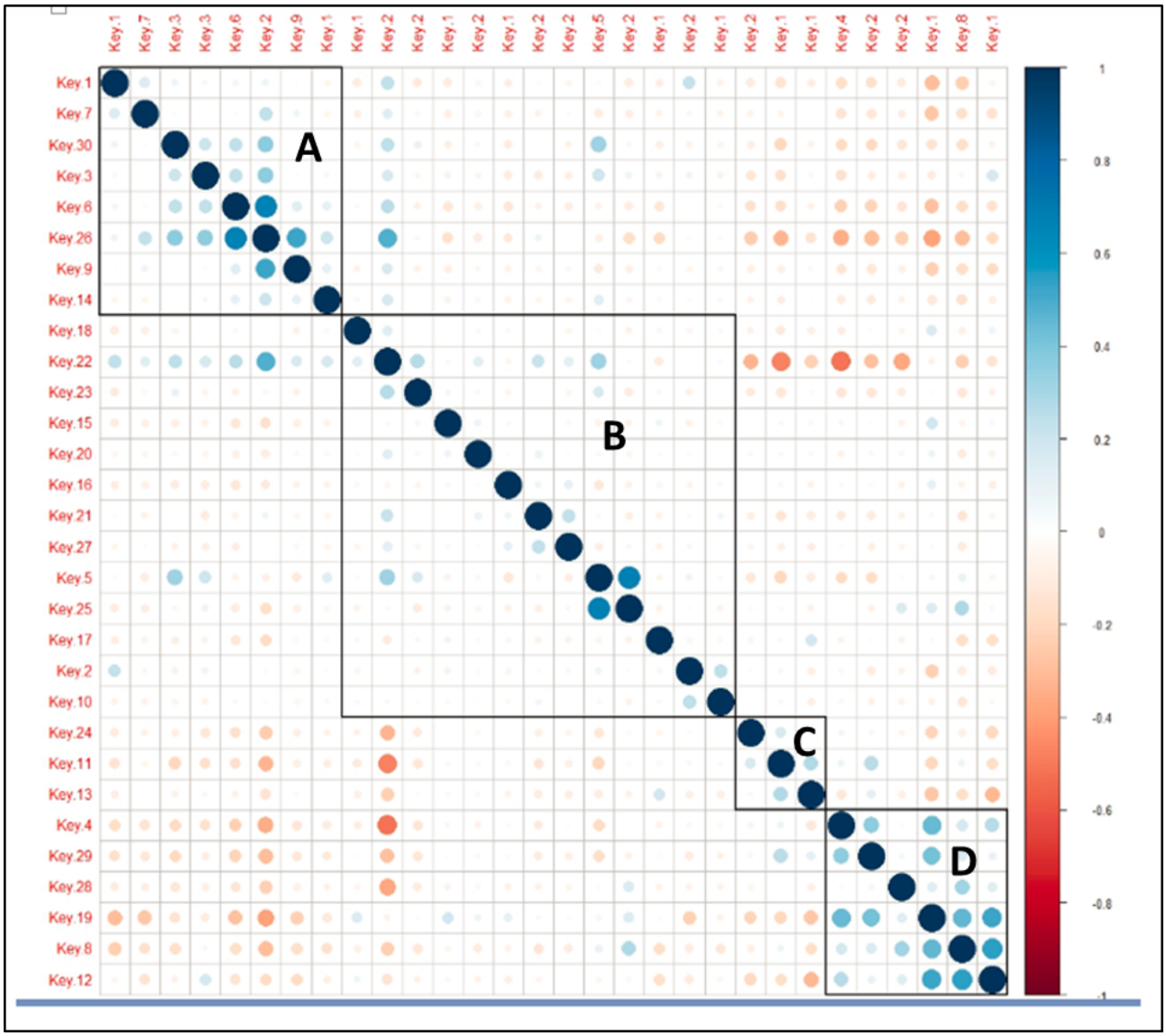


Figure 4 shows 4 clusters of highly correlated computationally modeled topics. Each cluster is labeled with a letter, and topics are labeled as *keys*. The topics with their consensus labels based on the semantic meaning of the word baskets are as follows:

Cluster A: Symptoms & Diagnosis

- *Topic1*: Diagnostic testing and waiting for result
- *Topic 7*: Genetic risk and testing
- *Topic 30*: Symptoms and diagnosis of recurrence
- *Topic 3*: Positive results after recurrence
- *Topic 26*: Positive results after treatment for recurrence
- *Topic 9*: Diagnostic and treatment observation for recurrence
- *Topic 14*: Medical drug treatment and long-term effects

Cluster B: Treatment

- *Topic18*: Chemotherapy side effects and change of treatment
- *Topic 22*: General feeling over time
- *Topic 23*: Medical or drug treatment and side effects
- *Topic15*: Physical activities during and after chemo
- *Topic 20*: Side effects of breast cancer treatment
- *Topic16*: Surgical treatments while in remission
- *Topic 21*: Lingering side effects while in remission
- *Topic27*: Surgical reconstruction during remission
- *Topic 5*: Prognosis about relapse or recurrence
- *Topic 25*: Support from caregivers and medical team for recovery long term
- *Topic 17*: Nutrition and supplements
- *Topic 2*: Side effect of inflammation and its treatment
- *Topic 10*: Radiation side effects and duration of the effects.

Cluster C: Financial

- *Topic 24*: Financial issues over time
- *Topic 11*: Forum-related discussion for support from people in similar circumstances
- *Topic 13:* Looking for clinical research and trials

Cluster D: Friends & Family

- *Topic 4*: Friends and fun
- *Topic 29*: Everyday life and breast cancer
- *Topic 28*: Spirituality and religion
- *Topic 19*: feeling back to normal
- *Topic 8*: Hope, love, family, and friends
- *Topic 12*: Feeling positive and support

XSL•FO
RenderX

Table 6. Most significant topics identified via multiple linear regression analysis.

Topic identifier	Topic label	Akaike information criterion values
21	Lingering side effects while in remission	−642.75
18	Chemotherapy side effects and change of treatment	−641.98
10	Radiation and side effects	−633.17
7	Genetic risk and testing	−620.41
25	Support from caregiver and medical team for recovery long term	−571.78
11	Looking for support from people in similar circumstances	−412.32

As can be deduced from the semantic labeling, each cluster describes a theme: cluster A is related to risk factors, diagnosis, and potential risk of recurrence, whereas cluster B describes different treatments and their side effects in the short and long term. Cluster C and D are less clinically and more oriented to patient contextual factors (ie, those that are typically ascertainable only outside of the clinic encounters).

Statistical Analyses: Multiple Linear Regression Analysis

Finally, MLR analysis was performed to identify the most significant topics (keys) across the 4 abovementioned clusters. The topics were arranged in a descending order based on the Akaike information criterion value, the most appropriate measure for the methodology. The most significant topics identified by the model were: Topic21> Topic18> Topic10> Topic7> Topic25> Topic11. Table 6 reports the most significant topics discussed among the forum participants, along with the respective topic labels assigned by the authors.

Discussion

Principal Findings

It is well known that many users share information online daily. Forum posts, blogs, or other social media activity reveal a rich diary of everyday life. Health information is revealed explicitly when an individual communicates about their well-being or when they ask for guidance, information on a very specific health issue, treatment, and other related topics. Our goal was to explore a method to enhance our ability to collect and interpret information from those social media sources. Our methodology allowed us to organize 4 million plus postings into 30 topics, consequently clustered into 4 groups.

The popularity of QCA-generated categories (as measured by the number of posts in their associated forums) showed a logarithmic-linear (*log-lin)* distribution, strongly suggesting that a few QCA-generated categories are disproportionately gravitated toward user self-selection, while most topics receive comparatively little attention. Moreover, it is of great interest that topic modeling analysis via MALLET showed that the overall LDA strengths of each topic among the forum documents (as seen in Figure 3) also followed a log-*lin* type of distribution, allowing for the same type of conclusion in objectively quantifying content with regards to the MALLET-generated topics.

In addition, a modest level of correlation was observed between the strongest (via MLR analysis) MALLET topics and the strength (by user posting) of manual QCA-generated categories. Topic 11, *looking for support from people in similar circumstances*, is almost semantically identical to the QCA-generated category of *Diagnosed—Psychosocial—Similar Patients.* The latter category encompassed 34.4% (41,972/121,688 of all threads on the Breastcancer.org Community website. Topics 10 and 18 (*Radiation and side effects* and *Chemotherapy side effects and change of treatments*, respectively), meanwhile, correlate in general to the manually generated category of *Diagnosed—Treatment* (top-level domain), which covered 38,698 threads on the site. Topic 7 (*genetic risk and testing*) is semantically similar to the QCA-generated category of *Not Diagnosed But Concerned—Testing*, noted in 4519 site user threads.

The computationally assigned importance of Topic 11 when combined with its equally significant manually generated category correlate demonstrates the need for health personnel to take into account the contextual (nonclinical) factors unlikely to be captured in conventional medical documentation and not supported by conventional clinic-based information technology resources. In particular, a greater emphasis on information-mediated psychosocial interventions is supported by the results of this research.

The computational topic modeling analysis via MALLET also demonstrated topics that did not arise via manual category generation. These topics, in particular Topic 21 (*lingering side effects while in remission)* and Topic 25 (*support from caregiver and medical team for recovery long term*), mirror breast cancer survivorship instead of the disorder itself. Therefore, it is suggested that computational topic modeling software such as MALLET is useful in future research on large bodies of patient-generated text and can generate topics similar in quality to those generated by expert QCA; furthermore, this type of software can detect significant but *hidden* topics (such as social and daily living issues dealt with during survivorship) that are not otherwise detectable when only the forum labels given by a site are analyzed qualitatively.

Visual analysis of the file-file (ie, forum-forum) similarity matrix (see Figure 4) shows a particular concentration of similar files across the diagonal axis, indicating that files numbered with close numbers tended to be more similar in content. This observation actually does strengthen the case for using computational topic modeling software such as MALLET because closely numbered files in the study at hand tended to originate from forums that resided in identical categories on the Breastcancer.org Community website. Similarly, topics that

XSL•FO
RenderX

closely correlated with each other were noted to have clinically significant correlates. It is important to note that many of these correlates may not have been intuitive at face value but were more explainable with the clinical expertise.

Overall, the research team was able to gain significant insight into the daily lives, clinical and otherwise, of patients affected by breast cancer; the onus to support survivors of breast cancer was also revealed. Furthermore, the research performed generated significant support for the use of computational topic modeling software such as MALLET to analyze patient-generated information for nonclinical issues revealed by patients with breast cancer over relevant disease-specific online forums.

Limitations and Future Considerations

The data could be better annotated in metadata-facilitated context as opposed to being in a purely free-text format; the granularity of the ontology in which the data are stored can be improved in future research. In particular, having the posts traceable to unique anonymized users would be of assistance. To achieve this granularity, forums in the future can be scraped in a manner that preserves the HTML source code of their content; structured information could be then extracted from the HTML. The granularity of time, if extractable from the HTML, could potentially facilitate the generation of individual patient records and potentially even allow for the capability for analyzing patient narratives in a longitudinal (time-wise) fashion.

Detailed analysis of similarity measures between files and clustering methods is an important part of potential future research and will require thorough analysis by clinical and patient health experts. This process is very time consuming and is out of the scope of this study as our goal in this work is to present a method for modeling a data in a meaningful format. Future work will focus more on further analysis of data to identify hidden relationships between files and topics that might reveal hidden aspects of breast cancer patients' challenges in their real life.

Conclusions

The importance of patient-generated data (including patient-generated information via online communications) is growing among scholars because of their value in identifying hidden aspects of patients' challenges and concerns. This study provides a reasonable amount of insight into the areas of interest or concern that patients with breast cancer discuss in social media and may need to be addressed to optimize patient disease and health management.

Acknowledgments

The authors would like to acknowledge Breastcancer.org Community (website) as well as the graduate students from the BioHealth Informatics Department who provided support and input in different aspects of the work. This research received no specific grant from any funding agency in public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The 80 forums and their groupings on the Breastcancer.org Community website.

[PDF File (Adobe PDF File), 112KB - medinform_v6i4e45_app1.pdf]

Multimedia Appendix 2

File-feature set.

[PDF File (Adobe PDF File), 107KB - medinform_v6i4e45_app2.pdf]

Multimedia Appendix 3

List of topics generated by MALLET in the 30 topic model.

[PDF File (Adobe PDF File), 117KB - medinform_v6i4e45_app3.pdf]

References

1. Esquivel A, Meric-Bernstam F, Bernstam EV. Accuracy and self correction of information received from an internet breast cancer list: content analysis. BMJ 2006 Apr 22;332(7547):939-942 [FREE Full text] [doi: 10.1136/bmj.38753.524201.7C] [Medline: 16513686]
2. Zhang S, Lin Q, Chen F, Zhang W, Yu Y, Elhadad N. We Make Choices We Think Are Going to Save Us: Debate and Stance Identification for Online Breast Cancer. In: Proceedings of 26th International Conference on World Wide Web Companion. We make choices we think are going to save us?: Debate and stance identification for online breast CAM

XSL•FO
RenderX

discussion. Proc Int World Wide Web Conf. 2017 Apr; 2017 Presented at: 26th International Conference on World Wide Web; 2017 April 3; Perth, Australia p. 1073-1081.
3. Office of the National Coordinator (ONC) for Health Information Technology. HealthIT.gov. Washington, DC, USA: United States Department of Health and Human Services; 2018 Mar 21. Patient-Generated Health Data URL: https://www.healthit.gov/policy-researchers-implementers/patient-generated-health-data [accessed 2018-03-02] [WebCite Cache ID 6xcjfJpno]
4. Kostick KM, Blumenthal-Barby JS, Wilhelms LA, Delgado ED, Bruce CR. Content Analysis of Social Media Related to Left Ventricular Assist Devices. Circ Cardiovasc Qual Outcomes 2015 Sep;8(5):517-523. [doi: 10.1161/CIRCOUTCOMES.115.002032] [Medline: 26219889]
5. Asch DA, Rader DJ, Merchant RM. Mining the social mediome. Trends Mol Med 2015 Sep;21(9):528-529 [FREE Full text] [doi: 10.1016/j.molmed.2015.06.004] [Medline: 26341614]
6. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. Journal of Medical Internet Research 2015;17(7):e188.
7. Breastcancer. URL: https://community.breastcancer.org/ [accessed 2018-04-19] [WebCite Cache ID 6yna7qQb6]
8. Breast Cancer Care. Breast Cancer Care. 2018. Breast Cancer Care Forums URL: http://forums.breastcancercare.co.uk/ [accessed 2018-04-19] [WebCite Cache ID 6ynb1syFD]
9. Susan G Komen Foundation. Susan G Komen Foundation (Website). 2018. Message Boards URL: https://apps.komen.org/forums/ [accessed 2018-04-19] [WebCite Cache ID 6ynephq1I]
10. Triple Negative Breast Cancer Foundation. Triple Negative Breast Cancer Foundation (Website). 2018. Triple Negative Forums URL: http://forum.tnbcfoundation.org/ [accessed 2018-04-19] [WebCite Cache ID 6ynbiJjXf]
11. No Surrender Breast Cancer Foundation. Website. 2018. No Surrender Breast Cancer Support Forum URL: http://nosurrenderbreastcancersupportforum.com/ [accessed 2018-04-19] [WebCite Cache ID 6yne8PLvp]
12. Institute of Medicine (National Academies of Science). Capturing Social and Behavioral Domains and Measures in Electronic Health Records (Phase 2). Washington, DC, USA: National Academies Press (USA); Jan 08, 2015.
13. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research 2003;2003(3):993-1022.
14. McCallum AK. MALLET: A Machine Learning Language Toolkit (Software Tool). Amherst, MA, USA: University of Massachusetts; 2002. URL: http://mallet.cs.umass.edu/ [accessed 2017-04-17] [WebCite Cache ID 6xcjxfqfW]
15. R Development Core Team, 2008-2017 (A2. R: A language and environment for statistical computing [Software tool]. 2008. URL: http://www.r-project.org/ [accessed 2017-04-17] [WebCite Cache ID 6xck6uupv]

Abbreviations

EDSM: Euclidean Distance Similarity Measures
LDA: Latent Dirichlet allocation
MALLET: Machine Learning Language Toolkit
MLR: multiple linear regression
NLP: natural language processing
QCA: qualitative content analysis

Edited by G Eysenbach; submitted 11.10.17; peer-reviewed by D Attai, D Hafez; comments to author 23.11.17; revised version received 14.03.18; accepted 21.06.18; published 29.11.18.

Please cite as:
Jones J, Pradhan M, Hosseini M, Kulanthaivel A, Hosseini M
Novel Approach to Cluster Patient-Generated Data Into Actionable Topics: Case Study of a Web-Based Breast Cancer Forum
JMIR Med Inform 2018;6(4):e45
URL: http://medinform.jmir.org/2018/4/e45/
doi:10.2196/medinform.9162
PMID:30497991

©Josette Jones, Meeta Pradhan, Masoud Hosseini, Anand Kulanthaivel, Mahmood Hosseini. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 29.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Extraction of Information Related to Adverse Drug Events from Electronic Health Record Notes: Design of an End-to-End Model Based on Deep Learning

Fei Li[1,2,3], PhD; Weisong Liu[1,2,3], PhD; Hong Yu[1,2,3,4], PhD

[1]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States

[2]Center for Healthcare Organization and Implementation Research, Bedford Veterans Affairs Medical Center, Bedford, MA, United States

[3]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States

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

Corresponding Author:
Hong Yu, PhD
Department of Computer Science
University of Massachusetts Lowell
1 University Avenue
Lowell, MA,
United States
Phone: 1 508 612 7292
Email: Hong_Yu@uml.edu

Abstract

Background: Pharmacovigilance and drug-safety surveillance are crucial for monitoring adverse drug events (ADEs), but the main ADE-reporting systems such as Food and Drug Administration Adverse Event Reporting System face challenges such as underreporting. Therefore, as complementary surveillance, data on ADEs are extracted from electronic health record (EHR) notes via natural language processing (NLP). As NLP develops, many up-to-date machine-learning techniques are introduced in this field, such as deep learning and multi-task learning (MTL). However, only a few studies have focused on employing such techniques to extract ADEs.

Objective: We aimed to design a deep learning model for extracting ADEs and related information such as medications and indications. Since extraction of ADE-related information includes two steps—named entity recognition and relation extraction—our second objective was to improve the deep learning model using multi-task learning between the two steps.

Methods: We employed the dataset from the Medication, Indication and Adverse Drug Events (MADE) 1.0 challenge to train and test our models. This dataset consists of 1089 EHR notes of cancer patients and includes 9 entity types such as Medication, Indication, and ADE and 7 types of relations between these entities. To extract information from the dataset, we proposed a deep-learning model that uses a bidirectional long short-term memory (BiLSTM) conditional random field network to recognize entities and a BiLSTM-Attention network to extract relations. To further improve the deep-learning model, we employed three typical MTL methods, namely, hard parameter sharing, parameter regularization, and task relation learning, to build three MTL models, called HardMTL, RegMTL, and LearnMTL, respectively.

Results: Since extraction of ADE-related information is a two-step task, the result of the second step (ie, relation extraction) was used to compare all models. We used microaveraged precision, recall, and F1 as evaluation metrics. Our deep learning model achieved state-of-the-art results (F1=65.9%), which is significantly higher than that (F1=61.7%) of the best system in the MADE1.0 challenge. HardMTL further improved the F1 by 0.8%, boosting the F1 to 66.7%, whereas RegMTL and LearnMTL failed to boost the performance.

Conclusions: Deep learning models can significantly improve the performance of ADE-related information extraction. MTL may be effective for named entity recognition and relation extraction, but it depends on the methods, data, and other factors. Our results can facilitate research on ADE detection, NLP, and machine learning.

(JMIR Med Inform 2018;6(4):e12159) doi:10.2196/12159

XSL•FO
RenderX

KEYWORDS

adverse drug event; deep learning; multi-task learning; named entity recognition; natural language processing; relation extraction

Introduction

Background

An adverse drug event (ADE) is an injury resulting from a medical drug intervention [1]. Previous studies reported that ADEs could account for up to 41% of all hospital admissions [2,3]. An ADE may cause a prolonged length of stay in the hospital and increase the economic burden [4]. The annual cost of ADEs for a 700-bed hospital is approximately $5.6 million [5]. Moreover, the total number of iatrogenic deaths can reach nearly 800,000 per year, which is higher than the death rate of heart disease or cancer [6]. In 2013, medical error, including ADEs, is the third most-common cause of death in the United States [7]. Therefore, ADE detection and report are crucial for pharmacovigilance and drug-safety surveillance [8,9].

Two methods are usually used to detect and report ADE. In premarketing surveillance, ADEs can be discovered during phase III clinical trials for drug development. In postmarketing surveillance, ADEs are discovered by patients and physicians using the Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS). These traditional methods are limited by the number of participants [10], underreporting [11], and missing patterns of drug exposure [12]; for example, underreporting is a known issue in FAERS and may occur due to several reasons. First, the objective and content of the report in FAERS change over time, which may confuse physicians and the general public. Second, patients may choose not to mention some reactions, due to which practitioners fail to report them. Third, ADEs with long latency or producing unusual symptoms may be unrecognized. Other reasons may include payments from pharmaceutical companies to doctors [13] and inefficient communication between patients, physicians, and pharmacists. Recently, the FDA made the FAERS data available through a public dashboard [14]. Since anyone can view ADE reports online, this venture may help the FDA receive feedback to improve the FAERS.

Many researchers have used other resources to identify ADEs, such as biomedical publications [15,16], social media [17,18], and electronic health record (EHR) notes [19-21]. The ADEs extracted from these resources are an important complement to traditional ADE-surveillance systems. However, manual collection of ADEs from these data is laborious [22]. As such, the use of computer systems is a good choice to automatically detect ADEs, but may fail since these data are often unstructured text. Therefore, natural language processing (NLP) techniques are employed for this significant task [15,20,21,23].

From the viewpoint of NLP, ADE detection is covered under the task of information extraction, which includes ADE extraction as well as information related to ADE, such as medications and indications. Extraction of ADE-related information can be casted as a two-step pipeline. The first step is named entity recognition (NER) [24], which recognizes a string of text as an entity (eg, medication or ADE) that is predefined by dataset annotators. The second step is relation extraction (RE) [15], which is a model that determines whether two entities have a specific relation (eg, medication and ADE).

Previous studies employed traditional machine-learning techniques [15,16,23,24] such as condition random field (CRF) [25] or support vector machine (SVM) [26]. Recently, deep learning attracted much attention in NLP due its numerous advantages such as better performances and less feature engineering compared to other systems [27,28]. However, only a few studies have addressed extraction of ADE-related information via deep learning. Since ADE detection can be divided into two tasks (ie, NER and RE), it is logical to incorporate multi-task learning (MTL) [29] into ADE detection. However, only limited prior work has investigated the impact of MTL on ADE detection.

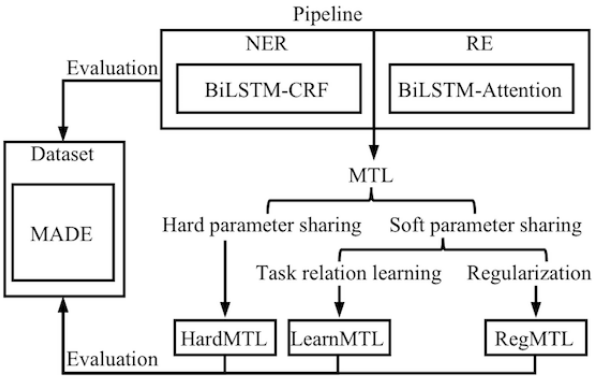
Relevant Literature

Adverse Drug Event Detection

Since ADEs play an important role in drug-safety surveillance, ADE detection receives increasing attention from both the federal regulation and the research community. Besides the official reporting system FAERS, there are other databases that collect data on known ADEs, such as the Comparative Toxicogenomics Database [30] and SIDER [31]. Various resources have been used to detect ADEs, such as biomedical publications [15,16], social media [17,18], and electronic health record (EHR) notes [19-21]. For example, Gurulingappa et al [16] leveraged medical case reports to build a corpus in order to support drug-related adverse effects. Wei et al [15] organized a challenge task to extract chemical-induced disease relations from the literature and created an annotated corpus from 1500 articles. With respect to the methods, both supervised and unsupervised methods were used. Ramesh et al [32] developed a supervised machine-learning model to extract adverse event entities from FAERS narratives. Xu and Wang [33] used a semisupervised bootstrapped method to construct a knowledge base for the drug-side-effect association. Liu et al [34] proposed a causality-analysis model based on structure learning for identifying factors that contribute to adverse drug reactions. Yildirim et al [35] applied the k-mean algorithm to identify adverse reactions. Xu et al [23] used SVM to extract ADEs between sentence-level and document-level drug-disease pairs. Recently, Munkhdalai et al [21] attempted to use deep learning to address ADE extraction, but their model was not end-to-end and relied on the entities.

XSL•FO
RenderX

Figure 1. Study overview. NER: named entity recognition. RE: relation extraction. BiLSTM: bidirectional long short-term-memory. CRF: conditional random field. MTL: multi-task learning. MADE: Medication, Indication, and Adverse Drug Events. HardMTL: multi-task learning model for hard parameter sharing. RegMTL: multi-task learning model for soft parameter sharing based on regularization. LearnMTL: multi-task learning model for soft parameter sharing based on task-relation learning.



Named Entity Recognition

NER is a standard task that has been studied for many years in NLP [25]. Many researchers made important contributions to dataset construction including the GENIA corpus [36], the NCBI disease corpus [37], and the ShARe/CLEF eHealth evaluation [38]. Early studies addressed NER by diverse approaches such as rule-based [39] and machine-learning approaches [40-42], among which CRF-based approaches predominated. For example, Campos et al [40] presented a CRF model to recognize biomedical names, which achieved state-of-the-art performance at the time by incorporating rich features. Tang et al [43] modified the label scheme of CRF to make it be capable of recognizing disjoint clinical concepts. However, such approaches need many efforts for feature engineering. Recently, a bidirectional LSTM (BiLSTM) model [44,45] was proposed and became a popular method for NER. In the biomedical domain, Jagannatha and Yu [20] used such a model to detect medical events from EHR notes.

Relation Extraction

RE has been widely studied, and some typical RE corpora in the biomedical domain include the 2010 i2b2/VA challenge [46] and BioCreative V CDR task [15]. Early work used rules and NLP structures such as dependency trees [47] and coreference chains [48] to help extract relations. Others usually leveraged machine-learning approaches such as SVM [49,50] and structured learning [51]. As deep learning developed, researchers proposed a number of neural network models to handle RE [52,53]. Our study is related to the joint or end-to-end entity and RE, which allows performance of NER and RE simultaneously. Miwa and Bansal [54] proposed an end-to-end model based on the sequence and tree LSTM. Similarly, Mehryary et al [55] proposed an end-to-end system to extract information about bacteria and their habitats.

Multi-Task Learning

MTL [29] refers to training the model for multiple related tasks. It is widely used in artificial intelligence research such as computer vision [56] and NLP [57]. Learning these tasks simultaneously may improve the performance as compared to learning the tasks individually. Prior MTL studies mostly focused on homogeneous MTL that consists of tasks with only one type such as classification or regression [58]. Some of their tasks are closely related, such as cross-lingual [59] and synthetic data [60]. Based on a previous study [58], MTL can be roughly divided into two categories—hard and soft parameter sharing. For hard parameter sharing, the lower layers are shared among multiple tasks and each task has its own higher layer [54]. For soft parameter sharing, each task has its own model with its own parameters. There are some representative methods for soft parameter sharing such as regularization [59] or learning task relations [56].

Objective

Since only a few prior studies have addressed end-to-end detection of ADE via deep learning, we aimed to design a two-step pipeline model that consists of two submodels: a BiLSTM [61] CRF [25] network for NER and a BiLSTM-Attention [62] network for RE. Since extraction of ADE-related information includes two steps, it is possible to study the impact of MTL on NER and RE. However, only limited prior work has focused on MTL with such heterogeneous and loosely related tasks. Therefore, our second objective was to fill this gap by proposing three MTL models and comparing them with the pipeline model. An overview of this study is shown in Figure 1.

Methods

Deep Learning Pipeline Model

BiLSTM-CRF Submodel for NER

Our NER submodel is presented in Figure 2. We extended the state-of-the-art BiLSTM-CRF model [44,45] by enriching its features. For a sentence, we extracted four kinds of features for each token, namely, its word, whether the initial character is capital, its part-of-speech (POS) tag, and its character representation. We employed a convolutional neural network to obtain character representation. After the token representations are obtained by concatenating the features, we fed them into a bidirectional LSTM layer to learn the hidden representations. Subsequently, the hidden representations were input into the CRF layer to determine the optimal labels for all

XSL•FO
RenderX

the tokens in the sentence. For labels, we use the BMES (Begin, Middle, End, Singular) label scheme [45] plus entity types. For example, the label of the token "Renal" is "B_Disease." The details of the BiLSTM-CRF submodel for NER are provided in Multimedia Appendix 1.

BiLSTM-Attention Submodel for RE

Our RE submodel is presented in Figure 3. A relation instance can be considered as a token sequence and two target entities. Here, the token sequence did not necessarily have to be one sentence, as we could also extract intersentence relations. For each token, we extracted four kinds of features, namely, its word, its POS tag, and the position embeddings [63]. Here, the character representation was not used, because it reduced the performance in our preliminary experiments. Similar to the case for NER, we employed a BiLSTM layer to generate the hidden representations. Subsequently, the attention method [62] was used to obtain context features.

Because only context features may not be enough to capture the semantic relation, we also employed other features for concision, which are not shown in Figure 3. Considering previous work [21], these features included words of two target entities, types of two target entities, the token number between two target entities, and the entity number between two target entities. Like the word or POS embeddings, these features can be represented as vectors. Therefore, the output layer considers the concatenation of all these features to determine the relation of target entities. The details of the BiLSTM-Attention submodel for RE are provided in Multimedia Appendix 2.

Multi-Task Learning Models

In this section, we propose three MTL models: one model used hard parameter sharing [54] and two models used soft parameter sharing, namely, regularization [59] and task relation learning [56].

HardMTL

Our MTL model for hard parameter sharing is presented in Figure 4. We employed the shared-private architecture [64] to make each submodel of each task retain its private parts and share some parts for multi-task learning. The NER and RE submodels had their own BiLSTM layers, namely, $LSTM^{ner}$ and $LSTM^{re}$, and shared a BiLSTM layer, $LSTM^{share}$. During training, the shared BiLSTM layer $LSTM^{share}$ was used by both the NER and RE submodels, so that it was tuned during the back-propagation by both submodels. Therefore, the model was able to learn useful knowledge from both tasks. The details of the HardMTL model are provided in Multimedia Appendix 3.

RegMTL

Our first MTL model for soft parameter sharing was based on regularization, and its architecture is presented in Figure 5. With reference to previous studies [59,60], we employed the L2 regularization in order to encourage the parameters of the NER and RE submodels to be similar instead of sharing some parts in the networks. Two BiLSTM layers were considered because different inputs of the NER and RE submodels lead to different dimensions of the first BiLSTM layer; therefore, L2 regularization of the parameters of the first BiLSTM layer was computationally intractable. We resolved this issue by performing L2 regularization in the second BiLSTM layer. The details of the RegMTL model are provided in Multimedia Appendix 3.

LearnMTL

Our second MTL model for soft parameter sharing was based on task relation learning [56], and its architecture is illustrated in Figure 6. After generating hidden representations from the BiLSTM and attention layers, we used a linear layer, W_5, to exchange information between the NER and RE submodels. To utilize task-specific and shared information, the concatenation of hidden representations of the BiLSTM and information exchange layers was fed into the upper decoders D^{ner} and D^{re}. The details of the LearnMTL model are provided in Multimedia Appendix 3.

Dataset

We used the MADE dataset from the MADE1.0 challenge for detecting medications and ADEs from EHR notes [65]. It consists of 1089 EHR notes of patients with cancer, from which data for 18 common Protected Health Information aspects were removed according to the Health Insurance Portability and Accountability Act. The dataset was separated into 876 notes for training and 213 notes for testing. In this dataset, the annotators annotated not only ADEs, but also other ADE-related information. They predefined 9 entity types, namely, Medication, Indication, Frequency, Severity, Dosage, Duration, Route, ADE, and SSLIF (any sign, symptom, and disease that is not an ADE or Indication). In addition, they predefined 7 relation types between these entity types, namely, Dosage-Medication, Route-Medication, Frequency-Medication, Duration-Medication, Medication-Indication, Medication-ADE, and Severity-ADE.

XSL•FO
RenderX

Figure 2. NER submodel. For simplicity, here we use “Renal Failure” to illustrate the architecture. For “Renal,” the word feature is “Renal,” the capital feature of the initial character is “R,” the POS feature is “JJ,” and the character representation is generated from CNN. NER: named entity recognition. CNN: convolutional neural network. CRF: condition random field. LSTM: long short-term memory. CNN: convolutional neural network. POS: part of speech.

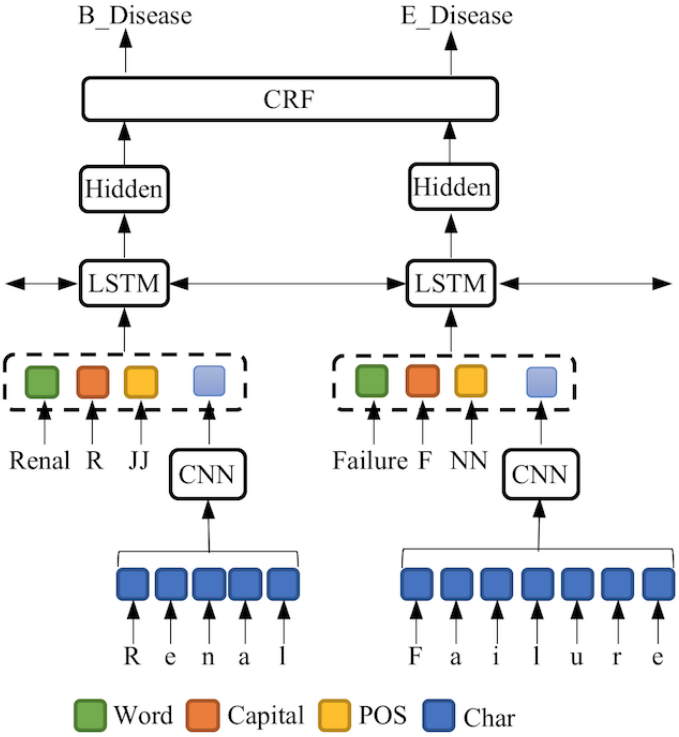
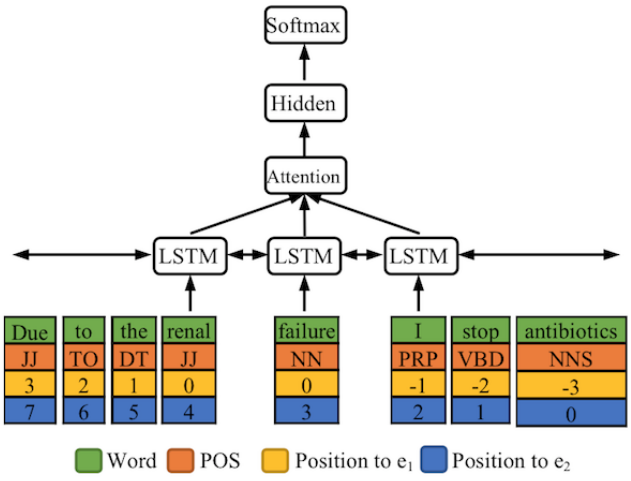


Figure 3. RE submodel. The target entities are “renal failure” (e_1) and “antibiotics” (e_2). Positions represent token distances to the target entities. RE: relation extraction. LSTM: long short-term memory. POS: part of speech.



XSL•FO
RenderX

Figure 4. The high-level view of HardMTL. For conciseness, "LSTM" indicates a BiLSTM layer, and the layers above the BiLSTM layer are denoted as D^{new} and D^{re}. The forward procedures for an NER instance and an RE instance are indicated by blue and green arrow lines, respectively. HardMTL: multi-task learning model for hard parameter sharing. LSTM: long short-term-memory. BiLSTM: bidirectional long short-term-memory. CRF: conditional random field. NER: named entity recognition. RE: relation extraction.

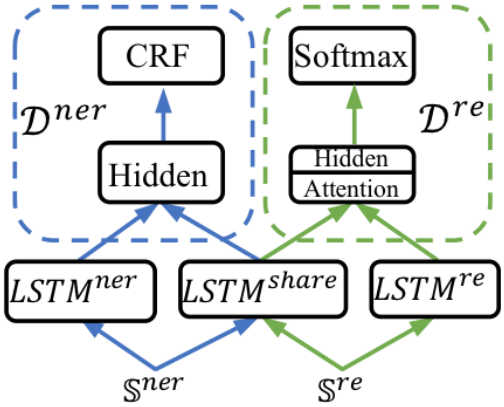
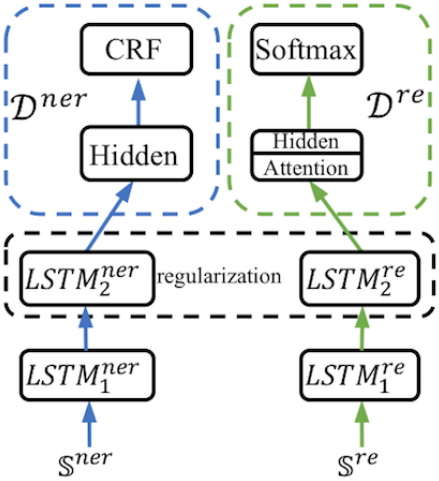
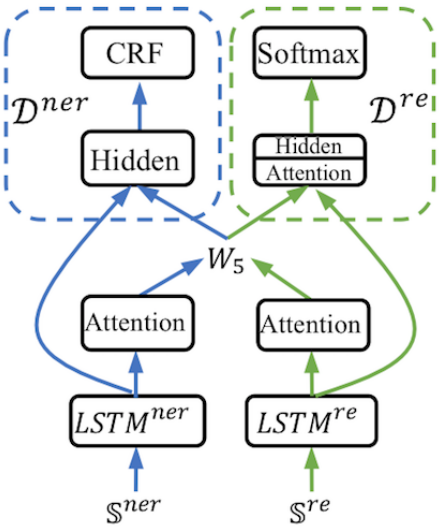


Figure 5. The high-level view of RegMTL. $LSTM_1^{ner}$ and $LSTM_2^{ner}$ indicate the first and second BiLSTM layers of the NER model. $LSTM_1^{re}$ and $LSTM_2^{re}$ indicate the first and second BiLSTM layers of the RE model. NER: named entity recognition. RE: relation extraction. RegMTL: multi-task learning model for soft parameter sharing based on regularization. BiLSTM: bidirectional long short-term-memory. CRF: conditional random field. LSTM: long short-term-memory.



XSL•FO
RenderX

Figure 6. The high-level view of LearnMTL. LearnMTL: multi-task learning model for soft parameter sharing based on task-relation learning. CRF: conditional random field. LSTM: long short-term-memory.



Results

The experimental settings used to obtain these results are provided in Multimedia Appendix 4.

Comparison Between Our Best Model and Existing Systems

We compared our models with the top three systems in the MADE1.0 challenge. Chapman et al [66] used CRF for NER and random forest for RE. Specifically, two random forest models were used—one for detecting whether relations exist between entities and the other for classifying what specific relation type exists. Xu et al [67] used BiLSTM-CRF for NER with word, prefix, suffix, and character features. For RE, they used SVM and designed features such as positions, distances, bag of words, and bag of entities. Dandala et al [68] also used BiLSTM-CRF for NER, but they input different features into the model such as words, POS tags, and characters. For RE, they employed the BiLSTM-Attention model that takes tokens, entity types, and positions as input.

Full neural systems ([68] and our study) achieve better performances with the MADE dataset than with other systems (Table 1). Although the main methods between the study of Dandala et al [68] and our study are similar, our model is much better, as it significantly improved the F1 for RE by 5%. The reasons for this superiority may be that we used more features than previous work, such as capital information and entity words, and our model attained approximately 0.8% improvement in F1 from MTL.

Comparison Between the Pipeline and MTL Models

The HardMTL model outperforms other models, achieving an F1 of 84.5% in NER and 66.7% in RE (Table 2); the pipeline model ranks second, with F1 values of 84.1% and 65.9%, respectively. The RegMTL model obtains the best recall in both NER (84.5%) and RE (63.6%). Surprisingly, the most-complex MTL model LearnMTL ranked last.

In our experiments, HardMTL successfully boosted the NER F1 by 0.4% (P=.003) and the RE F1 by 0.8% (P=.01), but RegMTL and LearnMTL failed to boost the performances. Thus, the effectiveness of different MTL methods depends on the selected tasks and data. For heterogenous and loosely related tasks such as NER and RE, it is more difficult for MTL to be effective.

Performance of Each Entity Type

Table 3 shows the performance of each entity type. Medication and Route (both F1>90%) were easier to recognize than other types. In contrast, ADE is the most-difficult type to recognize (F1=55%). Other entity types with lower performances included Indication and Duration.

Performance of Each Relation Type

Table 4 shows the performance of each relation type. Medication-ADE relations are the most-difficult type to extract (F1=45.5%). Severity-ADE ranks second (F1=54.1%), followed by Duration-Medication (F1=59.5%). In contrast, Route-Medication and Dosage-Medication extraction are relatively easier, with F1>80%.

Comparison Between the Pipeline Model and MedEx System

MedEx [69] is an end-to-end system used to identify medications and their attributes such as routes and dosages. Therefore, the final results of MedEx correspond to our results for extracting 4 kinds of relations: Route-Medication, Dosage-Medication, Duration-Medication, Frequency-Medication. Table 5 compares MedEx with our model. Our model significantly outperformed MedEx, which demonstrates that our model is a competitive system in this domain.

XSL•FO
RenderX

Table 1. Comparison of our model with the existing systems in the Medication, Indication, and Adverse Drug Events dataset. The microaveraged F1s of relation extraction are shown according to the official evaluation report.

System	Named entity recognition	Relation extraction	F1
Chapman et al [66]	CRF[a]	Random forest	59.2
Xu et al [67]	BiLSTM[b]-CRF	Support vector machine	59.9
Dandala et al [68]	BiLSTM-CRF	BiLSTM-Attention	61.7
Our Best (HardMTL[c])	BiLSTM-CRF	BiLSTM-Attention	66.7

[a]CRF: conditional random field

[b]BiLSTM: bidirectional long short-term memory

[c]HardMTL: multi-task learning model for hard parameter sharing

Table 2. Performances (%) of the pipeline and multi-task learning models. The values presented are the means of 5 runs of each model. The microaveraged P, R, and F1s of all entity or relation types are shown.

Method	Entity recognition			Relation extraction		
	P	R	F1	P	R	F1
Pipeline	85.0	83.2	84.1	69.8	62.4	65.9
HardMTL[a]	85.0	84.1	84.5	70.2	63.6	66.7
RegMTL[b]	84.5	84.5	84.5	66.7	63.6	65.1
LearnMTL[c]	84.5	82.8	83.6	67.2	61.5	64.2

[a]HardMTL: multi-task learning model for hard parameter sharing

[b]RegMTL: multi-task learning model for soft parameter sharing based on regularization

[c]LearnMTL: multi-task learning model for soft parameter sharing based on task relation learning

Table 3. Performance (%) of each entity type.

Entity type	P	R	F1
Medication	91.1	92.0	91.3
Indication	65.4	64.8	64.8
Frequency	87.1	86.5	86.3
Severity	84.6	84.7	84.7
Dosage	87.9	86.4	88.0
Duration	75.3	76.6	77.6
Route	91.6	91.9	91.9
Adverse drug events	59.5	57.6	55.4
SSLIF[a]	83.9	84.8	84.9

[a]SSLIF: any sign, symptom, and disease that is not an ADE or Indication

Table 4. Performance (%) of each relation type.

Relation type	P	R	F1
Severity-Adverse drug events	55.0	54.4	54.1
Route-Medication	81.0	82.5	82.1
Medication-Indication	53.9	52.5	52.9
Dosage-Medication	80.9	79.8	81.0
Duration-Medication	60.3	63.7	59.5
Frequency-Medication	77.7	78.6	78.4
Medication-Adverse drug events	50.4	47.6	45.5

XSL•FO
RenderX

Table 5. Results (%) of comparisons between our pipeline model and the MedEx system.

Entity type	MedEx system			Pipeline model		
	P	R	F1	P	R	F1
Route-Medication	71.9	47.9	57.5	81.0	82.5	82.1
Dosage-Medication	29.7	3.5	6.2	80.9	79.8	81.0
Duration-Medication	25.5	15.6	19.4	60.3	63.7	59.5
Frequency-Medication	52.5	36.2	42.8	77.7	78.6	78.4

Discussion

Principal Findings

Existing systems usually selected a two-step pipeline to address ADE-related information extraction: recognizing entities and extracting relations. BiLSTM-CRF is the most-popular model for NER, whereas the selections of RE models are mixed. All our models outperformed the existing systems in the MADE1.0 challenge, which may be because of the following reasons: First, our models benefited from deep learning that is able to learn better from the data. Second, we enriched the features of deep learning models; therefore, our model outperformed the system [68] that used similar deep learning models as ours.

For MTL, we found that the model using hard parameter sharing (HardMTL) performed better than the other two models using soft parameter sharing (RegMTL and LearnMTL) and that the most complex MTL model, LearnMTL, performed the worst in our data. Our results are not surprising, as different MTL methods depend on different tasks and data [54,56,59]. Overall, MTL more difficult between heterogeneous and loosely related tasks such as NER and RE.

In our experiments, the entity type "ADE" and relation type "Medication-ADE" were the most difficult information to be extracted. Based on our analysis, this is not only due to a lack of training data, but also the intrinsic character of ADEs. ADEs are often implicit in the context without any obvious pattern, which negatively affects the model (Example 1 in Multimedia Appendix 5). In contrast, some entity or relation types with obvious patterns (eg, Medication-Dosage) are easier to identify (Example 2 in Multimedia Appendix 5).

Finally, we found that the performance improved when we used the pretrained word embeddings in the biomedical domain [70] rather than those in the general domain. Furthermore, if the pretrained word embeddings were not tuned, our models would perform better. One likely reason for this is that such a method can alleviate the overfitting problem.

Error Analysis

We randomly sampled hundreds of error instances of NER and RE. Through the manual analyses, we found several sources of errors. For NER, the major false-negative errors are due to long expressions of entities (Examples 3 and 4 in Multimedia Appendix 5). These entities (eg, IgG kappa monoclonal protein) include multiple words; therefore, it is difficult to detect their boundaries. Moreover, the major false-positive errors for NER occur because some entity types are incorrectly recognized as SSLIF (Examples 5 and 6 in Multimedia Appendix 5). This may be because the training instances of SSLIF are ≥10 times those of other entity types such as ADE. Thus, imbalanced data distribution may lead to certain bias of our models.

With respect to RE, the major false-negative errors are due to long distances between target entities (Example 9 in Multimedia Appendix 5). The relation of two entities can be expressed through ≥6 sentences in EHRs; therefore, our model may miss such relations in a long context. In addition, the major false-positive errors for RE occur because relation expressions exist in the instance, but are not related to the target entities (Examples 7 and 8 in Multimedia Appendix 5). For instance, in Example 7 of Multimedia Appendix 5, "His current therapy includes [thalidomide]$_{Entity1}$ 50 mg a day for 2 weeks out of the month. He had been on Velcade, which was stopped secondary to increasing [peripheral neuropathy]$_{Entity2}$, " "peripheral neuropathy," and "thalidomide" have no Medication-ADE relation, but the model incorrectly predicts their relation due to the words "secondary to."

Contributions

The main contributions of this work are as follows: (1) We proposed an up-to-date deep learning model to perform ADE-related information extraction in an end-to-end manner. Our model achieved new state-of-the-art performance, improving the F1 by 4.2% (absolute value). (2) To our knowledge, this is the first attempt to investigate the impact of MTL on two heterogeneous and loosely related tasks (ie, NER and RE). One of our MTL models further improved the F1 by 0.8% (absolute value). (3) Our manually annotated dataset—Medication, Indication, and Adverse Drug Events (MADE) [65]—will be publicly available to support the research on extraction of ADE-related information.

Conclusions

We proposed a deep learning model to detect ADEs and related information. We also investigated MTL on two ADE-related tasks, NER and RE. Our models achieved state-of-the-art performance in an ADE-detection dataset. MTL can improve performance, but it depends on the methods and data used. In the future, we plan to evaluate our models with more related datasets.

XSL•FO
RenderX

Acknowledgments

This work was supported by a grant (5R01HL125089) from the National Institutes of Health and the Investigator Initiated Research grant (1I01HX001457-01) from the Health Services Research & Development Program of the US Department of Veterans Affairs.

Conflicts of Interest

None declared.

Multimedia Appendix 1

BiLSTM-CRF submodel for NER. BiLSTM: bidirectional long short-term-memory. CRF: conditional random field. NER: named entity recognition.

[PDF File (Adobe PDF File), 75KB - medinform_v6i4e12159_app1.pdf]

Multimedia Appendix 2

BiLSTM-Attention submodel for RE. BiLSTM: bidirectional long short-term-memory. RE: relation extraction.

[PDF File (Adobe PDF File), 69KB - medinform_v6i4e12159_app2.pdf]

Multimedia Appendix 3

MTL models. MTL: multi-task learning.

[PDF File (Adobe PDF File), 181KB - medinform_v6i4e12159_app3.pdf]

Multimedia Appendix 4

Experimental settings.

[PDF File (Adobe PDF File), 73KB - medinform_v6i4e12159_app4.pdf]

Multimedia Appendix 5

Examples.

[PDF File (Adobe PDF File), 22KB - medinform_v6i4e12159_app5.pdf]

References

1. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA 1995 Jul 05;274(1):29-34. [Medline: 7791255]
2. Nebeker JR, Hoffman JM, Weir CR, Bennett CL, Hurdle JF. High rates of adverse drug events in a highly computerized hospital. Arch Intern Med 2005 May 23;165(10):1111-1116. [doi: 10.1001/archinte.165.10.1111] [Medline: 15911723]
3. Fattinger K, Roos M, Vergères P, Holenstein C, Kind B, Masche U, et al. Epidemiology of drug exposure and adverse drug reactions in two swiss departments of internal medicine. Br J Clin Pharmacol 2000 Feb;49(2):158-167 [FREE Full text] [Medline: 10671911]
4. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. JAMA 1997;277(4):301-306. [Medline: 9002492]
5. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA 1997;277(4):307-311. [Medline: 9002493]
6. Null G, Dean C, Feldman M, Rasio D. Death by Medicine. Journal of Orthomolecular Medicine 2005;20(1):34 [FREE Full text]
7. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ 2016 Dec 03. [Medline: 27143499]
8. Handler SM, Altman RL, Perera S, Hanlon JT, Studenski SA, Bost JE, et al. A systematic review of the performance characteristics of clinical event monitor signals used to detect adverse drug events in the hospital setting. J Am Med Inform Assoc 2007;14(4):451-458 [FREE Full text] [doi: 10.1197/jamia.M2369] [Medline: 17460130]
9. Kaushal R, Jha AK, Franz C, Glaser J, Shetty KD, Jaggi T, BrighamWomen's Hospital CPOE Working Group. Return on investment for a computerized physician order entry system. J Am Med Inform Assoc 2006 May;13(3):261-266 [FREE Full text] [doi: 10.1197/jamia.M1984] [Medline: 16501178]
10. Haas JS, Iyer A, Orav EJ, Schiff GD, Bates DW. Participation in an ambulatory e-pharmacovigilance system. Pharmacoepidemiol Drug Saf 2010 Sep;19(9):961-969. [doi: 10.1002/pds.2006] [Medline: 20623512]

XSL•FO
RenderX

11. Edlavitch SA. Adverse Drug Event Reporting. Arch Intern Med 1988 Jul 01;148(7):1499. [doi: 10.1001/archinte.1988.00380070017005]
12. Begaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? British Journal of Clinical Pharmacology 2012 Jul 05;38(5):401-404. [doi: 10.1111/j.1365-2125.1994.tb04373.x]
13. ProPublica. Dollars for Docs URL: https://projects.propublica.org/docdollars/ [accessed 2018-10-28] [WebCite Cache ID 73UddA54z]
14. Kumar A. The Newly Available FAERS Public Dashboard: Implications for Health Care Professionals. Hospital Pharmacy 2018 Aug 23. [doi: 10.1177/0018578718795271]
15. Wei C, Peng Y, Leaman R, Davis AP, Mattingly CJ, Li J, et al. Assessing the state of the art in biomedical relation extraction: overview of the BioCreative V chemical-disease relation (CDR) task. Database (Oxford) 2016 [FREE Full text] [doi: 10.1093/database/baw032] [Medline: 26994911]
16. Gurulingappa H, Rajput AM, Roberts A, Fluck J, Hofmann-Apitius M, Toldo L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. J Biomed Inform 2012 Oct;45(5):885-892 [FREE Full text] [doi: 10.1016/j.jbi.2012.04.008] [Medline: 22554702]
17. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
18. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]
19. Gurwitz JH. Incidence and Preventability of Adverse Drug Events Among Older Persons in the Ambulatory Setting. JAMA 2003 Mar 05;289(9):1107. [doi: 10.1001/jama.289.9.1107]
20. Jagannatha AN, Yu H. Bidirectional RNN for Medical Event Detection in Electronic Health Records. Proc Conf 2016 Jun;2016:473-482 [FREE Full text] [Medline: 27885364]
21. Munkhdalai T, Liu F, Yu H. Clinical Relation Extraction Toward Drug Safety Surveillance Using Electronic Health Record Narratives: Classical Learning Versus Deep Learning. JMIR Public Health Surveill 2018 Apr 25;4(2):e29 [FREE Full text] [doi: 10.2196/publichealth.9361] [Medline: 29695376]
22. Hurdle JF, Weir CR, Roth B, Hoffman J, Nebeker JR. Critical gaps in the world's largest electronic medical record: Ad Hoc nursing narratives and invisible adverse drug events. AMIA Annu Symp Proc 2003:309-312 [FREE Full text] [Medline: 14728184]
23. Xu J, Wu Y, Zhang Y, Wang J, Lee H, Xu H. CD-REST: a system for extracting chemical-induced disease relation in literature. Database (Oxford) 2016 May 25. [doi: 10.1093/database/baw036] [Medline: 27016700]
24. Finkel J, Dingare S, Manning CD, Nissim M, Alex B, Grover C. Exploring the boundaries: gene and protein identification in biomedical text. BMC Bioinformatics 2005;6 Suppl 1:S5 [FREE Full text] [doi: 10.1186/1471-2105-6-S1-S5] [Medline: 15960839]
25. Lafferty J, McCallum A, Pereira F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. 2001 Presented at: The Eighteenth International Conference on Machine Learning; 2001; San Francisco, CA, USA p. 282-289 URL: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis_papers
26. Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer-Verlag, LLC; 2009.
27. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
28. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. Nature 2018. [doi: 10.1038/s41746-018-0029-1]
29. Caruana R. Multitask Learning. Machine Learning 1997 Jul;28(1):41-75. [doi: 10.1023/A:1007379606734]
30. Davis AP, Grondin CJ, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, et al. The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. Nucleic Acids Res 2015 Jan;43(Database issue):D914-D920 [FREE Full text] [doi: 10.1093/nar/gku935] [Medline: 25326323]
31. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 2010;6:343 [FREE Full text] [doi: 10.1038/msb.2009.98] [Medline: 20087340]
32. Polepalli RB, Belknap SM, Li Z, Frid N, West DP, Yu H. Automatically Recognizing Medication and Adverse Event Information From Food and Drug Administration's Adverse Event Reporting System Narratives. JMIR Med Inform 2014 Jun 27;2(1):e10 [FREE Full text] [doi: 10.2196/medinform.3022] [Medline: 25600332]
33. Xu R, Wang Q. Automatic construction of a large-scale and accurate drug-side-effect association knowledge base from biomedical literature. J Biomed Inform 2014 Oct;51:191-199 [FREE Full text] [doi: 10.1016/j.jbi.2014.05.013] [Medline: 24928448]
34. Liu M, Cai R, Hu Y, Matheny ME, Sun J, Hu J, et al. Determining molecular predictors of adverse drug reactions with causality analysis based on structure learning. J Am Med Inform Assoc 2014 Mar;21(2):245-251 [FREE Full text] [doi: 10.1136/amiajnl-2013-002051] [Medline: 24334612]

XSL•FO
RenderX

35. Yildirim P, Majnarić L, Ekmekci O, Holzinger A. Knowledge discovery of drug data on the example of adverse reaction prediction. BMC Bioinformatics 2014;15 Suppl 6:S7 [FREE Full text] [Medline: 25079450]
36. Kim J, Ohta T, Tateisi Y, Tsujii J. GENIA corpus--semantically annotated corpus for bio-textmining. Bioinformatics 2003;19 Suppl 1:i180-i182. [Medline: 12855455]
37. Doğan RI, Leaman R, Lu Z. NCBI disease corpus: a resource for disease name recognition and concept normalization. J Biomed Inform 2014 Feb;47:1-10 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.006] [Medline: 24393765]
38. Suominen H, Salanterä S, Velupillai S, Chapman W, Savova G, Elhadad N, et al. Overview of the ShARe/CLEF eHealth Evaluation Lab. 2013 Presented at: International Conference of the Cross-Language Evaluation Forum for European Languages; 2013; Berlin. [doi: 10.1007/978-3-642-40802-1_24]
39. Hanisch D, Fundel K, Mevissen H, Zimmer R, Fluck J. ProMiner: rule-based protein and gene entity recognition. BMC Bioinformatics 2005;6 Suppl 1:S14 [FREE Full text] [doi: 10.1186/1471-2105-6-S1-S14] [Medline: 15960826]
40. Campos D, Matos S, Oliveira JL. Gimli: open source and high-performance biomedical name recognition. BMC Bioinformatics 2013 Feb 15;14:54 [FREE Full text] [doi: 10.1186/1471-2105-14-54] [Medline: 23413997]
41. Hsu C, Chang Y, Kuo C, Lin Y, Huang H, Chung I. Integrating high dimensional bi-directional parsing models for gene mention tagging. Bioinformatics 2008 Jul 01;24(13):i286-i294 [FREE Full text] [doi: 10.1093/bioinformatics/btn183] [Medline: 18586726]
42. Zhou G, Zhang J, Su J, Shen D, Tan C. Recognizing names in biomedical texts: a machine learning approach. Bioinformatics 2004 May 01;20(7):1178-1190. [doi: 10.1093/bioinformatics/bth060] [Medline: 14871877]
43. Tang B, Chen Q, Wang X, Wu Y, Zhang Y, Jiang M, et al. Recognizing Disjoint Clinical Concepts in Clinical Text Using Machine Learning-based Methods. AMIA Annu Symp Proc 2015;2015:1184-1193 [FREE Full text] [Medline: 26958258]
44. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural Architectures for Named Entity Recognition. : Association for Computational Linguistics; 2016 Presented at: the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2016; San Diego, California. [doi: 10.18653/v1/N16-1030]
45. Yang J, Liang S, Zhang Y. Design Challenges and Misconceptions in Neural Sequence Labeling. 2018 Presented at: Proceedings of the 27th International Conference on Computational Linguistics; 2018; Santa Fe, New Mexico, USA p. 3879-3889 URL: http://aclweb.org/anthology/C18-1327
46. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011;18(5):552-556 [FREE Full text] [doi: 10.1136/amiajnl-2011-000203] [Medline: 21685143]
47. Fundel K, Küffner R, Zimmer R. RelEx--relation extraction using dependency parse trees. Bioinformatics 2007 Feb 01;23(3):365-371. [doi: 10.1093/bioinformatics/btl616] [Medline: 17142812]
48. Kilicoglu H, Rosemblat G, Fiszman M, Rindflesch TC. Sortal anaphora resolution to enhance relation extraction from biomedical literature. BMC Bioinformatics 2016 Apr 14;17(1). [doi: 10.1186/s12859-016-1009-6] [Medline: 27080229]
49. Lavergne T, Grouin C, Zweigenbaum P. The contribution of co-reference resolution to supervised relation detection between bacteria and biotopes entities. BMC Bioinformatics 2015;16 Suppl 10:S6 [FREE Full text] [doi: 10.1186/1471-2105-16-S10-S6] [Medline: 26201352]
50. Airola A, Pyysalo S, Björne J, Pahikkala T, Ginter F, Salakoski T. All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning. BMC Bioinformatics 2008 Nov 19;9 Suppl 11:S2 [FREE Full text] [doi: 10.1186/1471-2105-9-S11-S2] [Medline: 19025688]
51. Kordjamshidi P, Roth D, Moens M. Structured learning for spatial information extraction from biomedical text: bacteria biotopes. BMC Bioinformatics 2015 Apr 25;16:129 [FREE Full text] [doi: 10.1186/s12859-015-0542-z] [Medline: 25909637]
52. Luo Y, Cheng Y, Uzuner Ö, Szolovits P, Starren J. Segment convolutional neural networks (Seg-CNNs) for classifying relations in clinical notes. J Am Med Inform Assoc 2018 Jan 01;25(1):93-98. [doi: 10.1093/jamia/ocx090] [Medline: 29025149]
53. Verga P, Strubell E, McCallum A. Simultaneously Self-Attending to All Mentions for Full-Abstract Biological Relation Extraction. 2018 Presented at: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics; 2018; New Orleans, Louisiana p. 872-884 URL: http://aclweb.org/anthology/N18-1080
54. Miwa M, Bansal M. End-to-end Relation Extraction using LSTMs on Sequences and Tree Structures. 2016 Presented at: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics; 2016; Berlin, Germany p. 1105-1116. [doi: 10.18653/v1/P16-1105]
55. Mehryary F, Hakala K, Kaewphan S, Björne J, Salakoski T, Ginter F. End-to-End System for Bacteria Habitat Extraction. 2017 Presented at: BioNLP; 2017; Vancouver, Canada p. 80-90 URL: http://www.aclweb.org/anthology/W17-2310
56. Misra I, Shrivastava A, Gupta A, Hebert M. Cross-Stitch Networks for Multi-task Learning. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; Las Vegas, Nevada p. 3994-4003 URL: http://openaccess.thecvf.com/content_cvpr_2016/papers/Misra_Cross-Stitch_Networks_for_CVPR_2016_paper.pdf [doi: 10.1109/CVPR.2016.433]
57. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. The Journal of Machine Learning Research 2011;12:2537 [FREE Full text]

XSL•FO
RenderX

58. Ruder S. arXiv. 2017. An Overview of Multi-Task Learning in Deep Neural Networks URL: https://arxiv.org/abs/1706.05098 [accessed 2018-11-15] [WebCite Cache ID 73wxJmB0m]
59. Duong L, Cohn T, Bird S, Cook P. Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser. 2015 Presented at: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics; 2015; Beijing, China p. 845-850 URL: http://anthology.aclweb.org/P/P15/P15-2139.pdf
60. Argyriou A, Evgeniou T, Pontil M. Multi-task Feature Learning. 2006 Presented at: Proceedings of the 19th International Conference on Neural Information Processing Systems; 2006; Cambridge, MA p. 41-48 URL: https://papers.nips.cc/paper/3143-multi-task-feature-learning.pdf
61. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation 1997 Nov;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735]
62. Luong T, Pham H, Manning C. Effective Approaches to Attention-based Neural Machine Translation. 2015 Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2015; Lisbon, Portugal p. 1412-1421 URL: http://aclweb.org/anthology/D15-1166
63. Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation Classification via Convolutional Deep Neural Network. 2014 Presented at: Proceedings of the 25th International Conference on Computational Linguistics; 2014; Dublin, Ireland p. 2335-2344 URL: http://www.aclweb.org/anthology/C14-1220
64. Chen X, Cardie C. Multinomial Adversarial Networks for Multi-Domain Text Classification. 2018 Presented at: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics; 2018; New Orleans, Louisiana p. 1226-1240 URL: http://www.aclweb.org/anthology/N18-1111
65. Jagannatha A, Liu F, Liu W, Yu H. Overview of the First Natural Language Processing Challenge for Extracting Medication, Indication and Adverse Drug Events from Electronic Health Record Notes (MADE1.0). Drug Safety 2018 [FREE Full text]
66. Chapman A, Peterson K, Alba P, DuVall S, Patterson O. Hybrid system for adverse drug event detection. 2018 Presented at: International Workshop on Medication and Adverse Drug Event Detection; 2018; Worcester, MA, United States p. 16-24 URL: http://proceedings.mlr.press/v90/chapman18a/chapman18a.pdf
67. Xu D, Yadav V, Bethard S. UArizona at the MADE1.0 NLP Challenge. 2018 Presented at: International Workshop on Medication and Adverse Drug Event Detection; 2018; Worcester, MA, United States p. 57-65 URL: http://proceedings.mlr.press/v90/xu18a/xu18a.pdf
68. Dandala B, Joopudi V, Devarakonda M. IBM Research System at MADE 2018: Detecting Adverse Drug Events from Electronic Health Records. 2018 Presented at: International Workshop on Medication and Adverse Drug Event Detection; 2018; Worcester, MA, United States p. 39-47 URL: http://proceedings.mlr.press/v90/dandala18a/dandala18a.pdf
69. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc 2010;17(1):19-24 [FREE Full text] [doi: 10.1197/jamia.M3378] [Medline: 20064797]
70. Pyysalo S, Ginter F, Moen H, Salakoski T, Ananiadou S. Distributional semantics resources for biomedical text processing. 2013 Presented at: the 5th International Symposium on Languages in Biology and Medicine; 2013; Tokyo, Japan p. 39-43 URL: http://bio.nlplab.org/pdf/pyysalo13literature.pdf

Abbreviations

ADE: adverse drug event
BiLSTM: bidirectional long short-term-memory
CRF: conditional random field
EHR: electronic health record
FAERS: Food and Drug Administration Adverse Event Reporting System
LSTM: long short-term-memory
MADE: Medication, Indication, and Adverse Drug Events
MTL: multi-task learning
NER: named entity recognition
NLP: natural language processing
RE: relation extraction
POS: part of speech
SVM: support vector machine

XSL•FO
RenderX

Edited by G Eysenbach; submitted 12.09.18; peer-reviewed by H Miller, M Liu; comments to author 08.10.18; revised version received 31.10.18; accepted 09.11.18; published 26.11.18.

Please cite as:
Li F, Liu W, Yu H
Extraction of Information Related to Adverse Drug Events from Electronic Health Record Notes: Design of an End-to-End Model Based on Deep Learning
JMIR Med Inform 2018;6(4):e12159
URL: http://medinform.jmir.org/2018/4/e12159/
doi:10.2196/12159
PMID:30478023

©Fei Li, Weisong Liu, Hong Yu. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 26.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predictive Modeling of 30-Day Emergency Hospital Transport of Patients Using a Personal Emergency Response System: Prognostic Retrospective Study

Jorn op den Buijs[1], PhD; Mariana Simons[1], PhD; Sara Golas[2], MA; Nils Fischer[2], MPH; Jennifer Felsted[2,3], PhD; Linda Schertzer[4]; Stephen Agboola[2,3,5], MPH, MD; Joseph Kvedar[3,5,6], MD; Kamal Jethwani[2,3,5], MPH, MD

[1]Philips Research, Eindhoven, Netherlands

[2]Partners HealthCare Pivot Labs, Partners HealthCare, Boston, MA, United States

[3]Department of Dermatology, Harvard Medical School, Boston, MA, United States

[4]Philips Lifeline, Framingham, MA, United States

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

[6]Partners Connected Health, Partners HealthCare, Boston, MA, United States

Corresponding Author:
Jorn op den Buijs, PhD
Philips Research
High Tech Campus 34
Eindhoven, 5656 AE
Netherlands
Phone: 31 631926890
Email: jorn.op.den.buijs@philips.com

Abstract

Background: Telehealth programs have been successful in reducing 30-day readmissions and emergency department visits. However, such programs often focus on the costliest patients with multiple morbidities and last for only 30 to 60 days postdischarge. Inexpensive monitoring of elderly patients via a personal emergency response system (PERS) to identify those at high risk for emergency hospital transport could be used to target interventions and prevent avoidable use of costly readmissions and emergency department visits after 30 to 60 days of telehealth use.

Objective: The objectives of this study were to (1) develop and validate a predictive model of 30-day emergency hospital transport based on PERS data; and (2) compare the model's predictions with clinical outcomes derived from the electronic health record (EHR).

Methods: We used deidentified medical alert pattern data from 290,434 subscribers to a PERS service to build a gradient tree boosting-based predictive model of 30-day hospital transport, which included predictors derived from subscriber demographics, self-reported medical conditions, caregiver network information, and up to 2 years of retrospective PERS medical alert data. We evaluated the model's performance on an independent validation cohort (n=289,426). We linked EHR and PERS records for 1815 patients from a home health care program to compare PERS–based risk scores with rates of emergency encounters as recorded in the EHR.

Results: In the validation cohort, 2.22% (6411/289,426) of patients had 1 or more emergency transports in 30 days. The performance of the predictive model of emergency hospital transport, as evaluated by the area under the receiver operating characteristic curve, was 0.779 (95% CI 0.774-0.785). Among the top 1% of predicted high-risk patients, 25.5% had 1 or more emergency hospital transports in the next 30 days. Comparison with clinical outcomes from the EHR showed 3.9 times more emergency encounters among predicted high-risk patients than low-risk patients in the year following the prediction date.

Conclusions: Patient data collected remotely via PERS can be used to reliably predict 30-day emergency hospital transport. Clinical observations from the EHR showed that predicted high-risk patients had nearly four times higher rates of emergency encounters than did low-risk patients. Health care providers could benefit from our validated predictive model by targeting timely preventive interventions to high-risk patients. This could lead to overall improved patient experience, higher quality of care, and more efficient resource utilization.

XSL•FO
RenderX

(JMIR Med Inform 2018;6(4):e49) doi:10.2196/medinform.9907

KEYWORDS

accountable care organizations; decision support techniques; emergency medical dispatch; machine learning; population health

Introduction

Background

With the worldwide increase in the elderly population [1], chronic diseases and associated health care utilization, such as costly emergency department (ED) visits and subsequent hospitalizations, are also on the rise. In the United States, nearly 40% of patients making ED visits arrived by emergency ambulance transport [2,3], and 80% of unscheduled hospital admissions were through the ED [4]. Preventing avoidable ED visits and admissions among elderly patients is becoming a global priority [5,6], since emergency hospitalizations may be particularly distressing for older people and have been associated with adverse events such as hospital-acquired infections, loss of functional independence, and falls [7].

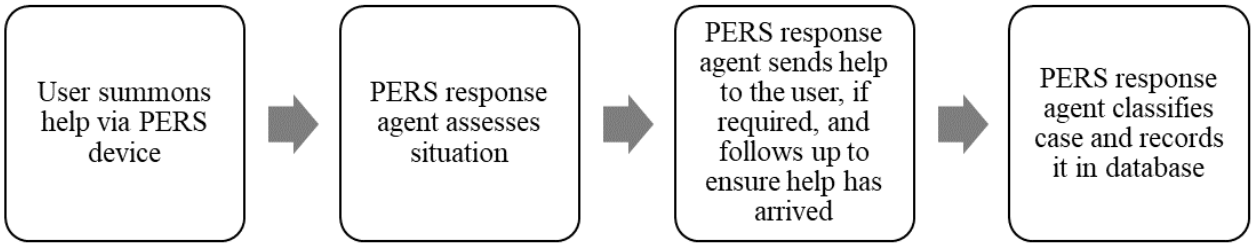
Various telehealth programs have been demonstrated to be effective in reducing readmissions, ED visits, and mortality for patients with congestive heart failure (CHF), stroke, and chronic obstructive pulmonary disease (COPD) [8]. For instance, a large nonprofit health care system in the United States introduced a home-based telehealth program that lasts 30 to 60 days after discharge from hospital, targeting the top 5% high-cost patients with multiple chronic conditions [9]. In addition, our recent work unveiled cost-saving opportunities by managing the patients in the lower-cost segments that are at risk of becoming more costly in the long term, such as beyond institutional settings and 30- to 60-day telehealth services [10]. Population health management programs may benefit from data analytics that integrate home monitoring devices to monitor patients after discharge and find out whether they have issues before ED utilization. These devices include personal emergency response systems (PERSs), which can help older adults get immediate assistance when a serious home-based accident occurs and where delayed response may result in preventable ED utilization [11]. Furthermore, our preliminary findings suggested that predictive analysis using PERS data could be useful for identifying patients at high risk for imminent emergency health care utilization [12].

A PERS service enables older adults to get help in a situation that potentially requires emergency transport by ambulance to the hospital, such as a sudden worsening of their chronic condition or a fall [13,14]. PERS is a widely used wearable technology with a help button that is worn as either a bracelet or a pendant. Patients may press the help button at any time to activate an in-home communication system that connects to a 24/7 call center. The call center associate may contact an informal responder (eg, a neighbor or a family member) or emergency medical services (EMSs) based on the patient's specific situation and follows up with the patient to confirm that help has arrived. The call center associate records notes from the conversations with the subscribers in an electronic record and classifies the type, situation, and outcome of the case (Figure 1). In combination with user enrollment data, such as demographic, caregiver network, and medical condition data, these case data provide valuable information about the patient's status.

Progress in big-data analysis techniques, such as predictive modeling to identify patients at risk of worsening health conditions, may contribute to cost-effectively reducing potentially avoidable health care utilization [15]. Previous efforts in the field of risk prediction include predictive modeling of hospital readmission [16], repeat ED visits [17,18], and the use of specialized discharge services [19]. For example, the LACE index (length of stay; acuity of the admission; comorbidity of the patient, measured with the Charlson comorbidity index score; and emergency department use, measured as the number of visits in the 6 months before admission) was designed for the prediction of death or unplanned readmissions after hospitalization [20], achieving a predictive performance of the area under the receiver operating characteristic curve (AUC) of 0.68. Another model, HOSPITAL (hemoglobin, discharge from an oncology service, sodium level, procedure performed, index type of admission [urgent vs elective], number of admissions in the last year, and length of stay), is a risk score for predicting 30-day potentially avoidable readmission with a performance of AUC=0.72 as evaluated in 9 hospitals in 4 countries [21]. Yet another study used 1-year retrospective electronic health record (EHR) data to predict 30-day ED revisits, with predictive power of AUC=0.70, in a prospective validation cohort [18]. These predictive models rely only on clinical data collected before or during the clinical encounters or at the time of discharge, but not on data from home monitoring postdischarge. In contrast, PERS services collect information while the patient is back at home. This information includes the details—time stamp, type, situation, and outcome—of incidents such as falls, respiratory issues, chest pain, or general pain, as well as other check-in or social calls [13]. Such events may indicate a decline in patient status, and some patients may eventually request emergency transport to the hospital via the PERS service. Thus, patient decline may be captured earlier with PERS–based prediction models than with models based on clinical data. Therefore, the hypothesis of this study is that the medical alert pattern data collected via the PERS service may be used to predict imminent risk for emergency transport to the hospital.

XSL•FO
RenderX

Figure 1. Overview of the personal emergency response system (PERS) process and data collection.



Objectives

The objectives of this study were to (1) develop and validate a predictive model of 30-day emergency hospital transport based on PERS data; and (2) compare the model's predictions with clinical outcomes derived from the EHR, namely outpatient and inpatient emergency encounters. We used a large, deidentified PERS dataset of more than 580,000 individuals to develop and validate the predictive model. For a subpopulation of 1815 patients, we linked PERS and EHR data so that we could compare PERS–predicted risk of hospital transport with observed clinical outcomes.

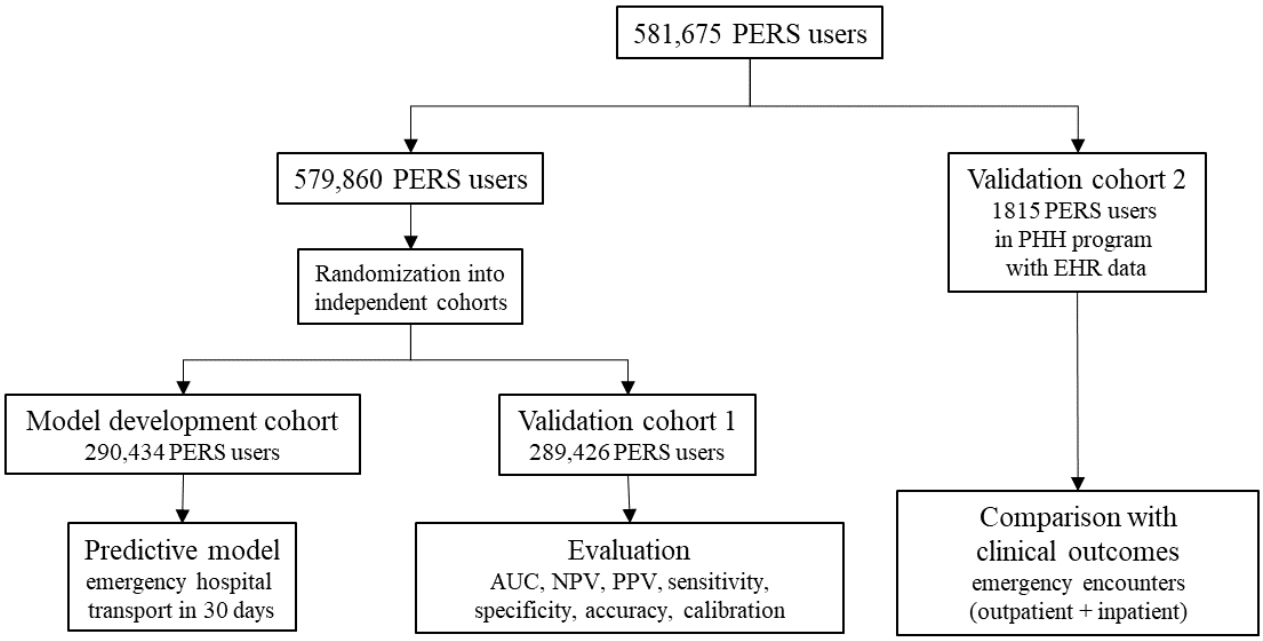
Methods

General Overview

In this study, we followed the guidelines by Luo et al for developing and reporting machine learning predictive models in biomedical research [22]. This was a prognostic retrospective predictive modeling study of emergency hospital transport in the next 30 days. Figure 2 presents an overview of the methodology for predictive model development and evaluation. We used retrospective, deidentified data of 581,675 subscribers of a commercial PERS service (Philips Lifeline, Framingham, MA, USA). We developed and validated the predictive model on cohort 1 using data from 579,860 subscribers. For 1815 subscribers (validation cohort 2), who were patients of the Partners HealthCare at Home (PHH; Partners HealthCare at Home, Inc, Waltham, MA, USA) program in the greater Boston, Massachusetts area, we also collected clinical outcomes in the year following prediction to evaluate how prediction of emergency hospital transport compared with rates of outpatient and inpatient emergency encounters. Patients in the model development cohort and both validation cohorts were mutually exclusive.

This study was approved by the Internal Committee on Biomedical Experiments of Philips Research (ICBE-2-16049). Furthermore, we obtained approval for linkage of PERS and EHR records from the Partners HealthCare institutional review board.

Figure 2. Overview of the study design to develop and evaluate the predictive model of emergency hospital transport. AUC: area under the receiver operating characteristic curve; EHR: electronic health record; NPV: negative predictive value; PERS: personal emergency response system; PHH: Partners HealthCare at Home; PPV: positive predictive value.



XSL•FO
RenderX

Study Cohorts

Model Development Cohort and Validation Cohort 1

We included patients if they were subscribers of the Lifeline PERS service on January 1, 2014 (model development cohort) or February 1, 2014 (model validation cohort 1) and were between the ages of 18 and 100 years. We excluded patients if they appeared to be more than 10 years on the service due to the use of a different back-end system prior to 2004 where some patients may have left the service in the meantime without an indication recorded in the database. Also, patients who left the PERS service for any reason in 30 days from the prediction date were excluded from the analysis. Reasons for leaving the service included death, moving to long-term care, and financial reasons. We excluded patients from the model development cohort and validation cohort 1 if they were part of the PPH program, as we included these patients into validation cohort 2. Furthermore, 3.27% (19,886/606,848) of total patients had missing demographic data and were excluded from the analysis.

The primary PERS data source for model development was the Philips Lifeline database, which contained historical data such as demographics, patients' living situation and caregiver network, self-reported medical conditions, and medical alert data. We collected data on self-reported medical conditions, including medication allergies, using a custom coding system, which consisted of a drop-down list of common disease categories, such as COPD, CHF, and diabetes. Caregiver network information included the number of responders, the number of people who lived with the patients, and persons to be notified after an incident occurred. Medical alert data from the PERS device included all information gathered during the interactions of the patients with the Lifeline call center associates. Call center associates categorized the calls as incidents or nonincidents. For all calls, the situation (eg, fall, respiratory problems, chest pain, and social call) and the outcome (eg, subscriber okay, responder assistance, and emergency hospital transport) was collected by the associate via custom-made software.

Validation Cohort 2

Validation cohort 2 included PERS users who were on the service on February 1, 2014 and received care at PHH, a homecare management service that offers general care as well as specialized services to help patients manage chronic conditions at home. PHH uses a variety of technological innovations to remotely monitor their patients, including the Philips Lifeline PERS, which is a service routinely recommended for elderly or chronically ill patients who are at risk of falls or other health-related emergencies. We excluded patients if they left the PERS service in the 30 days after the prediction date.

The primary data source for comparison of PERS–based predicted risk for emergency hospital transport with clinical outcomes in validation cohort 2 was the Enterprise Data Warehouse, an electronic medical record data repository of hospitals within the Partners HealthCare System. The data include demographic information, medical conditions, and hospital utilization. We combined longitudinal clinical data from the EHR (from February 1, 2014 to January 31, 2015) for 1815 individuals who met the inclusion criteria with the PERS–based predictive score of emergency hospital transport. All data were deidentified before analysis. It should be noted that, in this study, we did not use EHR data to train the predictive model; rather, we compared PERS–based predictions of 30-day emergency hospital transport with clinical observations from the EHR.

Data Processing

We processed PERS data for input into the predictive model using the statistical programming language R [23]. Different database tables were extracted from an operational PERS database, deidentified, and made accessible to the research team.

One table included subscriber demographics, including age, sex, and region in the United States, subscription type, and enrollment information. A few test accounts, used for demonstration purposes and training of call center agents, were removed prior to the predictive analysis. We used enrollment information to determine which subscribers were active on the prediction date. Subscribers with missing enrollment dates (79,764/1,221,073, 6.53% of all subscribers in the initial data extract) were already deactivated or were pending a PERS installation and were therefore excluded from further analysis. We derived time on the PERS service as the number of days between enrollment and prediction dates.

One table consisted of the subscribers' caregiver network information. This included the number of responders (eg, son, daughter, or neighbor) and number of emergency service providers (eg, ambulance or fire department) listed by the subscriber at the time of enrollment.

One table included self-reported medical conditions and medication allergies. Medical conditions and medication allergies were recorded at enrollment by the call center agent via a custom-made drop-down list for the purpose of informing EMSs in case of an incident. We extracted the 50 most frequently occurring medical conditions and medication allergies from the database for use in the prediction model. Up to 3 of the most common medical conditions and medication allergies were listed for various categories (Table 1).

XSL•FO
RenderX

Table 1. The most common self-reported medical conditions and medication allergies by personal emergency response system subscribers per category (N=581,675).

Category and condition description	Subscribers, n (%)
Auditory	
Hearing loss	66,102 (11.36)
Hearing aid	20,802 (3.58)
Cancer	
Cancer	11,507 (1.98)
Cardiovascular	
High blood pressure	221,416 (38.07)
Heart condition	80,113 (13.77)
History of stroke	41,925 (7.21)
Endocrine	
Diabetes	133,694 (22.98)
Thyroid disease	10,435 (1.79)
Medication allergy	
Penicillin	85,982 (14.78)
Sulfa drugs	59,327 (10.20)
Codeine	42,581 (7.32)
Musculoskeletal, falls, or movement problems	
Cane, crutches, or walker	136,215 (23.42)
Arthritis	83,565 (14.37)
History of falls	83,395 (14.34)
Neurological	
Balance problems or unsteady gait	38,672 (6.65)
Dementia	20,360 (3.50)
Dizziness or vertigo	17,557 (3.02)
Psychiatric	
Depression	19,788 (3.40)
Anxiety	12,848 (2.21)
Pulmonary	
Chronic obstructive pulmonary disease	40,868 (7.03)
Oxygen dependent	28,190 (4.85)
Asthma	21,905 (3.77)
Visual	
Impaired vision	20,055 (3.45)
Glasses	15,078 (2.59)
Macular degeneration	13,207 (2.27)

One table consisted of up to 2 years of historical PERS service use data, which totaled to more than 25 million data points. We categorized cases by case type, situation, and outcome (Table 2). For each case, we constructed 2 features for input to the predictive model, namely frequency and recency. Frequency was the sum of cases experienced by the subscriber prior to the prediction date. Recency was the number of days between the most recent case prior to the prediction date and the prediction date.

XSL•FO
RenderX

Table 2. Case types, situations, and outcomes for which frequency and recency features were derived for input into the predictive model. Examples are given per category.

Case category and classification example	Description
Types	
Incident	Case where help is sent to the subscriber
Accidental	Subscriber accidentally pushed the help button
Check-in	Social call by subscriber
Situations	
Breathing problems, chest pain, dizziness, fall, illness, other pain	Incident with subscriber experiencing breathing problems, chest pain, dizziness, a fall, illness, or other pain
Outcomes	
Check-in	Social call
EMS[a]—transport	EMS sent to subscriber's house—subscriber transported to hospital
Responder—no transport	Responder sent to subscriber's house—subscriber not transported to hospital

[a]EMS: emergency medical service.

We derived the dependent variable—having 1 or more emergency hospital transports in the 30 days from the prediction date—from the PERS service use data. As the prediction problem was treated as a classification problem, the dependent variable was rendered binary, with 1 if the patient had 1 or more emergency hospital transports in the 30 days following the prediction date and 0 otherwise.

After preprocessing, we merged the data into 1 large table with 129 columns representing predictive model features and 1 column for the dependent variable. Each of the rows corresponded to 1 PERS user. For 1815 PERS users in the PHH program, clinical outcomes from the EHR were available, and we assigned these to validation cohort 2. We randomly assigned the remaining PERS users on a 1:1 basis to the model development cohort and the validation cohort 1.

Predictive Model of Emergency Hospital Transport

We then developed a predictive model using a boosted regression trees approach called extreme gradient boosting within R [24]. We chose extreme gradient boosting because it was adopted by more than half of the winning competitions on the Kaggle platform for data science competitions in 2015 [24]. This methodology was also carried out within the commercial PERS–based prediction system CareSage (Philips Lifeline). During initial model development, we also considered logistic regression as a candidate modeling technique but it resulted in significantly lower AUC values in the validation cohort.

The frequency and recency features based on medical alert pattern data are highly skewed. While this may pose a problem for certain predictive modeling methods, the tree-based method we used did not make any assumptions on normality of the features. Monotonic transformations of the features to render them more normal did not affect the predictive performance of the models. We determined variable importance according to the gain, a measure of the relative contribution of the corresponding variable to the predictive model, calculated by taking the improvement in accuracy brought by a variable to the branches it is on. We reduced the number of features by selecting only those with non–zero gain values for the final model, resulting in a total of 121 features in the final model. The boosted regression model also involved tuning hyperparameters of the learning algorithm, such as the number of trees, the maximum depth of the trees, and the learning rate. We achieved this optimization using 5-fold cross-validation on the model development set with the optimization metric determined by the AUC in the test fold.

Predictive Model Evaluation

We evaluated the discriminatory accuracy of the predictive models using the AUC, which indicates the probability of the predictive model ranking a randomly selected patient with 30-day emergency transport higher than a randomly selected patient without the event. Furthermore, the negative predictive value (NPV) is the percentage of patients not having emergency transport in the group classified as negative, while the positive predictive value (PPV) indicates the percentage of patients having emergency transport in the group classified as positive. We varied the threshold for classifying patients as positive using risk scores above the 90th, 95th, and 99th percentiles such that 10%, 5%, and 1%, respectively, of patients were classified as high risk. For these thresholds, we computed the NPV, PPV, sensitivity, specificity, and accuracy. We derived 95% confidence intervals for performance metrics using a stratified bootstrapping method with 1000 bootstrap replicates.

While we used AUC to assess the ability of the predictive model to discriminate between patients with and without transport, regularization, as done in extreme gradient boosting, may have created bias in the predictive model [22]. Therefore, we also checked the model for calibration—the agreement between predictions made by the model and the outcome—by plotting the observed outcome in ranges of the predicted probabilities (0%-20%, 20%-40%, …, 80%-100%). We also tested goodness-of-fit using the Hosmer-Lemeshow test. This is a systematic way of assessing whether the observed outcomes match the predicted probabilities in subgroups of the model population.

XSL•FO
RenderX

Statistical Analysis

We tested differences in age between the model development and 2 validation cohorts using analysis of variance followed by Tukey post hoc test. Pairwise comparison between pairs of proportions with Holm correction for multiple testing was applied to the patients' sex, self-reported medical conditions, and 30-day emergency transport.

To determine how PERS predicted risk for 30-day emergency hospital transport compared with clinical outcomes, we derived the rates of emergency outpatient and inpatient encounters from the EHR at 30, 60, 90, 180, and 365 days following prediction for validation cohort 2. We split the cohort into low, medium, and high risk according to the risk thresholds that corresponded to the 0 to 50th percentile, 51st to 95th percentile, and greater than 95th percentile. The numbers of emergency encounters in these 3 risk groups were normalized as the number of emergency encounters per 100 patients. Emergency encounters were statistically compared using a pairwise Wilcoxon rank sum test.

Results

Patient Characteristics and Prevalence of the Dependent Variable

Table 3 presents the characteristics of the model development and both validation cohorts. PERS users were on average 81 years old and most were female (234,817/290,434, 80.85%) in the model development cohort and validation cohort 1. There was no statistically significant difference in age between the model development and validation cohort 1. Average age was slightly, but statistically significantly, younger in validation cohort 2 than in the model development cohort.

The distribution of the number of self-reported conditions was different in validation cohort 2, with only 6.89% (125/1815) of users not reporting any conditions, compared with 20.58% (59,769/290,434) in the model development cohort. One reason may be that these were patients receiving care for 1 or more specific conditions through a home health care program. Regarding the dependent variable of having 1 or more emergency hospital transports in 30 days, 2.20% (40/1815) of patients had emergency hospital transport in the 30 days following the prediction date in validation cohorts 1 and 2. This was not significantly different from the model development cohort (6686/290,434, 2.30%). The ratio of positive to negative classes was 0.024 for the model development cohort and 0.023 for both validation cohorts. According to the EHR data, 509/1815 patients of validation cohort 2 (28.04%) had 1 or more emergency encounters in the year following the prediction date.

Model Performance

Table 4 details the performance of the predictive model in both validation cohorts for various risk thresholds. The AUC for emergency hospital transport in 30 days was 0.779 (95% CI 0.774-0.785) in validation cohort 1 and 0.766 (95% CI 0.686-0.845) in validation cohort 2. Nonsignificant Hosmer-Lemeshow test results (P=.99 for validation cohort 1 and P=.78 for validation cohort 2) showed that predicted probabilities matched observed outcomes.

Table 3. Patient characteristics and prevalence of the dependent variable in the model development and validation cohorts.

Variable	Model development cohort, (n=290,434)	Validation cohort 1 (n=289,426)		Validation cohort 2 (n=1815)	
		Statistic	*P* value	Statistic	*P* value
Prediction date	January 1, 2014	February 1, 2014	N/A[a]	February 1, 2014	N/A
Age (years), mean (SD)	81.3 (11.5)	81.2 (11.5)	.20	79.8 (11.5)	<.001
Female patients, n (%)	234,817 (80.85)	233,692 (80.74)	.66	1461 (80.50)	.91
PERS[b] self-reported medical conditions, n (%)					
0	59,769 (20.58)	61,685 (21.31)	<.001	125 (6.89)	<.001
1-2	72,067 (24.81)	71,094 (24.56)	.02	656 (36.14)	<.001
3-4	77,739 (26.77)	76,384 (26.39)	.27	519 (28.60)	.31
≥5	80,859 (27.84)	80,263 (27.73)	.25	515 (28.37)	.84
Patients with 30-day emergency hospital transport, n (%)	6686 (2.30)	6411 (2.22)	.08	40 (2.20)	.99
Patients with ≥ 1 emergency encounters at 3 time points after the prediction date, n (%)					
30 days	N/A	N/A	N/A	64 (3.53)	N/A
180 days	N/A	N/A	N/A	321 (17.69)	N/A
365 days	N/A	N/A	N/A	509 (28.04)	N/A

[a]N/A: not available.

[b]PERS: personal emergency response system.

XSL•FO
RenderX

Table 4. Performance of the predictive model evaluated by the area under the receiver operating characteristic curve (AUC), negative predictive value (NPV), positive predictive value (PPV), sensitivity, specificity, and accuracy using the 90th, 95th, and 99th percentiles as thresholds in the 2 validation cohorts.

Characteristic	Prediction score threshold		
	>90th percentile	>95th percentile	>99th percentile
Validation cohort 1 (n=289,426)[a]			
AUC (95% CI)	0.779 (0.774-0.785)	0.779 (0.774-0.785)	0.779 (0.774-0.785)
NPV (95% CI)	98.6% (98.6%-98.7%)	98.4% (98.4%-98.4%)	98.0% (98.0%-98.0%)
PPV (95% CI)	9.6% (9.3%-9.9%)	13.5% (13.1%-14.0%)	25.5% (24.1%-27.2%)
Sensitivity (95% CI)	43.8% (42.5%-45.0%)	30.5% (29.3%-31.7%)	11.5% (10.7%-12.3%)
Specificity (95% CI)	90.7% (90.6%-90.8%)	95.6% (95.5%-95.7%)	99.2% (99.2%-99.3%)
Accuracy (95% CI)	89.6% (89.5%-89.7%)	94.1% (94.1%-94.2%)	97.3% (97.3%-97.3%)
Validation cohort 2 (n=1815)[b]			
AUC (95% CI)	0.766 (0.686-0.845)	0.766 (0.686-0.845)	0.766 (0.686-0.845)
NPV (95% CI)	98.8% (98.4%-99.1%)	98.3% (98.0%-98.7%)	98.0% (97.8%-98.3%)
PPV (95% CI)	8.3% (5.9%-10.7%)	9.2% (5.1%-13.9%)	16.7% (4.3%-29.6%)
Sensitivity (95% CI)	52.5% (37.5%-67.5%)	30.0% (15.0%-45.0%)	12.5% (2.5%-22.5%)
Specificity (95% CI)	86.9% (85.3%-88.4%)	93.4% (92.2%-94.5%)	98.6% (98.0%-99.1%)
Accuracy (95% CI)	86.2% (84.4%-87.7%)	92.0% (90.7%-93.2%)	96.7% (96.1%-97.3%)

[a]Hosmer-Lemeshow test: $\chi^2_{98}=13.1$; $P=.99$.

[b]Hosmer-Lemeshow test: $\chi^2_{8}=4.8$; $P=.78$.

PPVs were on the low side due to the low prevalence of 30-day emergency hospital transport, which was around 2.2% in both validation cohorts (Table 3). By increasing the prediction score threshold, PPV increased but at the expense of decreased sensitivity. In validation cohort 1, at a risk score threshold corresponding to the 90th percentile, the predictive model identified 43.8% (95% CI 42.5%-45.0%) of the patients who had emergency transport in the 30 days following the prediction date (sensitivity); however, only 9.6% (95% CI 9.3%-9.9%) of flagged patients had emergency transport in the following 30 days (PPV) at this threshold. At thresholds corresponding to the 95th and 99th percentiles, the sensitivity dropped to 30.5% (95% CI 29.3%-31.7%) and 11.5% (95% CI 10.7%-12.3%), respectively, while the PPV increased to 13.5% (95% CI 13.1%-14.0%) and 25.5% (95% CI 24.1%-27.2%), respectively. This analysis using different thresholds illustrates the trade-off between sensitivity and PPV, that is, trying to find as many positive cases as possible with an acceptable false-positive rate. A similar trade-off between sensitivity and PPV was also observed in validation cohort 1.

The predictive model produced a probability (from 0% to 100%) for each patient to assess the risk of 30-day emergency hospital transport. The actually observed percentage of patients with emergency hospital transport increased as the predicted probabilities increased (Figure 3). At predicted probabilities between 80% and 100%, 80% (4/5) of patients had 30-day emergency hospital transport in validation cohort 1.

Predictor Variables

Table 5 provides the number of predictors and total gain, a measure of predictability, for each broad category of predictors. Predictors from the medical alert data formed the most important predictor category for the predictive model, as they accounted for 87.7% of the total gain. A total of 8 predictor variables with zero gain did not contribute to predictability. These included certain uncommon medication allergies and self-reported medical conditions.

The predictive model of 30-day emergency hospital transport included 121 variables with nonzero values for the gain. Figure 4 shows the 5 most important predictors for each category. For self-reported medical conditions, COPD, CHF, and heart conditions are among the 5 most important predictors. Other important predictors include age, sex, and the number of responders.

XSL•FO
RenderX

Figure 3. Observed percentage of patients needing 30-day emergency hospital transport versus model-predicted probability in validation cohort 1.

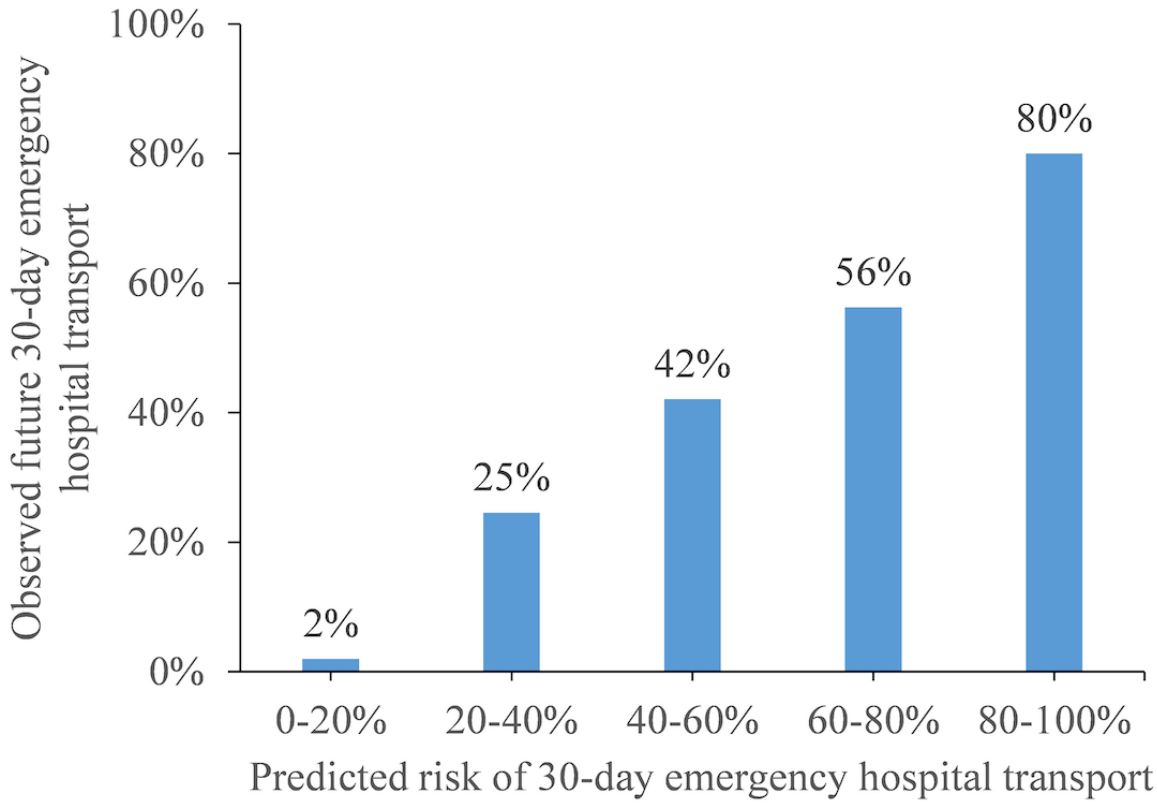


Table 5. Number of predictors and total gain per predictor category.

Predictor category	Number of predictors	Total gain (%)
Medical alert pattern-based predictors	62	87.7
Self-reported medical conditions and medication allergies	44	3.7
Other predictors	15	8.7

Comparison With Clinical Outcomes

We then used the predictive model of 30-day emergency hospital transport to segment validation cohort 2 into high risk (>95th percentile), medium risk (51st-95th percentile), and low risk (≤50th percentile) after ranking them according to the predicted probabilities of 30-day emergency hospital transport. We normalized the number of emergency encounters (both outpatient visits and inpatient admissions) in the year following the prediction date to 100 patients (Figure 5). Compared with emergency encounters in the low-risk group, patients in the high-risk group had significantly more emergency encounters at 30 (P=.001), 60 (P<.001), 90 (P<.001), 180 (P<.001), and 365 (P<.001) days. The medium-risk group had significantly more emergency encounters at 90 (P=.01), 180 (P<.001), and 365 (P<.001) days. After 365 days, there were 3.9 times more emergency encounters in the high-risk group than in the low-risk group.

XSL•FO
RenderX

Figure 4. The 5 most important variables in the predictive model for 3 categories of predictors: predictors derived from medical alert data, self-reported medical conditions, and other predictors. Predictor importance as measured by the gain is reported for validation cohort 1. COPD: chronic obstructive pulmonary disease; PERS: personal emergency response system.

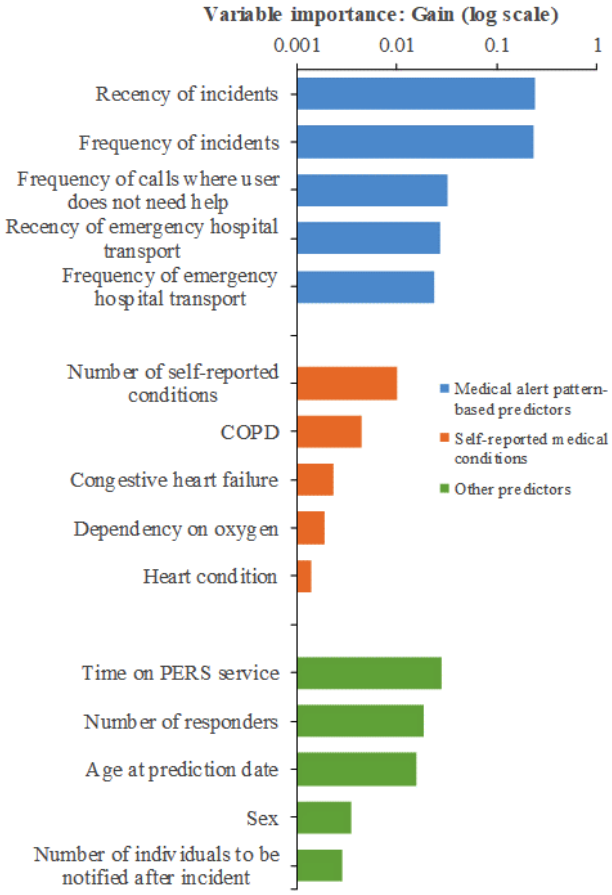
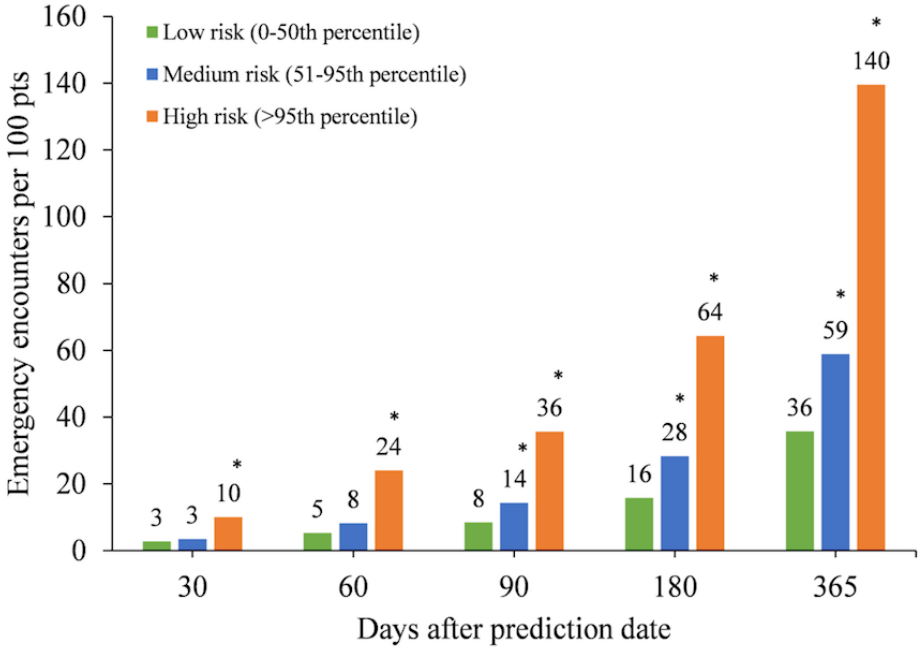


Figure 5. Emergency encounters per 100 patients (pts) in low-, medium-, and high-risk groups in the year following the prediction date. Data shown are for validation cohort 2. *P<.05 compared with low risk, pairwise Wilcoxon rank sum test.



XSL•FO
RenderX

Discussion

Principal Findings

Retrospective validation demonstrated the effectiveness of the PERS medical alert data-based predictive model of 30-day emergency hospital transport in identifying patients at risk of hospital transport with good discrimination (AUC=0.779). A goodness-of-fit test and calibration plot indicated that the model-predicted probabilities matched with observed outcomes across ranges of predicted risk. We evaluated the trade-off between finding many true-positive cases (ie, a high sensitivity) and reducing the number of false-positives (ie, a high PPV) by varying the threshold to classify patients as positive. At a risk threshold corresponding to the 99th percentile, 1 out of 4 patients flagged at high risk needed hospital transport in the next 30 days (PPV=25.5%), while at this threshold the sensitivity was rather low at 11.5%. PPV was nearly 12 times higher than the prevalence of 30-day emergency transport in validation cohort 1, which was 2.20%. The trade-off between sensitivity and PPV was also reported by other predictive modeling studies of emergency health care utilization [18,25].

Recent studies have reported on the development and validation of predictive models that can be used to systematically identify individuals at high risk for an unfavorable outcome in a hospital setting [26,27]. Such risk assessment models have great potential to inform treatment decisions and improve the quality of care delivered to patients [26-28]. As innovative connected health technologies and value-based care policy influence the evolution of geriatric models of care to deliver more individualized, multifaceted management strategies, home health programs for community-dwelling seniors may benefit from enhanced risk assessment of patients. For instance, a 32-center randomized controlled trial of the guided care model used health care encounter–based predictive modeling software to identify the 25% of patients with the highest risk of needing complex health care in the coming year. The study provided them with geriatric assessment and specialized support, resulting in significantly reduced health care utilization by home health care patients compared with a group that followed usual care [29]. By using remotely collected PERS data to predict the risk of emergency health care utilization, our predictive model of 30-day emergency hospital transport presents a unique opportunity to efficiently allocate limited health care resources to patients who need them the most and thereby reduce costs associated with excessive use.

This predictive modeling study was, to the best of our knowledge, unique, as it used PERS data to predict emergency hospital transport. Most state laws in the United States require that patients in need of emergency medical care must be taken to the nearest appropriate health care facility capable of treating the patient, which may be an out-of-network facility. Furthermore, in case of ED crowding [30], ambulance diversion to other potentially out-of-network facilities may be requested. In both cases, the resulting ED visit may not be captured in the EHR of the in-network organization. Therefore, predictive models of emergency health care utilization developed on EHR data of a particular health care provider may be missing outcome data. We minimized this limitation in our study, since response agents from the PERS service recorded information about any ambulance transport after an incident that required help from an EMS in the electronic case record of the user.

Linkage of PERS and EHR data in a cohort of 1815 PERS users who were patients of an accountable care organization in the greater Boston area group enabled comparison of PERS–based predicted risk of 30-day emergency hospital transport with emergency encounters derived from the clinical EHR. The patients in the high-risk group—that is, those above the 95th percentile range of risk for emergency transport—also had 3.9 times more emergency encounters per 100 patients in the year following the prediction date than did patients classified as having low risk. Here we specifically analyzed rates of health care encounters, as these are the events that could ultimately be avoided with the appropriate interventions. These results suggest that prediction of emergency health care utilization based on PERS data may be a good alternative to EHR-based prediction models, which could be especially helpful for continuous monitoring of patients after discharge and where patients have missing or limited EHR records.

Our previous study on health care utilization by PERS users indicated that 21% of emergency admissions were considered potentially avoidable [10]. Therefore, we believe that prediction of risk for emergency transport, in combination with appropriate interventions, could potentially reduce emergency hospitalizations. Clinical and financial implications of predictive models largely depend on how well case managers and health professionals can integrate risk prediction of patients into clinical workflows. It is vital to have a detailed guideline that clarifies how the algorithm will inform care [15]. An example of such an approach is a predictive model to detect impending deterioration of patients outside the intensive care unit [31], which was implemented in routine clinical care in 2 community hospitals [32]. In an ongoing randomized clinical trial, we are developing workflows that integrate daily PERS–based risk of 30-day emergency hospital transport with care pathways [33]. In this randomized clinical trial, we are using the predictive model described herein to predict patients' risk for 30-day emergency transport, followed by a nurse assessment and tailored interventions for high-risk patients. The number of patients who will ultimately benefit from a combination of prediction and intervention will depend on various factors, including the population size and the prevalence of emergency health care utilization, the performance of the predictive model and the risk threshold above which patients are considered to be high risk, and the efficacy of the interventions provided to high-risk patients.

Limitations

This study had a few limitations. The PERS population is mostly old and primarily female, and the service is predominantly privately paid for by patients and not covered by their health insurance. This may limit the generalizability of the study to older women who can afford the service. Interpretation of the predictive model may have been influenced by confounding of unobserved variables, including when and where users wore the PERS device [34]. While the second objective of the study

XSL•FO
RenderX

was to validate the PERS–based prediction model in a specific health system, we believe it is more broadly applicable, since the prediction model was developed on PERS–data only and did not rely on specific inputs from the EHR.

More PERS subscribers in validation cohort 2 self-reported on medical conditions than in the development and validation cohorts. It is likely that this cohort was a biased patient group with higher risk and more medical conditions, as they were receiving care from a health care provider. Given that the AUC for 30-day transport was not significantly lower in validation cohort 2, we do not expect that this affected the predictions. Furthermore, patients with emergency transport in validation cohort 2 may not have ended up in the ED of one of the hospitals of the Partners HealthCare network, but in a different hospital system, such that emergency health care utilization was not recorded in the EHR data that we used. Additionally, patients may have initiated emergency hospital transport outside of the PERS service. Both above-mentioned limitations may have affected the correlation between PERS–based risk for emergency hospital transport and emergency encounters derived from the EHR.

Conclusions

This study showed that remotely collected patient data using a PERS service can be used to predict 30-day hospital transport. Furthermore, linking these data to clinical observations from the EHR showed that predicted high-risk patients had nearly four times higher rates of emergency encounters in the year following the prediction date compared with low-risk patients. Health care providers could benefit from our validated predictive model by estimating the risk of 30-day emergency hospital transport for individual patients and target timely preventive interventions to high-risk patients. We are testing this hypothesis in a randomized clinical trial where risk predictions are combined with a stepped intervention pathway. This approach could lead to overall improved patient experience, higher quality of care, and more efficient resource utilization. Future studies should explore the impact of combined EHR and PERS data on predictive accuracy.

Authors' Contributions

JodB, MS, SG, LS, and SA designed the research. JodB, MS, and SG conducted the analyses. JF, LS, JK, and KJ provided feedback on analyses and interpretation of results. JodB, MS, SG, NF, JF, LS, and SA wrote the paper, and JodB had primary responsibility for the final content. All authors read and approved the final manuscript.

Conflicts of Interest

Philips funded the study. JodB, MS, and LS are employed by Philips.

References

1. Suzman R, Beard JR, Boerma T, Chatterji S. Health in an ageing world--what do we know? Lancet 2015 Feb 07;385(9967):484-486. [doi: 10.1016/S0140-6736(14)61597-X] [Medline: 25468156]
2. Ellis G, Marshall T, Ritchie C. Comprehensive geriatric assessment in the emergency department. Clin Interv Aging 2014 Nov;9:2033-2043 [FREE Full text] [doi: 10.2147/CIA.S29662] [Medline: 25473275]
3. Albert M, McCaig LF, Ashman JJ. Emergency department visits by persons aged 65 and over: United States, 2009-2010. NCHS Data Brief 2013 Oct(130):1-8 [FREE Full text] [Medline: 24152709]
4. Kocher KE, Dimick JB, Nallamothu BK. Changes in the source of unscheduled hospitalizations in the United States. Med Care 2013 Aug;51(8):689-698. [doi: 10.1097/MLR.0b013e3182992c7b] [Medline: 23752257]
5. Iwata M, Kuzuya M, Kitagawa Y, Iguchi A. Emergency department use by nonagenarian patients. Geriatr Gerontol Int 2006 Mar;6(1):25-31. [doi: 10.1111/j.1447-0594.2006.00310.x]
6. Downing A, Wilson R. Older people's use of accident and emergency services. Age Ageing 2005 Jan;34(1):24-30. [doi: 10.1093/ageing/afh214] [Medline: 15496462]
7. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008 Jun;17(3):216-223 [FREE Full text] [doi: 10.1136/qshc.2007.023622] [Medline: 18519629]
8. Bashshur RL, Shannon GW, Smith BR, Alverson DC, Antoniotti N, Barsan WG, et al. The empirical foundations of telemedicine interventions for chronic disease management. Telemed J E Health 2014 Sep;20(9):769-800 [FREE Full text] [doi: 10.1089/tmj.2014.9981] [Medline: 24968105]
9. Dahl D, Reisetter JA, Zismann N. People, technology, and process meet the triple aim. Nurs Adm Q 2014;38(1):13-21. [doi: 10.1097/NAQ.0000000000000006] [Medline: 24317028]
10. Agboola S, Simons M, Golas S, op den Buijs J, Felsted J, Fischer N, et al. Health care cost analyses for exploring cost savings opportunities in older patients: longitudinal retrospective study. JMIR Aging 2018 Aug 01;1(2):e10254. [doi: 10.2196/10254]
11. Bakk L. Does PERS and falls prevention information reduce stress, anxiety, falls risk, hospital admission and emergency room use for older adults on the wait list for home-and community-based services? Final Report. Southfield, MI: Area Agency on Aging 1-B; 2011 May. Accessed 2018-10-09 URL: http://www.aaa1b.com/wp-content/uploads/2010/07/PERS-Report-Final-May-2011.pdf [accessed 2018-10-09] [WebCite Cache ID 732e1k5Jz]

XSL•FO
RenderX

12. op den Buijs J, Smits T, Pijl M, Simons M, Schertzer L. Predictive modeling of emergency hospital transport using medical alert pattern data: retrospective cohort study. iProceedings 2015 Oct 27;1(1):e19. [doi: 10.2196/iproc.4772]
13. Agboola S, Golas S, Fischer N, Nikolova-Simons M, op den Buijs J, Schertzer L, et al. Healthcare utilization in older patients using personal emergency response systems: an analysis of electronic health records and medical alert data. BMC Health Serv Res 2017 Apr 18;17(1):282. [doi: 10.1186/s12913-017-2196-1] [Medline: 28420358]
14. Redd JL, Zura RD, Tanner AE, Walk EE, Wu MM, Edlich RF. Personal emergency response systems. J Burn Care Rehabil 1992;13(4):453-459. [Medline: 1429819]
15. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]
16. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, et al. Risk prediction models for hospital readmission: a systematic review. JAMA 2011 Oct 19;306(15):1688-1698 [FREE Full text] [doi: 10.1001/jama.2011.1515] [Medline: 22009101]
17. Meldon SW, Mion LC, Palmer RM, Drew BL, Connor JT, Lewicki LJ, et al. A brief risk-stratification tool to predict repeat emergency department visits and hospitalizations in older patients discharged from the emergency department. Acad Emerg Med 2003 Mar;10(3):224-232 [FREE Full text] [Medline: 12615588]
18. Hao S, Jin B, Shin AY, Zhao Y, Zhu C, Li Z, et al. Risk prediction of emergency department revisit 30 days post discharge: a prospective study. PLoS One 2014;9(11):e112944 [FREE Full text] [doi: 10.1371/journal.pone.0112944] [Medline: 25393305]
19. Holland DE, Harris MR, Leibson CL, Pankratz VS, Krichbaum KE. Development and validation of a screen for specialized discharge planning services. Nurs Res 2006;55(1):62-71. [Medline: 16439930]
20. van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 2010 Apr 06;182(6):551-557 [FREE Full text] [doi: 10.1503/cmaj.091117] [Medline: 20194559]
21. Donzé JD, Williams MV, Robinson EJ, Zimlichman E, Aujesky D, Vasilevskis EE, et al. International validity of the HOSPITAL score to predict 30-day potentially avoidable hospital readmissions. JAMA Intern Med 2016 Apr;176(4):496-502 [FREE Full text] [doi: 10.1001/jamainternmed.2015.8462] [Medline: 26954698]
22. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
23. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation; 2015.
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system [arXiv:1603.02754]. 2016 Mar 09. URL: https://arxiv.org/pdf/1603.02754v2.pdf [accessed 2018-10-08] [WebCite Cache ID 7319sAmvk]
25. Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. BMJ 2006 Aug 12;333(7563):327 [FREE Full text] [doi: 10.1136/bmj.38870.657917.AE] [Medline: 16815882]
26. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study. BMJ Open 2015 Mar 17;5(3):e007825-e007825. [doi: 10.1136/bmjopen-2015-007825] [Medline: 25783428]
27. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ 2017 May 23:j2099. [doi: 10.1136/bmj.j2099] [Medline: 28536104]
28. LaValley MP, Lo GH, Price LL, Driban JB, Eaton CB, McAlindon TE. Development of a clinical prediction algorithm for knee osteoarthritis structural progression in a cohort study: value of adding measurement of subchondral bone density. Arthritis Res Ther 2017 May 16;19(1):95 [FREE Full text] [doi: 10.1186/s13075-017-1291-3] [Medline: 28511690]
29. Boyd CM, Reider L, Frey K, Scharfstein D, Leff B, Wolff J, et al. The effects of guided care on the perceived quality of health care for multi-morbid older persons: 18-month outcomes from a cluster-randomized controlled trial. J Gen Intern Med 2010 Mar;25(3):235-242 [FREE Full text] [doi: 10.1007/s11606-009-1192-5] [Medline: 20033622]
30. Richardson LD, Asplin BR, Lowe RA. Emergency department crowding as a health policy issue: past development, future directions. Ann Emerg Med 2002 Oct;40(4):388-393. [Medline: 12239493]
31. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D. Early detection of impending physiologic deterioration among patients who are not in intensive care: development of predictive models using data from an automated electronic medical record. J Hosp Med 2012 Mar;7(5):388-395. [doi: 10.1002/jhm.1929] [Medline: 22447632]
32. Dummett BA, Adams C, Scruth E, Liu V, Guo M, Escobar GJ. Incorporating an early detection system into routine clinical practice in two community hospitals. J Hosp Med 2016 Nov;11 Suppl 1:S25-S31 [FREE Full text] [doi: 10.1002/jhm.2661] [Medline: 27805798]
33. Palacholla RS, Fischer NC, Agboola S, Nikolova-Simons M, Odametey S, Golas SB, et al. Evaluating the impact of a web-based risk assessment system (CareSage) and tailored interventions on health care utilization: protocol for a randomized controlled trial. JMIR Res Protoc 2018 May 09;7(5):e10045 [FREE Full text] [doi: 10.2196/10045] [Medline: 29743156]

XSL•FO
RenderX

34. Stokke R. The personal emergency response system as a technology innovation in primary health care services: an integrative review. J Med Internet Res 2016 Jul 14;18(7):e187 [FREE Full text] [doi: 10.2196/jmir.5727] [Medline: 27417422]

Abbreviations

AUC: area under the receiver operating characteristic curve
CHF: congestive heart failure
COPD: chronic obstructive pulmonary disease
ED: emergency department
EHR: electronic health record
EMS: emergency medical service
NPV: negative predictive value
PERS: personal emergency response system
PHH: Partners HealthCare at Home
PPV: positive predictive value

Edited by G Eysenbach; submitted 22.01.18; peer-reviewed by S Khairat, T Guetterman, SH Kang; comments to author 21.03.18; revised version received 20.07.18; accepted 07.08.18; published 27.11.18.

Please cite as:
op den Buijs J, Simons M, Golas S, Fischer N, Felsted J, Schertzer L, Agboola S, Kvedar J, Jethwani K
Predictive Modeling of 30-Day Emergency Hospital Transport of Patients Using a Personal Emergency Response System: Prognostic Retrospective Study
JMIR Med Inform 2018;6(4):e49
URL: http://medinform.jmir.org/2018/4/e49/
doi:10.2196/medinform.9907
PMID:30482741

©Jorn op den Buijs, Mariana Simons, Sara Golas, Nils Fischer, Jennifer Felsted, Linda Schertzer, Stephen Agboola, Joseph Kvedar, Kamal Jethwani. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 27.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Utilization of Electronic Medical Records and Biomedical Literature to Support the Diagnosis of Rare Diseases Using Data Fusion and Collaborative Filtering Approaches

Feichen Shen[1], PhD; Sijia Liu[1], MS; Yanshan Wang[1], PhD; Andrew Wen[1], MS; Liwei Wang[1], MD, PhD; Hongfang Liu[1], PhD

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

Corresponding Author:
Hongfang Liu, PhD
Department of Health Sciences Research
Mayo Clinic
200 1st Street SW
Rochester, MN, 55905
United States
Phone: 1 5072930057
Fax: 1 5072841516
Email: liu.hongfang@mayo.edu

Abstract

Background: In the United States, a rare disease is characterized as the one affecting no more than 200,000 patients at a certain period. Patients suffering from rare diseases are often either misdiagnosed or left undiagnosed, possibly due to insufficient knowledge or experience with the rare disease on the part of clinical practitioners. With an exponentially growing volume of electronically accessible medical data, a large volume of information on thousands of rare diseases and their potentially associated diagnostic information is buried in electronic medical records (EMRs) and medical literature.

Objective: This study aimed to leverage information contained in heterogeneous datasets to assist rare disease diagnosis. Phenotypic information of patients existed in EMRs and biomedical literature could be fully leveraged to speed up diagnosis of diseases.

Methods: In our previous work, we advanced the use of a collaborative filtering recommendation system to support rare disease diagnostic decision making based on phenotypes derived solely from EMR data. However, the influence of using heterogeneous data with collaborative filtering was not discussed, which is an essential problem while facing large volumes of data from various resources. In this study, to further investigate the performance of collaborative filtering on heterogeneous datasets, we studied EMR data generated at Mayo Clinic as well as published article abstracts retrieved from the Semantic MEDLINE Database. Specifically, in this study, we designed different data fusion strategies from heterogeneous resources and integrated them with the collaborative filtering model.

Results: We evaluated performance of the proposed system using characterizations derived from various combinations of EMR data and literature, as well as with sole EMR data. We extracted nearly 13 million EMRs from the patient cohort generated between 2010 and 2015 at Mayo Clinic and retrieved all article abstracts from the semistructured Semantic MEDLINE Database that were published till the end of 2016. We applied a collaborative filtering model and compared the performance generated by different metrics. Log likelihood ratio similarity combined with k-nearest neighbor on heterogeneous datasets showed the optimal performance in patient recommendation with area under the precision-recall curve (PRAUC) 0.475 (string match), 0.511 (systematized nomenclature of medicine [SNOMED] match), and 0.752 (Genetic and Rare Diseases Information Center [GARD] match). Log likelihood ratio similarity also performed the best with mean average precision 0.465 (string match), 0.5 (SNOMED match), and 0.749 (GARD match). Performance of rare disease prediction was also demonstrated by using the optimal algorithm. Macro-average *F*-measure for string, SNOMED, and GARD match were 0.32, 0.42, and 0.63, respectively.

Conclusions: This study demonstrated potential utilization of heterogeneous datasets in a collaborative filtering model to support rare disease diagnosis. In addition to phenotypic-based analysis, in the future, we plan to further resolve the heterogeneity issue and reduce miscommunication between EMR and literature by mining genotypic information to establish a comprehensive disease-phenotype-gene network for rare disease diagnosis.

XSL•FO
RenderX

(JMIR Med Inform 2018;6(4):e11301) doi:10.2196/11301

KEYWORDS

electronic medical record; literature; text mining; rare diseases

Introduction

Background

In the United States, a rare disease is described as the one affecting no more than 200,000 patients at a certain time [1]. Currently, there are nearly 10% Americans suffering from rare diseases [2]. However, patients often are misdiagnosed or left undiagnosed because of insufficient clinical knowledge and experience. Furthermore, merely 5% of these diseases have treatment plans [2]. Therefore, accelerating rare disease diagnostic decision support is crucial and urgent.

The very initial step in diagnosing rare disease is to stratify patients into subgroups with similar phenotypic characterizations. In addition, with computationally accessible medical data growing at an exponential rate, an abundance of rare disease-related phenotypic information is believed to be buried in electronic medical records (EMRs) and medical literature. Therefore, we hypothesize that patients' phenotypic information available among these resources can be leveraged to accelerate disease diagnosis. Few studies focus on phenotypic characterization of diseases and the analysis of phenotype-disease associations from free-text data such as EMRs and medical literature. One of the most representative efforts, the Human Phenotype Ontology (HPO) [3] was built to collect human phenotypic information for the differential diagnosis of rare diseases. In our previous work, we leveraged the HPO to annotate a large collection of clinical narratives and demonstrated a use case by using an annotation pipeline to perform knowledge discovery on Wilson disease [4]. We also proposed the use of collaborative filtering in our previous study for rare disease diagnosis [5], as making diagnostic decision making for a patient based on phenotype is similar to recommending a similar online product according to customers' previous purchases in e-commerce [6-8].

Since all datasets are flawed, it is important to prepare data with good quality, as machine learning depends heavily on data [9]. Especially for collaborative filtering algorithm, a proper preparation of data can largely avoid key information loss and improve learning performance [10]. More challenges come into the picture while feeding heterogeneous data into collaborative filtering model.

Previous Work

One of these challenges is the alignment of semantic heterogeneity. Semantic heterogeneity is referred to as a situation where 2 or more datasets are provided by different parties with various perspectives and purposes [11]. For structured data, the fusion of heterogeneous data is difficult due to inconsistent data models, data schemas, query languages, and terminology [12]. For unstructured or semistructured data, such issues are exacerbated as schemas must become much more flexible to accommodate the nonstandardized data and as such semantic drift becomes a more significant problem. Some studies have focused on making good semantic alignment across heterogeneous data. For example, MedKDD is a system for integrating and aligning heterogeneous biomedical ontologies [13]. Bache et al targeted on identifying patient cohort from heterogeneous resources by developing an adaptive query model [14]. Bleich et al made a comparison between integrated and interfaced hospital systems [15]. Burkle et al conducted a study to transfer data stored in one electronic patient record to another health care information system [16]. EHR4CR demonstrates an interoperable way to reuse electronic health records [17]. Mate et al conducted a study on integrating ontology data between clinical and research systems [18]. SHRINE provides a platform for disease studies across multiple health care institutions [19]. Ohmann et al proposed an overview of studies on data interoperability of basic research, clinical research, and medical data [20].

Another challenge is to get benefit from heterogeneous data to improve performance of machine learning. To investigate this, Lewis et al applied support vector machine on heterogeneous biological data to infer gene function [21]. Yu et al introduced a l2-norm multiple kernel learning algorithm and applied it on biomedical data fusion [22]. Ye et al showed a study on Alzheimer disease using heterogeneous data fusion [23]. Wang et al made a comparison among clinical notes, biomedical literature, and their combination to test their performances with word embeddings [24]. Torii et al showed the performance for concept extraction using machine learning taggers across narratives from heterogeneous data sources [25]. A GOstruct extension was developed to annotate protein functions from heterogeneous data [26].

Objective

According to the aforementioned related work, although some success was demonstrated, the issue regarding semantic heterogeneity is still an unsolved puzzle. Moreover, to the best of our knowledge, no study has paid attention to the impact of applying collaborative filtering on heterogeneous data, especially in biomedical domain. Therefore, it is interesting to investigate how data fusion strategies on heterogeneous resources can work with collaborative filtering for an optimal recommendation.

In this work, we developed a new framework based on our previous designed collaborative filtering system to incorporate heterogeneous data sources with different data fusion strategies to assist in diagnosing rare diseases. We extracted Unified Medical Language System concepts with MetaMap [27] and applied the HPO with the Genetic and Rare Diseases Information Center (GARD) [28] to annotate clinical notes at the Mayo Clinic generated from 2010 to 2015 as well as research articles stored in the Semantic MEDLINE Database (SemMedDB) [29] published up to December 2016. We integrated different data fusion strategies with collaborative filtering and evaluated their performances for patient recommendation and rare disease prediction.

XSL•FO
RenderX

Methods

Data Collection

For the EMR dataset, we collected clinical notes generated at the Mayo Clinic from 2010 to 2015. The extracted corpus maintained about 13 million unstructured clinical notes for over 700,000 patients. We only annotated sections with problems and diagnoses. For the medical literature dataset, we extracted abstracts of research articles from the SemMedDB. We then used HPO and GARD terms to match either subject or object for each predication [29] and finally came up with 91,680 phenotype-rare disease associations to process.

Collaborative Filtering Model for Rare Disease Recommendation

In e-commerce, collaborative filtering techniques [30] are popularly applied to recommend products to a customer based on customers with similar purchase preferences and other interests. Diagnosing a patient with a disease based on patients' phenotypic information is very similar to recommending a product to a customer; therefore, it is natural to propose the use of collaborative filtering for disease diagnosis.

In our previous work, we developed a collaborative filtering model based on a cohort of rare disease patients to stratify patients into subgroups and accelerate the diagnosis of rare diseases. Here, we treated patient profiles with their respective phenotypes as binary inputs, which means that the patient either has or does not have a phenotype. For the patients with a confirmed rare disease diagnosis, we used their phenotypes as input and treated their rare disease diagnosis as labels to train the collaborative filtering model.

Specifically, we applied the Tanimoto coefficient similarity (TANI), overlap coefficient similarity (OL), Fager & McGowan coefficient similarity (FMG), and log likelihood ratio similarity (LL) as 4 measurements to compute patient similarity [5]. For any 2 patients m and n, $|P_m|$ and $|P_n|$ denote the number of phenotypes each patient has, and TANI, OL, FMG, and LL are described as shown in Equations 1, 2, 3, and 4, respectively.

We also applied 2 neighborhood algorithms to provide recommendations: k-nearest neighbors (KNN) and threshold patient neighbor (TPN) [5]. Detailed steps of identifying neighborhood for 2 approaches are shown in Textbox 1.

Semantic MEDLINE Database

SemMedDB is a repository of semantic predications (ie, subject-predicate-object triples) extracted from the titles and abstracts of all PubMed citations [29,31-33]. In this study, we used SemMedDB Version 25, which contains more than 84 million predications (ie, associations) between concepts retrieved from abstracts of over 25 million PubMed-indexed publications [34].

Human Phenotype Ontology

The HPO is a standardized vocabulary for phenotypic terms, and it is built based on collecting phenotypic knowledge from various biomedical literature as well as databases. In this study, we used HPO released in September 2016 to annotate phenotypic terms.

Genetic and Rare Diseases Information Center

The GARD is a database that contains information on rare diseases. It groups collected 4560 diseases into 32 disease categories. In this study, we used the GARD to extract rare disease terms.

Textbox 1. Algorithm 1-Neighborhood identification.

Input: Sorted Similarity Score Map S (Neighbor_Patient, Score) for each patient, number of neighbor k, similarity threshold t

Output for KNN: Neighbor List LK

Output for TPN: Neighbor List LT

1. FOR each neighbor_patient NP in S
2. $score_{np}$=S.get(NP)
3. IF (LK.size()<k)
4. add NP into LK
5. IF ($score_{np}$>t)
6. add NP into LT
7. RETURN LK, LT

Learning Methods

Figure 1 illustrates the system workflow of our study. The proposed system is able to absorb heterogeneous data sources, and it adapted the collaborative filtering model on any type of input for rare disease recommendation in a general manner. For EMRs, we leveraged the developed annotation pipeline to collect all phenotypic information mentioned within 1 year of the first appearance of the rare disease [4]. For medical literature, we first retrieved all predications from SemMedDB and saved them with PMID. We looked up HPO and GARD glossaries to check each predication (subject, predicate, object) and filtered out those sentences in which neither subject nor object could be found. To exclude disease-disease and phenotype-phenotype associations, we also filtered out those predications in which both the subject and the object could be mapped to the same

XSL•FO
RenderX

vocabulary (GARD or HPO) and only kept the associations between phenotypes and rare diseases.

The format of input data is composed of patient identification or PMID and unique phenotypes manifested by each patient or article. We treated a positive diagnosis of a rare disease as a gold standard for association tasks involving patients and PMID to rare disease mentions as a gold standard for literature association tasks. We used 3 different data fusion strategies to prepare homogeneous and heterogeneous resources:

1. EMR only: Only patient-phenotype information extracted from the EMR was used.
2. EMR and literature (EMR+L): We first conducted a treatment on medical literature. Since each publication might only mention 1 phenotype with 1 rare disease, to strengthen the evidence power provided by the literature, we merged multiple literature sources together as 1 large document if those sources shared the same rare disease. Therefore, the number of documents used will be less than 91,680. We then mixed patient-phenotype association with literature-phenotype information and randomly permuted them without any additional treatment. Detailed steps of this process are shown as case 1 in Textbox 2.
3. EMR and pruned literature (EMR+PL). A similar approach as EMR+L was followed, but some phenotype-rare disease associations mined from literature were additionally filtered out if they did not appear in the EMR. In this case, we tried to enhance the correlation and coexisting evidence between phenotypes and rare diseases a bit further to provide a better prediction output. Case 2 in Textbox 2 demonstrates this pruned process.

Different phenotype-disease associations with 3 different data fusion strategies were imported to collaborative filtering model and the final recommendation outputs based on 3 data inputs would be given. For example, if a new patient has phenotypes *crystalline retinopathy*, *optic neuropathy*, *nephrocalcinosis*, and *cysteine stones*, 3 different disease recommendations (*kidney stone*, *calcium oxalate nephrolithiasis*, and *primary hyperoxaluria*) will be made, and we compared them with such patient's true diagnostic results for evaluation purpose.

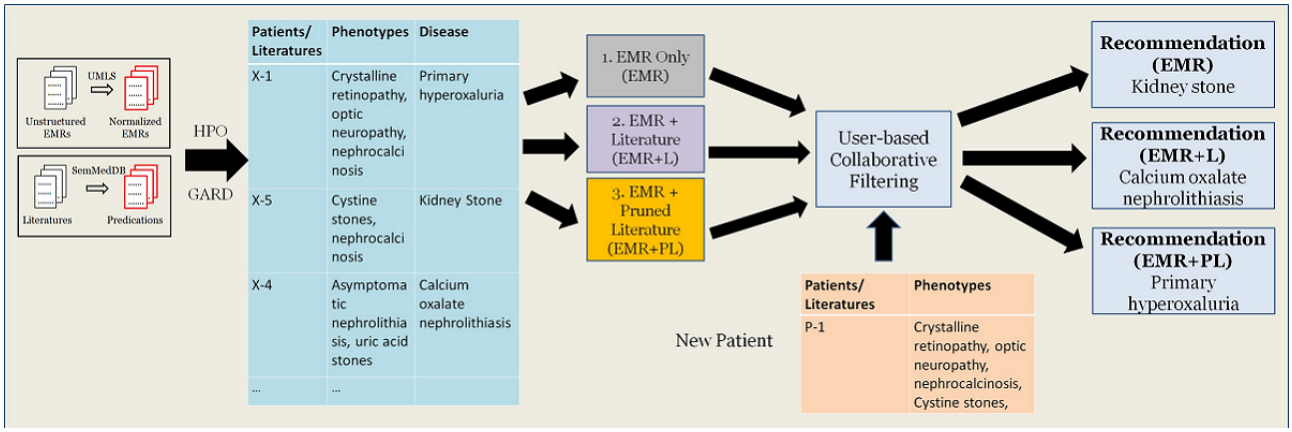
Evaluation

We evaluated 24 various evaluation groups as: (1) TANI with KNN on EMR; (2) TANI with KNN on EMR and literature; (3) TANI with KNN on EMR and pruned literature; (4) TANI with TPN on EMR; (5) TANI with TPN on EMR and literature; (6) TANI with TPN on EMR and pruned literature; (7) LL with KNN on EMR; (8) LL with KNN on EMR and literature; (9) LL with KNN on EMR and pruned literature; (10) LL with TPN on EMR; (11) LL with TPN on EMR and literature; (12) LL with TPN on EMR and pruned literature; (13) OL with KNN on EMR; (14) OL with KNN on EMR and literature; (15) OL with KNN on EMR and pruned literature; (16) OL with TPN on EMR; (17) OL with TPN on EMR and literature; (18) OL with TPN on EMR and pruned literature; (19) FMG with KNN on EMR; (20) FMG with KNN on EMR and literature; (21) FMG with KNN on EMR and pruned literature; (22) FMG with TPN on EMR; (23) FMG with TPN on EMR and literature; and (24) FMG with TPN on EMR and pruned literature.

We used the same metrics adopted in our previous work to evaluate system performance. Specifically, we applied root mean square error (RMSE) [35] to determine the optimal thresholds for KNN and TPN. We evaluated performances of patient recommendations using precision-recall curve and mean average precision [5]. We also evaluated disease prediction performance with precision recall and *F*-measure [5].

Similar to our previous study, we used 3 matching strategies to measure the similarity between any 2 rare diseases: string matching, systematized nomenclature of medicine-clinical terms (SNOMED) matching, and GARD matching to provide different levels of relaxation on predicting rare diseases [5].

Figure 1. System workflow. EMR: electronic medical record; UMLS: Unified Medical Language System.



XSL•FO
RenderX

Textbox 2. Algorithm 2-Heterogeneous Data Fusion for electronic medical record and literature.

```
Input: Map A (PMID, Rare Disease), Map B (PMID, Map(Rare Disease, List(Phenotypes)))
Output for Case 1: Merged literature with same rare disease, stored rare diseases along with their associated phenotypes in Map C
Output for Case 2: Pruned Map C'
Case 1: EMR+L
1. FOR each PMID and Rare Disease RD in A
2.     retrieve all relevant phenotypes {P} for RD and PMID from B
3.     IF C does not contain RD
4.         create new document_ID
5.         add {P} to list L
6.         add (document_ID, (RD, L)) to C
7.     ELSE
8.         List L=A.retrieve(document_ID)
9.         add nonduplicate elements from {P} to list L
10.        add (document_ID, (RD, L)) to C
11. RETURN C
Case 2: EMR+PL
12. C'=C
13. FOR each phenotype-disease pair PD1 in Map E
14.     FOR each phenotype-disease pair PD2 in Map C'
15.         IF (PD1 !=PD2)
16.         remove PD2 from C'
17. RETURN C'
```

Results

As shown in Table 1, after eliminating rare diseases that affect only 1 patient, there were 38,607 patients for EMR only, 40,241 patients for EMR and literature, and 39,677 patients for EMR and pruned literature datasets. Since EMR+L is mixed data without any refinement, the total number of phenotypes, rare diseases, and their associations are larger than the other 2 outputs. In addition, of the 32 possible GARD categories, we found that the number of GARD categories covered for each of the 3 outputs were 28, 31, and 28, respectively.

Threshold Selection With Root Mean Square Error

For KNN combined with different similarity measurements, Figure 2 plots the curve to illustrate the change of RMSE associated with different number of selected neighbors. We observed that for LL and OL, RMSE was more sensitive to EMR+L and EMR+PL than to EMR only, which shows that adding sources of literature might affect the results in either a positive or negative way. On the other hand, the change of RMSE for TANI and FMG was minimal among these 3 datasets, which indicates that literature enrichment did not reflect markedly on the performance for these 2 algorithms.

Table 1. Statistics for prepared datasets.

Datasets	EMR[a] only, n	EMR and literature (EMR+L), n	EMR and pruned literature (EMR+PL), n
Patients or literature sources	38,607	40,241	39,677
Phenotypes	3271	3818	3271
Rare diseases	1074	1634	1074
Phenotype-disease associations	141,036	154,802	141,036
GARD[b] categories covered	28	31	28

[a]EMR: electronic medical record.

[b]GARD: Genetic and Rare Diseases Information Center.

XSL•FO
RenderX

Figure 2. Root mean square error (RMSE) for k-nearest neighbors (KNN) with four similarity measurements. EMR: electronic medical record; FMG: Fager and McGowan coefficient similarity; L: literature; LL: log likelihood ratio similarity; OL: overlap coefficient similarity; PL: pruned literature; TANI: Tanimoto coefficient similarity.

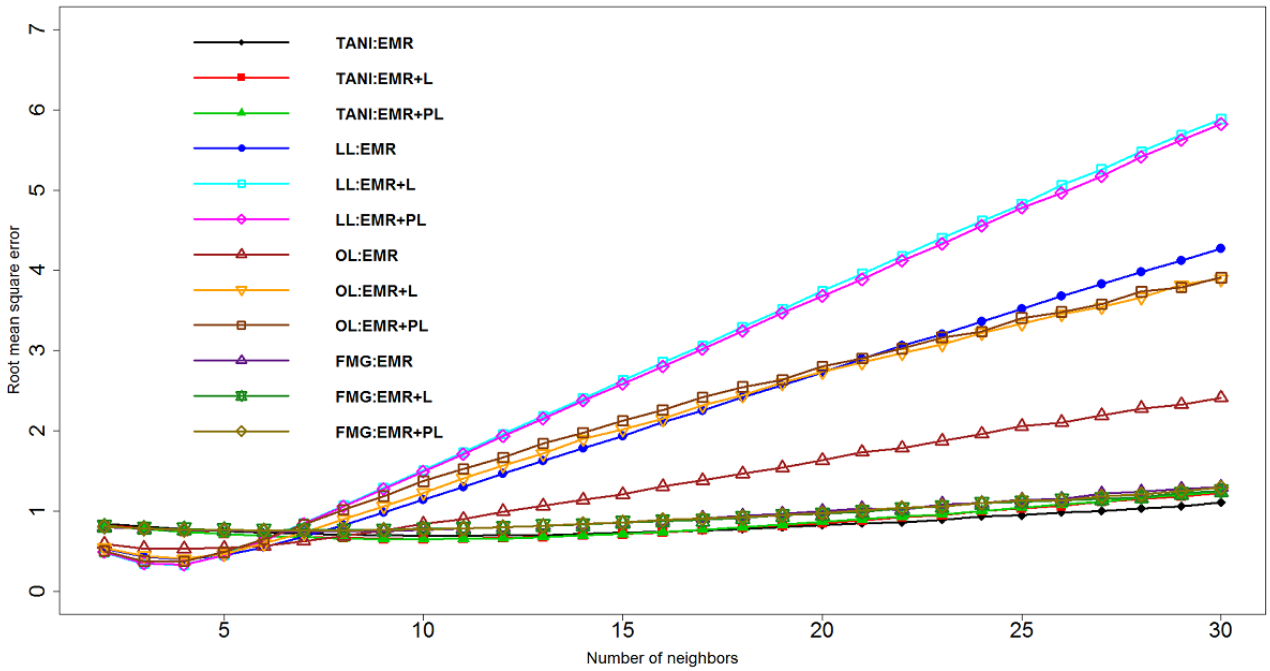


Figure 3 describes how RMSE changes for the tested coefficient similarities with TPN. We found that TANI is sensitive to a smaller threshold but tends to be balanced with the threshold getting larger for all 3 datasets. RMSE for LL remained balanced until similarity threshold became larger, denoting that LL is not sensitive to the similarity threshold. All 4 algorithms held a higher average RMSE with EMR+L but a lower average RMSE with EMR and EMR+PL, indicating that the random mix of EMR and literature might not be able to provide a strong scheme for rare disease prediction and recommendation. Specifically for OL, EMR only performed better than the other 2 datasets, which showed that OL is not a very suitable measurement for heterogeneous datasets. Table 2 summarizes the optimal threshold selection for different evaluation groups.

Performance for Patient Recommendation

We plotted precision-recall curves for each of the 24 experiments and area under the precision-recall curve (PRAUC) for each matching criterion. Overall, we observed that GARD matching contributed to the optimal performance among all matching criteria, and SNOMED semantic matching was always a suboptimal strategy. Figure 4 shows the performance of TANI with KNN and TPN on different datasets and matching criteria. We observed that there are no considerable differences between TANI+KNN and TANI+TPN for 3 matching criteria with 3 datasets. Although the difference seems subtle, TANI+TPN with EMR+PL yielded the optimal PRAUC score for string, SNOMED, and GARD matching, respectively. Figure 5 shows the performance of LL. Compared with TANI, LL performed worse with TPN for all datasets and matching criteria. However, when using KNN, although LL performed worse with EMR data only, it outperformed for both EMR+L and EMR+PL. This result indicates that LL is more suitable for mining knowledge from heterogeneous datasets than TANI. Figure 6 illustrates the performance of OL. Compared with TANI and LL, this measurement produced considerably lower PRAUC for either neighborhood algorithm. Additionally, OL yielded better performance with EMR data only but worse performance with combined datasets, which indicates that OL may be more suitable for a single dataset, and it suggests that combined datasets might possess too much noise for OL to make an accurate judgment. Although OL cannot handle literature-enriched data well, we observed that pruned literature still performed better than nonpruned literature. Figure 7 depicts the reaction of FMG to different neighborhood algorithms and combinations of datasets. Similar to OL, FMG with EMR data only outperformed EMR+L and EMR+PL in all 3 matching criteria. However, unlike OL, although FMG with EMR+L had the worst performance with both KNN and TPN, pruned literature slightly increased the performance, and no substantial difference exists between using FMG with EMR only and EMR+PL. Tables 3-6 show MAP for all patients' recommendations, which showed a consistent performance with PRAUC evaluation, indicating that TANI and LL performed better and are more suitable for integrated EMR and literature, whereas OL and FMG performed worse and are not suitable for fused datasets. In general, optimal performance produced by LL indicated the potential of combining EMR and literature to increase patient stratification.

XSL•FO
RenderX

Figure 3. Root mean square error (RMSE) for threshold patient neighbor (TPN) with four similarity measurements. EMR: electronic medical record; FMG: Fager and McGowan coefficient similarity; L: literature; LL: log likelihood ratio similarity; OL: overlap coefficient similarity; PL: pruned literature; TANI: Tanimoto coefficient similarity.

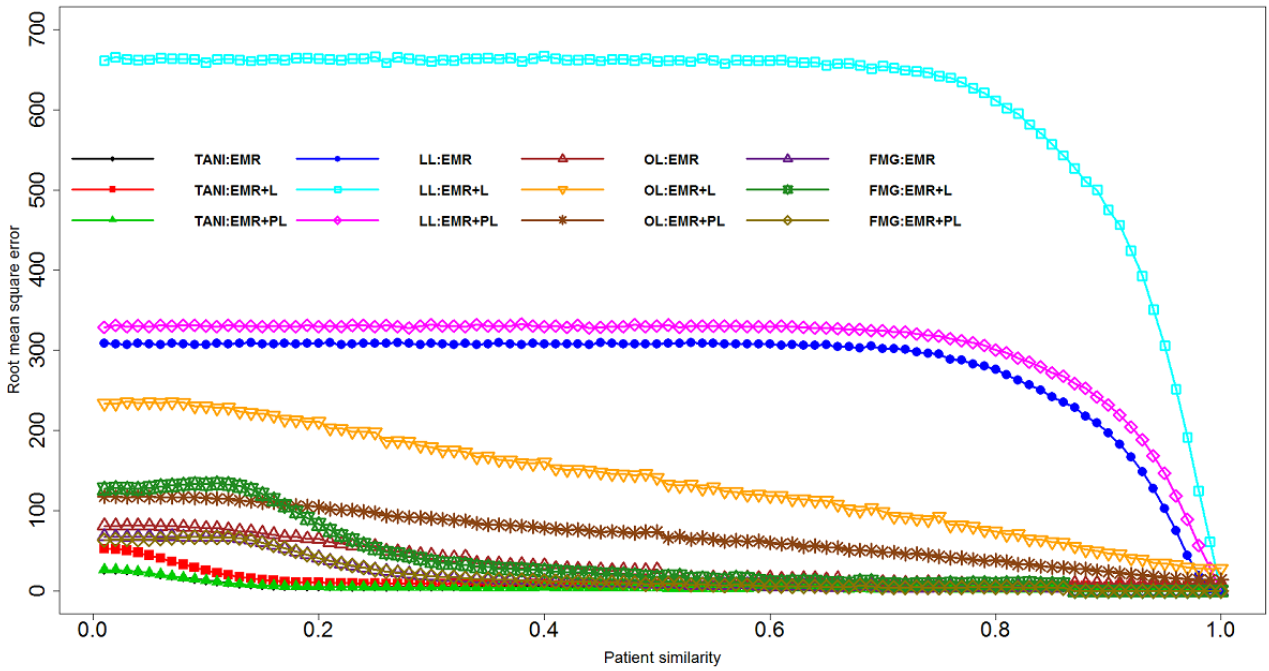


Table 2. Optimal thresholds for different evaluation groups.

Optimal parameters	TANI[a]			LL[b]			OL[c]			FMG[d]		
	EMR[e]	EMR+L[f]	EMR+PL[g]	EMR	EMR+L	EMR+PL	EMR	EMR+L	EMR+PL	EMR	EMR+L	EMR+PL
Optimal *k* (KNN[h])	11	10	9	4	4	4	4	4	4	7	6	6
Optimal *t* (TPN[i])	0.19	0.19	0.2	0.72	0.73	0.76	0.51	0.49	0.51	0.12	0.11	0.12

[a]TANI: Tanimoto coefficient similarity.

[b]LL: log likelihood ratio similarity.

[c]OL: overlap coefficient similarity.

[d]FMG: Fager and McGowan coefficient similarity.

[e]EMR: electronic medical record.

[f]L: literature.

[g]PL: pruned literature.

[h]KNN: k-nearest neighbors.

[i]TPN: threshold patient neighbor.

XSL•FO
RenderX

Figure 4. Precision-recall curves and area under the precision-recall curve (PRAUC) for Tanimoto coefficient similarity (TANI) with k-nearest neighbors (KNN) and threshold patient neighbors (TPN). EMR: electronic medical record; GARD: Genetic and Rare Diseases Information Center; KNN: k-nearest neighbors; SNOMED: systematized nomenclature of medicine; TANI: Tanimoto coefficient similarity.

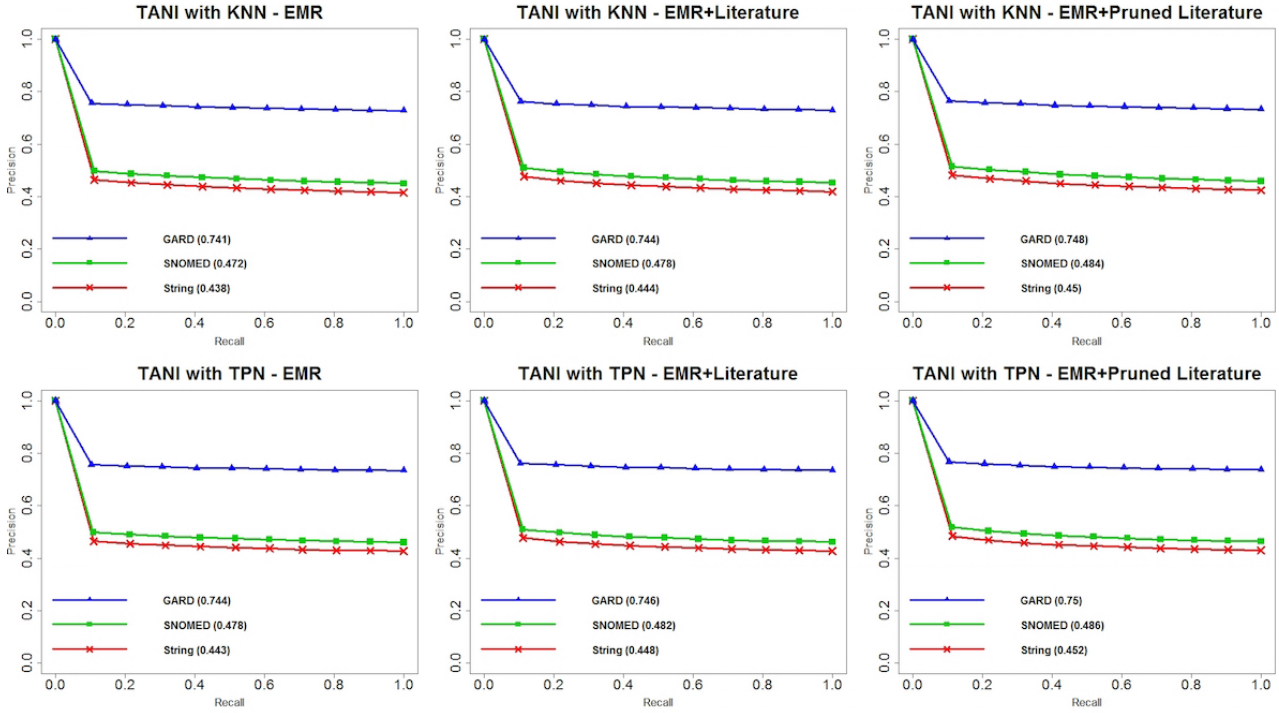
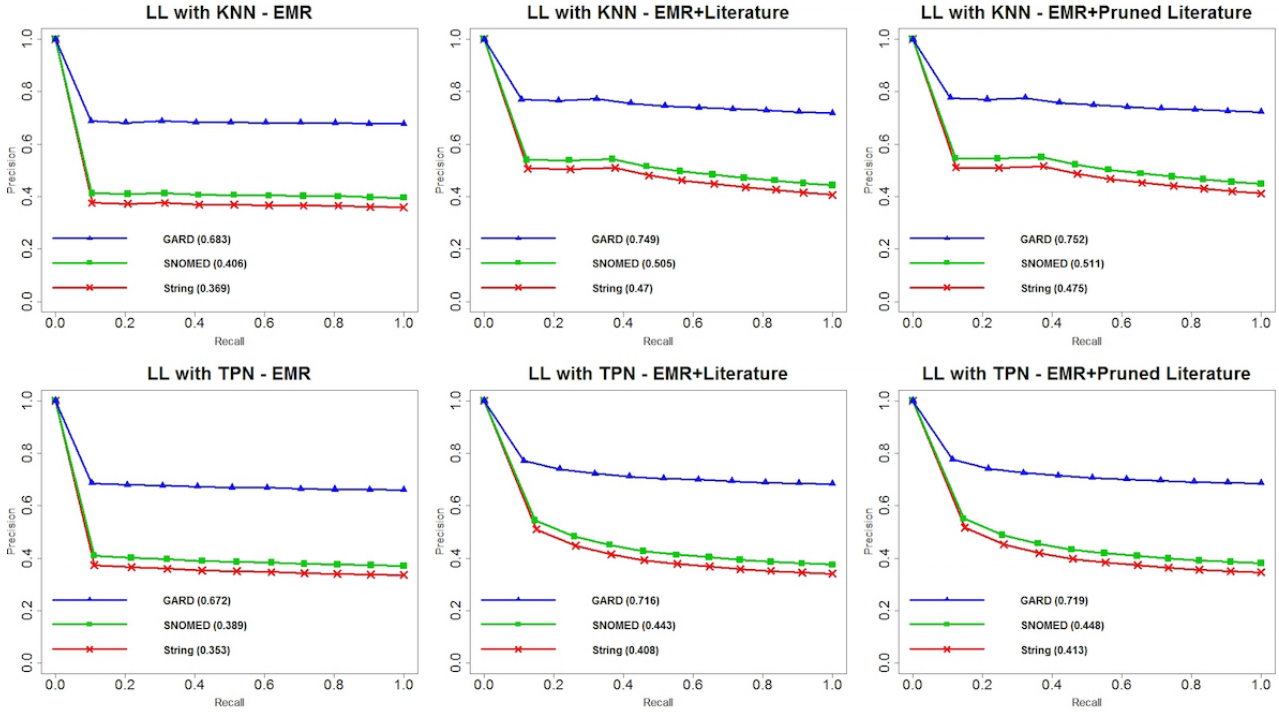


Figure 5. Precision-recall curves and area under the precision-recall curve (PRAUC) for log likelihood ratio similarity with k-nearest neighbors and threshold patient neighbors. EMR: electronic medical record; GARD: Genetic and Rare Diseases Information Center; KNN: k-nearest neighbors; LL: log likelihood ratio similarity; SNOMED: systematized nomenclature of medicine; TPN: threshold patient neighbor.



XSL•FO
RenderX

Figure 6. Precision-recall curves and area under the precision-recall curve (PRAUC) for overlap coefficient similarity with k-nearest neighbors and threshold patient neighbors. EMR: electronic medical record; GARD: Genetic and Rare Diseases Information Center; KNN: k-nearest neighbors; OL: overlap coefficient similarity; SNOMED: systematized nomenclature of medicine; TPN: threshold patient neighbor.

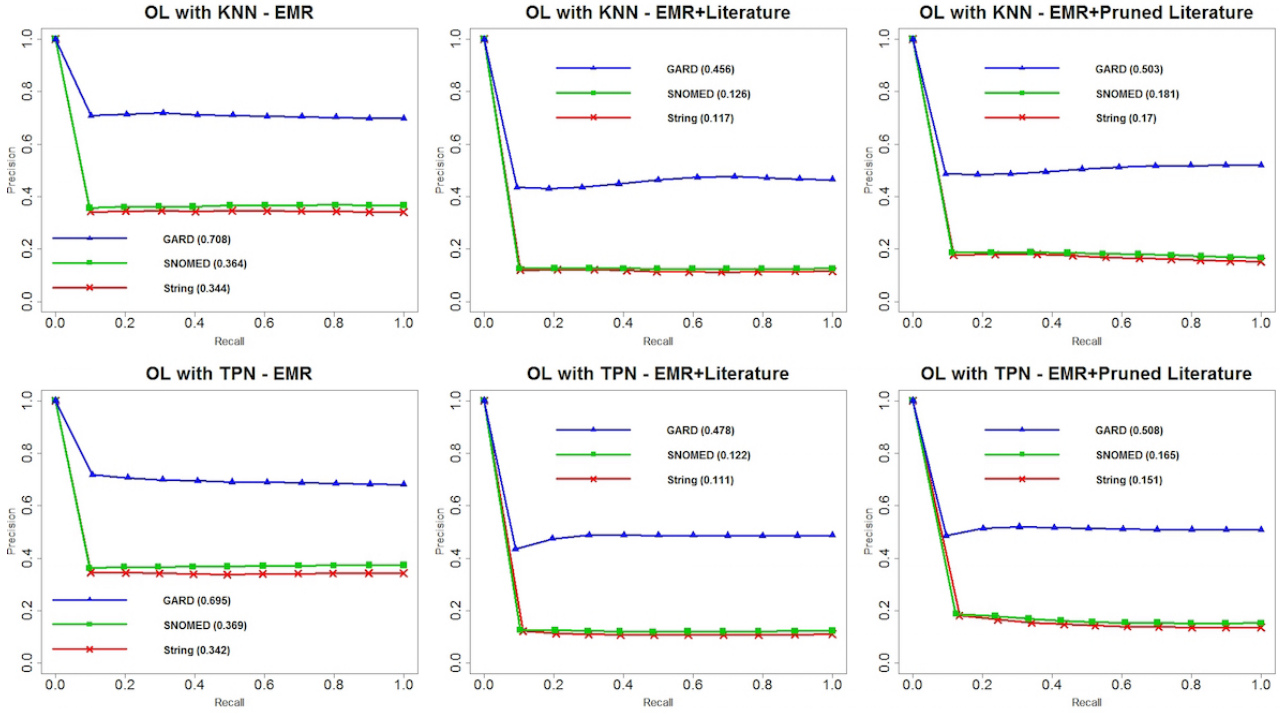
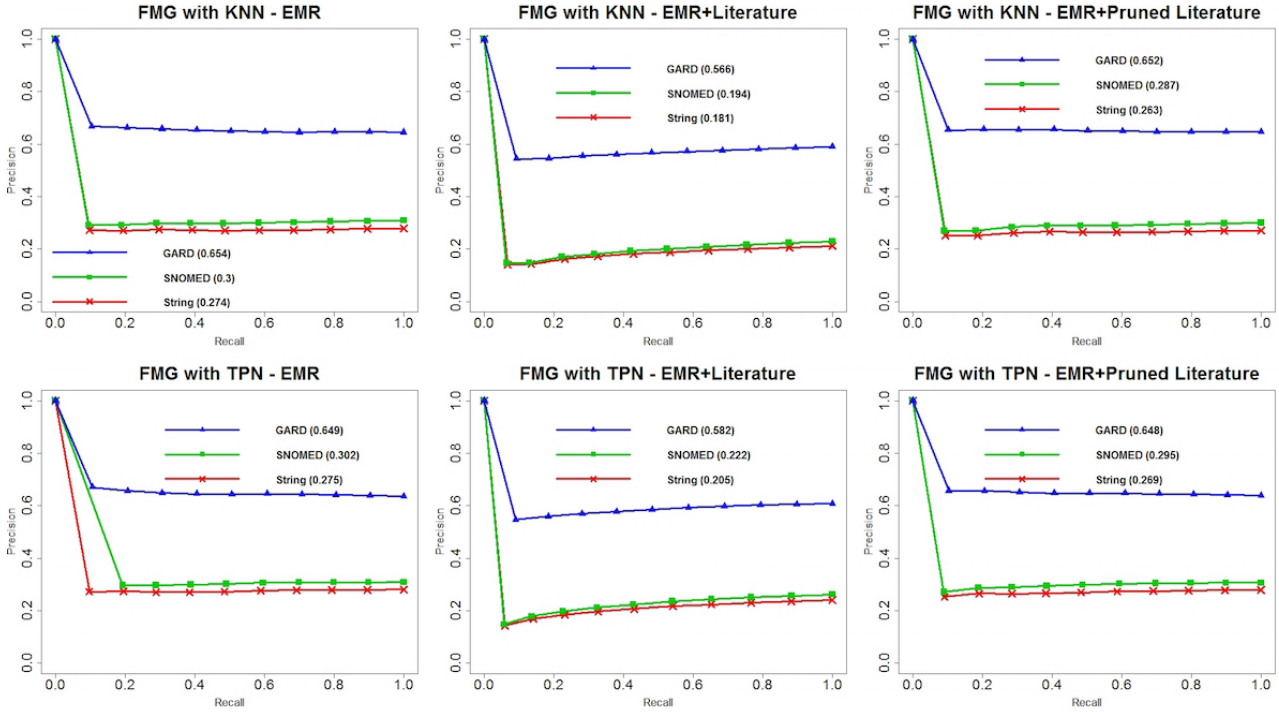


Figure 7. Precision-recall curves and area under the precision-recall curve (PRAUC) for Fager and McGowan coefficient similarity with k-nearest neighbors and threshold patient neighbors. EMR: electronic medical record; FMG: Fager and McGowan coefficient similarity; GARD: Genetic and Rare Diseases Information Center; KNN: k-nearest neighbors; SNOMED: systematized nomenclature of medicine; TPN: threshold patient neighbor.



XSL•FO
RenderX

Table 3. Mean average precision for TANI[a] with EMR[b], EMR+L[c], and EMR+PL[d] (optimal in italics).

Matching criterion	EMR		EMR+L		EMR+PL	
	KNN[e]	TPN[f]	KNN	TPN	KNN	TPN
String	0.435	0.441	0.436	0.445	*0.446*	*0.448*
SNOMED[g]	0.469	0.475	0.474	0.479	*0.481*	*0.483*
GARD[h]	0.739	0.742	0.742	0.745	*0.746*	*0.748*

[a]TANI: Tanimoto coefficient similarity.
[b]EMR: electronic medical record.
[c]L: literature.
[d]PL: pruned literature.
[e]KNN: k-nearest neighbors.
[f]TPN: threshold patient neighbor.
[g]SNOMED: systematized nomenclature of medicine.
[h]GARD: Genetic and Rare Diseases Information Center.

Table 4. Mean average precision for LL[a] with EMR[b], EMR+L[c], and EMR+PL[d] (optimal in italics).

Matching criterion	EMR		EMR+L		EMR+PL	
	KNN[e]	TPN[f]	KNN	TPN	KNN	TPN
String	0.368	0.351	0.46	0.391	*0.465*	*0.396*
SNOMED[g]	0.405	0.386	0.495	0.426	*0.5*	*0.431*
GARD[h]	0.683	0.67	0.745	0.71	*0.749*	*0.713*

[a]LL: log likelihood ratio similarity.
[b]EMR: electronic medical record.
[c]L: literature.
[d]PL: pruned literature.
[e]KNN: k-nearest neighbors.
[f]TPN: threshold patient neighbor.
[g]SNOMED: systematized nomenclature of medicine.
[h]GARD: Genetic and Rare Diseases Information Center.

Table 5. Mean average precision for OL[a] with EMR[b], EMR+L[c], and EMR+PL[d] (optimal in italics).

Matching criterion	EMR		EMR+L		EMR+PL	
	KNN[e]	TPN[f]	KNN	TPN	KNN	TPN
String	*0.344*	*0.342*	0.117	0.11	0.167	0.148
SNOMED[g]	*0.365*	*0.369*	0.126	0.122	0.179	0.162
GARD[h]	*0.708*	*0.693*	0.457	0.48	0.505	0.509

[a]OL: overlap coefficient similarity.
[b]EMR: electronic medical record.
[c]L: literature.
[d]PL: pruned literature.
[e]KNN: k-nearest neighbors.
[f]TPN: threshold patient neighbor.
[g]SNOMED: systematized nomenclature of medicine.
[h]GARD: Genetic and Rare Diseases Information Center.

XSL•FO
RenderX

Table 6. Mean average precision for FMG[a] with EMR[b], EMR+L[c], and EMR+PL[d] (optimal in italics).

Matching criterion	EMR		EMR+L		EMR+PL	
	KNN[e]	TPN[f]	KNN	TPN	KNN	TPN
String	*0.274*	*0.275*	0.18	0.205	0.264	0.27
SNOMED[g]	*0.3*	*0.302*	0.192	0.221	0.288	0.297
GARD[h]	*0.653*	*0.647*	0.568	0.584	0.651	*0.647*

[a]FMG: Fager and McGowan coefficient similarity.

[b]EMR: electronic medical record.

[c]L: literature.

[d]PL: pruned literature.

[e]KNN: k-nearest neighbors.

[f]TPN: threshold patient neighbor.

[g]SNOMED: systematized nomenclature of medicine.

[h]GARD: Genetic and Rare Diseases Information Center.

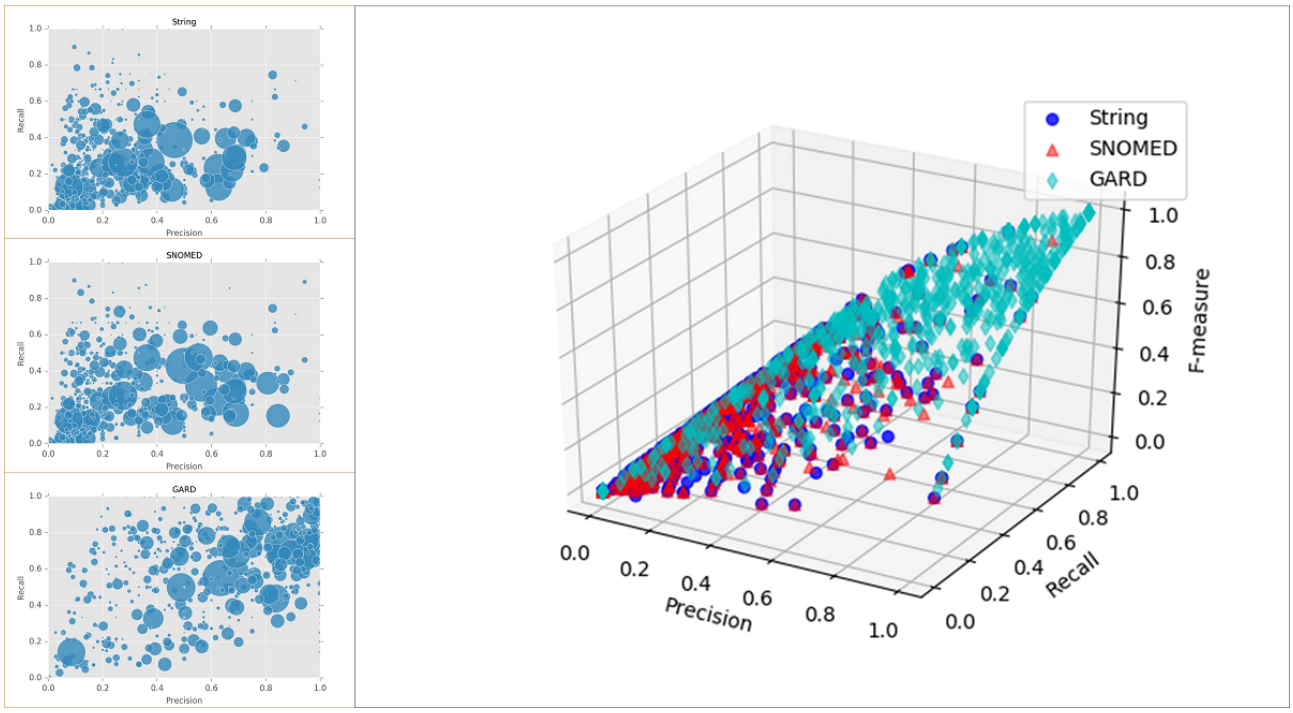
Performance on Rare Disease Prediction With Log Likelihood Ratio Similarity

We selected LL with KNN as the optimal metric, trained it with EMR+PL, and applied it on 44,060 patients with only 1 rare disease. We only selected rare diseases with at least 3 affected patients, which resulted in 702 rare diseases in total. Prediction performances for different matching criteria are described as shown in Figure 8. The circle size in two-dimensional scatter plots is proportional to the number of affected patients. Three-dimensional plot for precision, recall, and *F*-measure in Figure 8 clearly depicts that GARD outperformed SNOMED matching, and string matching yielded the worst performance. Macro-average *F*-measure for string, SNOMED, and GARD matching for the tested diseases were 0.32, 0.42, and 0.63, respectively.

In Table 7, we selected 9 diseases for each matching criterion for LL with KNN. Specifically, we picked 3 with high *F*-measures, 3 with medium to high *F*-measures, and 3 with relatively low *F*-measures. For any rare disease affecting no more than 10 cases, we marked them as <10.

For string matching, *holoprosencephaly*, *Huntington disease*, and *juvenile polyposis syndrome* contributed to higher *F*-measures and do not have a large number of affected patients. However, since they are unique, performance of recommendation was promising. *Sacrococcygeal teratoma*, *frontotemporal dementia*, and *polycystic liver disease* were well predicted but with some missed cases. Taking *sacrococcygeal teratoma* as an example, we found *neurogenic bladder*, *constipation*, and *diarrhea* to be the most common phenotypes that also occurred in patients with the rare disease *microcephaly*.

Figure 8. Prediction performance for rare diseases. GARD: Genetic and Rare Diseases Information Center; SNOMED: systematized nomenclature of medicine.



XSL•FO
RenderX

Table 7. Recommendation performance for selected rare diseases (3 high, 3 medium to high, 3 low).

Approaches and top diseases	Number of patients affected	Precision	Recall	*F*-measure
LL[a]+KNN[b] with string matching				
Holoprosencephaly	<10	0.75	1	0.86
Huntington disease	<10	1	0.67	0.8
Juvenile polyposis syndrome	<10	0.91	0.71	0.8
Sacrococcygeal teratoma	15	0.83	0.67	0.74
Frontotemporal dementia	202	0.69	0.58	0.63
Polycystic liver disease	72	0.64	0.58	0.61
Hemicrania continua	36	0.08	0.25	0.12
Intrahepatic cholangiocarcinoma	94	0.08	0.22	0.12
Neuromyelitis optica	50	0.16	0.1	0.12
LL+KNN with SNOMED[c] matching				
Myxoid liposarcoma	37	0.94	0.89	0.91
Linear scleroderma	16	0.91	0.71	0.8
Migraine with brainstem aura	15	0.75	1	0.86
Hypophosphatemic rickets	<10	0.83	0.75	0.79
Congenital radio ulnar synostosis	14	0.67	0.86	0.75
Spasmodic dysphonia	177	0.83	0.67	0.74
Acute graft-versus-host disease	20	0.1	0.5	0.15
Cryptogenic organizing pneumonia	37	0.14	0.17	0.15
Cerebellar degeneration	29	0.14	0.17	0.15
LL+KNN with GARD[d] matching				
Acrospiroma	<10	1	1	1
Birt-Hogg-Dube syndrome	<10	1	1	1
Dendritic cell tumor	<10	1	1	1
Acute promyelocytic leukemia	15	0.97	0.93	0.95
Migraine with brainstem aura	15	1	0.88	0.93
Thyroid cancer, anaplastic	30	1	0.86	0.92
Addison disease	34	0.88	0.45	0.6
Encephalocele	56	0.4	0.59	0.48
Mixed connective tissue disease	78	0.4	0.48	0.43

[a]LL: log likelihood ratio similarity.

[b]KNN: k-nearest neighbors.

[c]SNOMED: Systematized Nomenclature of Medicine.

[d]GARD: Genetic and Rare Diseases Information Center.

In our EMR, *sacrococcygeal teratoma* patients and *microcephaly* patients reported 140 cases of *neurogenic bladder*, 84 cases of *constipation*, and 84 cases of *diarrhea*. In medical literature, *neurogenic bladder*, *constipation*, and *diarrhea* are also the 3 top phenotypes found in microcephaly, and they appeared 34, 32, and 32 times, respectively. Considering evidence from heterogeneous data sources, *sacrococcygeal teratoma* was often predicted as *microcephaly*. *Hemicrania continua*, *intrahepatic cholangiocarcinoma*, and *neuromyelitis optica* are 3 diseases with relatively low *F*-measures. Although the number of affected patients for them is not small, they lack a unique group of phenotypic patterns to differentiate them from other diseases with similar phenotypes.

For SNOMED matching, the top predicted diseases are *myxoid liposarcoma*, *linear scleroderma*, and *microscopic polyangiitis*. Since we used the SNOMED semantic hierarchy to measure the similarity between 2 diseases, the prediction performance was slightly better than using string matching only. For example, in our results, *myxoid liposarcoma* was semantically the same as *liposarcoma*, *linear scleroderma* had the same meaning as

XSL•FO
RenderX

morphea, and *microscopic polyangiitis* was treated equally with *granulomatosis with polyangitis*. These 3 diseases are related to unique phenotypes, and as such, the prediction results were positive. For example, the phenotypes *soft tissue sarcoma* and *lymphedema* have a tight relationship with *myxoid liposarcoma*. In addition, 65 phenotypes from EMRs and literature were closely related to *linear scleroderma*, and we found that *headache* and *hemiatrophy* frequently appeared. We also found that *vasculitis* and *glomerulonephritis* often appeared along with *cicatricial pemphigoid*. *Hypophosphatemic rickets*, *congenital radio ulnar synostosis*, and *spasmodic dysphonia* also contributed to the positive recommendation results but with some minor prediction errors. For example, *hypophosphatemic rickets* was considered to be *lung adenocarcinoma* in a few cases, *congenital radio ulnar synostosis* was misdiagnosed as *esophageal atresia*, and *spasmodic dysphonia* and *trigeminal neuralgia* were sometimes mismatched. Finally, patient profiles from EMRs and literature content regarding *acute graft-versus-host disease*, *cryptogenic organizing pneumonia*, and *cerebellar degeneration* were not good enough for our model to conduct the prediction.

Since GARD matching was able to have a broader recommendation based on system categories of rare diseases, it usually yielded a better prediction performance than the other 2 strategies. *Acrospiroma*, *Birt-Hogg-Dube syndrome*, and *dendritic cell tumor* all had a 100% prediction rate, though the number of affected patients was small. By using GARD matching, for example, *acrospiroma* can be inferred as *fibrosarcoma*, *Birt-Hogg-Dube syndrome* can be recommended as syndrome *adenocarcinoma of the appendix*, and *dendritic cell tumor* can be predicted as *large granular lymphocyte leukemia*. The reason for this is that all of these pairs can be categorized as rare cancers according to GARD. Similarly, in *acute promyelocytic leukemia*, *migraine with brainstem aura*, and *thyroid cancer*, *anaplastic disease* can also be recommended to other diseases within the same rare disease system. For some rare diseases, the GARD matching did not perform well. In the case of *Addison disease*, for example, although we found some recommendations by GARD matching from our datasets, such as *isolated ACTH deficiency* (categorized in Endocrine Diseases), *x-linked adrenal hypoplasia congenital* (categorized in Congenital and Genetic Diseases), *fibrous dysplasia* (categorized in Congenital and Genetic Diseases), and *syringomyelia* (categorized in Congenital and Genetic Diseases), there are still many nonrelevant results found by our system caused by general phenotypes that are related to numerous diseases. In general, the prediction of similar kind of rare diseases can still provide suggestions and clues for physicians' decision making.

Discussion

Limitations

This study demonstrates the potential to provide decision support on rare diseases for differential diagnosis. With more comprehensive knowledge extracted from clinical notes and literature, collaborative filtering performed better on both patient recommendation and rare disease prediction. The current clinical decision support (CDS) system is limited to a narrow area of clinical practice due to the inability to utilize information embedded in clinical narratives and challenges in making good semantic alignment among precision medicine knowledge and clinical data stored in various formats and heterogeneous resources. Therefore, there exists a huge opportunity to integrate our proposed work into current CDS system for a better rare disease differential diagnosis in clinical practice.

For homogeneous data, LL performance would be depressed when compared with TANI (eg, EMR only). On the other hand, LL is good at dealing with heterogeneous data, and as phenotype-rare disease associations extracted from EMR and medical literature share different perspectives, such flexibility can help us find more patterns compared with TANI. Therefore, it is not surprising that patient recommendation performance improved when we combined EMR and literature randomly, and performance improved further after we used pruned literature. OL and FMG, however, performed worse than TANI and LL. We found that OL gives too much weight to patient similarities even with few shared phenotypes. Such strict similarity measurements have difficulty finding semantic relationships and lack the ability to stratify patients well. This is possibly an explanation for the better performance of OL for single EMR data with high homogeneity but poor performance for combined datasets with high heterogeneity. Similar to OL, FMG is not good at dealing with heterogeneous data; nevertheless, it yielded a better patient recommendation performance than OL in the EMR+L and EMR+PL datasets. Furthermore, we observed that LL is sensitive to the selection of KNN or TPN, especially for combined datasets, which infers that making a good balance between KNN and TPN has the potential ability to optimize overall performance and eliminate bias with idealized neighbors and similarity at the same time.

The combination of EMR and literature did not always contribute to optimal performance in patient recommendations. The reason for this is that some biases exist when physicians or researchers documented phenotype-disease associations. For EMRs, each document is recorded based on individual physician instinct and experiences starting from a clinical perspective, and for literature, phenotypes and rare diseases with positive relationships are reported based on a large number of gene tests from a biomedical experimental perspective, which may increase the gap between these two sources. Collaborative filtering with different similarity measurements and neighborhood algorithms can remedy this problem to some extent. In the future, we plan to investigate on gene level to reduce miscommunication and balance the heterogeneity between different datasets. Besides the use of literature only, it would also be interesting to integrate cross-institutional EMRs with balanced heterogeneity to acquire diagnostic experience and knowledge from multiple hospitals and health care institutions to build a more general system for rare disease diagnostic decision support.

Conclusion and Future Work

We investigated the application of a patient-based collaborative filtering model on heterogeneous EMRs and literature with different similarity measurements and neighborhood algorithms.

XSL•FO
RenderX

Results demonstrated the potential of combining heterogeneous datasets to support diagnostic decision making for rare diseases.

In the future, we are going to fully utilize the graph structure provided by the HPO and leverage its node embeddings [5,36,37] to provide coefficient similarities from various perspectives to improve performance of disease recommendation. We also plan to resolve the heterogeneity issue and reduce miscommunication between EMR and literature by mining genotypic information to establish a comprehensive disease-phenotype-gene network for rare disease diagnosis.

Acknowledgments

This work has been supported by the National Institute of Health grants OT3TR002019, R01EB19403, R01LM011934, R01GM102282, and TR02062.

Conflicts of Interest

None declared.

References

1. Orphadata. Orphanet Rare Disease Ontology URL: http://www.orphadata.org/cgi-bin/inc/ordo_orphanet.inc.php [accessed 2018-09-24] [WebCite Cache ID 72fvJHyob]
2. Rare disease day. URL: http://rarediseaseday.us/ [WebCite Cache ID 72fvQ7paf]
3. Robinson PN, Mundlos S. The human phenotype ontology. Clin Genet 2010 Jun;77(6):525-534. [doi: 10.1111/j.1399-0004.2010.01436.x] [Medline: 20412080]
4. Shen F, Wang L, Liu H. Phenotypic analysis of clinical narratives using human phenotype ontology. Stud Health Technol Inform 2017;245:581-585. [Medline: 29295162]
5. Shen F, Liu S, Wang Y, Wang L, Afzal N, Liu H. Leveraging collaborative filtering to accelerate rare disease diagnosis. AMIA Annu Symp Proc 2017;2017:1554-1563 [FREE Full text] [Medline: 29854225]
6. Balabanovic M, Shoham Y. Fab: content-based, collaborative recommendation. Commun ACM 1997 Mar;40(3):66-72.
7. Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput 2003:76-80. [doi: 10.1109/MIC.2003.1167344]
8. Zhao Z, Shang M. User-based collaborative-filtering recommendation algorithms on hadoop. In: 2010 Third International Conference on Knowledge Discovery and Data Mining. 2010 Presented at: IEEE International Conference on Knowledge Discovery and Data Mining; January 9-10, 2010; Phuket.
9. Zhang S, Zhang C, Yang Q. Data preparation for data mining. Appl Artif Intell 2003;17(5-6):375-381. [doi: 10.1080/713827180]
10. Mobasher B, Dai H, Luo T, Nakagawa M. Improving the effectiveness of collaborative filtering on anonymous web usage data. 2001 Presented at: In Proceeding of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization; 2001; Seattle URL: https://pdfs.semanticscholar.org/4a70/b2209dfe9d7e5e084a314f1aba41402b1be9.pdf
11. Shen F, Liu H, Sohn S, Larson DW, Lee Y. BmQGen: Biomedical query generator for knowledge discovery. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2015 Presented at: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); November 9, 2015; Washington, DC p. 1092-1097 URL: https://queue.acm.org/detail.cfm?id=1103836 [doi: 10.1109/BIBM.2015.7359833]
12. Sujansky W. Heterogeneous database integration in biomedicine. J Biomed Inform 2001 Aug;34(4):285-298 [FREE Full text] [Medline: 11977810]
13. Shen F, Lee Y. Knowledge Discovery from Biomedical Ontologies in Cross Domains. PLoS One 2016 Sep;11(8):e0160005-e0160206 [FREE Full text] [doi: 10.1371/journal.pone.0160005] [Medline: 27548262]
14. Bache R, Miles S, Taweel A. An adaptable architecture for patient cohort identification from diverse data sources. J Am Med Inform Assoc 2013 Dec;20(e2):e327-e333 [FREE Full text] [doi: 10.1136/amiajnl-2013-001858] [Medline: 24064442]
15. Bleich H, Slack W. Designing a hospital information system: a comparison of interfaced and integrated systems. MD Comput 1992;9(5):293-296. [Medline: 1522791]
16. Bürkle T, Schweiger R, Altmann U, Holena M, Blobel B, Dudeck J. Transferring data from one EPR to another: content--syntax--semantic. Methods Inf Med 1999 Dec;38(4-5):321-325. [Medline: 10805022]
17. Hussain S, Ouagne D, Sadou E, Dart T, Jaulent M, De VB. EHR4CR: A semantic web based interoperability approach for reusing electronic healthcare records in protocol feasibility studies. EHR4CR; 2012 Presented at: Semantic Web Applications and Tools for Health Care and Life Sciences; November 30, 2012; Paris URL: http://ceur-ws.org/Vol-952/paper_31.pdf
18. Mate S, Köpcke F, Toddenroth D, Martin M, Prokosch H, Bürkle T, et al. Ontology-based data integration between clinical and research systems. PLoS One 2015;10(1):e0116656 [FREE Full text] [doi: 10.1371/journal.pone.0116656] [Medline: 25588043]

XSL•FO
RenderX

19. McMurry AJ, Murphy SN, MacFadden D, Weber G, Simons WW, Orechia J, et al. SHRINE: enabling nationally scalable multi-site disease studies. PLoS One 2013;8(3):e55811 [FREE Full text] [doi: 10.1371/journal.pone.0055811] [Medline: 23533569]
20. Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. Methods Inf Med 2009;48(1):45-54. [Medline: 19151883]
21. Lewis DP, Jebara T, Noble WS. Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure. Bioinformatics 2006 Nov 15;22(22):2753-2760. [doi: 10.1093/bioinformatics/btl475] [Medline: 16966363]
22. Yu S, Falck T, Daemen A, Tranchevent L, Suykens JA, De Moor B, et al. L2-norm multiple kernel learning and its application to biomedical data fusion. BMC Bioinformatics 2010 Jun 08;11:309 [FREE Full text] [doi: 10.1186/1471-2105-11-309] [Medline: 20529363]
23. Ye J, Chen K, Wu T, Li J, Zhao Z, Patel R, et al. Heterogeneous data fusion for alzheimer's disease study. 2008 Presented at: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining; 2008; Las Vegas p. 1025-1033. [doi: 10.1145/1401890.1402012]
24. 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 Sep 11. [doi: 10.1016/j.jbi.2018.09.008] [Medline: 30217670]
25. Torii M, Wagholikar K, Liu H. Using machine learning for concept extraction on clinical documents from multiple data sources. J Am Med Inform Assoc 2011;18(5):580-587 [FREE Full text] [doi: 10.1136/amiajnl-2011-000155] [Medline: 21709161]
26. Sokolov A, Funk C, Graim K, Verspoor K, Ben-Hur A. Combining heterogeneous data sources for accurate functional annotation of proteins. BMC Bioinformatics 2013;14 Suppl 3:S10 [FREE Full text] [doi: 10.1186/1471-2105-14-S3-S10] [Medline: 23514123]
27. Aronson AR. Metamap: Mapping text to the UMLS Metathesaurus. URL: https://ii.nlm.nih.gov/Publications/Papers/metamap06.pdf [WebCite Cache ID 72j44oz9v]
28. Genetic and Rare Diseases Information Center. URL: https://rarediseases.info.nih.gov/ [WebCite Cache ID 72fzD2m7G]
29. Kilicoglu H, Shin D, Fiszman M, Rosemblat G, Rindflesch T. SemMedDB: a PubMed-scale repository of biomedical semantic predications. Bioinformatics 2012 Dec 01;28(23):3158-3160 [FREE Full text] [doi: 10.1093/bioinformatics/bts591] [Medline: 23044550]
30. Breese JS, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. 1998 Presented at: Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence; 1998; Madison p. 43-52.
31. Kilicoglu H, Fiszman M, Rodriguez A, Shin D, Ripple A, Rindflesch T. Semantic MEDLINE: A Web Application for Managing the Results of PubMed Searches. 2008 Presented at: Proceedings of the third international symposium for semantic mining in biomedicine; September 1-3, 2008; Turku p. 69-76.
32. Kilicoglu H, Rosemblat G, Fiszman M, Rindflesch TC. Constructing a semantic predication gold standard from the biomedical literature. BMC Bioinformatics 2011 Dec 20;12:486 [FREE Full text] [doi: 10.1186/1471-2105-12-486] [Medline: 22185221]
33. Rindflesch T, Kilicoglu H, Fiszman M, Rosemblat G, Shin D. Semantic MEDLINE: an advanced information management application for biomedicine. Inf Serv Use 2011;31(1-2):15-21 [FREE Full text]
34. Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform 2003 Dec;36(6):462-477 [FREE Full text] [doi: 10.1016/j.jbi.2003.11.003] [Medline: 14759819]
35. Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering. 2007 Presented at: Proceedings of the 24th international conference on Machine learning; 2007; New York p. 791-798.
36. Grover A, Leskovec J. node2vec: scalable feature learning for networks. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining; 2016; San Francisco p. 855-864.
37. Shen F, Liu S, Wang Y, Wang L, Wen A, Limper A. Constructing Node Embeddings for Human Phenotype Ontology to Assist Phenotypic Similarity Measurement. 2018 Presented at: IEEE International Conference on Healthcare Informatics; 2018; New York City.

Abbreviations

CDS: clinical decision support
EMR: electronic medical record
FMG: Fager and McGowan coefficient similarity
GARD: Genetic and Rare Diseases Information Center
HPO: Human Phenotype Ontology
KNN: K-nearest neighbors
LL: log likelihood ratio similarity
OL: overlap coefficient similarity
PRAUC: area under the precision-recall curve

XSL•FO
RenderX

RMSE: root mean square error
SNOMED: systematized nomenclature of medicine
TANI: Tanimoto coefficient similarity
TPN: threshold patient neighbor

Edited by G Eysenbach; submitted 15.06.18; peer-reviewed by H Guo, A Kulanthaivel, B Polepalli Ramesh; comments to author 17.07.18; revised version received 19.08.18; accepted 08.09.18; published 10.10.18.

Please cite as:
Shen F, Liu S, Wang Y, Wen A, Wang L, Liu H
Utilization of Electronic Medical Records and Biomedical Literature to Support the Diagnosis of Rare Diseases Using Data Fusion and Collaborative Filtering Approaches
JMIR Med Inform 2018;6(4):e11301
URL: http://medinform.jmir.org/2018/4/e11301/
doi:10.2196/11301
PMID:30305261

©Feichen Shen, Sijia Liu, Yanshan Wang, Andrew Wen, Liwei Wang, Hongfang Liu. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 10.10.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predicting Current Glycated Hemoglobin Values in Adults: Development of an Algorithm From the Electronic Health Record

Brian J Wells[1], MD, PhD; Kristin M Lenoir[1], MPH; Jose-Franck Diaz-Garelli[2], PhD; Wendell Futrell[3], BS; Elizabeth Lockerman[4], MD; Kevin M Pantalone[5], DO; Michael W Kattan[6], PhD

[1]Division of Public Health Sciences, Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[2]Department of Physiology and Pharmacology, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[3]Clinical and Translational Science Institute, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[4]Department of Internal Medicine, Loyola University Medical Center, Maywood, IL, United States

[5]Endocrinology and Metabolism Institute, Department of Endocrinology, Diabetes and Metabolism, Cleveland Clinic, Cleveland, OH, United States

[6]Lerner Research Institute, Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, United States

Corresponding Author:
Brian J Wells, MD, PhD
Division of Public Health Sciences
Department of Biostatistics and Data Science
Wake Forest University School of Medicine
525 Vine Street, 4th Floor
Medical Center Bouldevard
Winston-Salem, NC,
United States
Phone: 1 336 416 5185
Email: bjwells@wakehealth.edu

Abstract

Background: Electronic, personalized clinical decision support tools to optimize glycated hemoglobin (HbA_{1c}) screening are lacking. Current screening guidelines are based on simple, categorical rules developed for populations of patients. Although personalized diabetes risk calculators have been created, none are designed to predict current glycemic status using structured data commonly available in electronic health records (EHRs).

Objective: The goal of this project was to create a mathematical equation for predicting the probability of current elevations in HbA_{1c} (≥5.7%) among patients with no history of hyperglycemia using readily available variables that will allow integration with EHR systems.

Methods: The reduced model was compared head-to-head with calculators created by Baan and Griffin. Ten-fold cross-validation was used to calculate the bias-adjusted prediction accuracy of the new model. Statistical analyses were performed in R version 3.2.5 (The R Foundation for Statistical Computing) using the rms (Regression Modeling Strategies) package.

Results: The final model to predict an elevated HbA_{1c} based on 22,635 patient records contained the following variables in order from most to least importance according to their impact on the discriminating accuracy of the model: age, body mass index, random glucose, race, serum non–high-density lipoprotein, serum total cholesterol, estimated glomerular filtration rate, and smoking status. The new model achieved a concordance statistic of 0.77 which was statistically significantly better than prior models. The model appeared to be well calibrated according to a plot of the predicted probabilities versus the prevalence of the outcome at different probabilities.

Conclusions: The calculator created for predicting the probability of having an elevated HbA_{1c} significantly outperformed the existing calculators. The personalized prediction model presented in this paper could improve the efficiency of HbA_{1c} screening initiatives.

(JMIR Med Inform 2018;6(4):e10780) doi:10.2196/10780

XSL•FO
RenderX

KEYWORDS

electronic health records; risk prediction; clinical decision support; hemoglobin A1c; diabetes

Introduction

Many prediction tools have been created to assess the risk of undiagnosed diabetes and related outcomes such as impaired glucose tolerance, prediabetes, risk of future diabetes, and hyperinsulinemia. Most of these tools are not practical in the setting of the electronic health record (EHR) because they include predictor variables not readily available in structured formats [1-20]. Examples of impractical variables include waist circumference, fasting time, physical activity, review of systems, diet, pregnancy-related variables, and detailed ethnicity. Tools typically leverage fasting glucose level as a predictor, which is simple to obtain in practice, but documentation of fasting time in a structured fashion in EHRs is generally absent. The authors identified two tools that accurately predict the presence of diabetes using structured variables routinely present in the EHR [21,22].

Current guidelines from the United States Preventive Services Task Force (USPSTF) recommend screening for abnormal blood glucose in adults aged 40 to 70 years who are overweight or obese. The USPSTF acknowledges that patients with other high-risk characteristics (eg, family history of diabetes, personal history of gestational diabetes) may need to be screened sooner but this is left up to the physician's discretion [23]. The guidelines published by the American Diabetes Association (ADA) recommend glucose screening of adult patients with an elevated body mass index (BMI; ≥25 kg/m^2) plus another risk factor (eg, hypertension, physical inactivity, family history of diabetes) at any age and for all patients beginning at 45 years of age at 3-year intervals [24].

Current approaches do not take advantage of advanced statistical modeling. Hyperglycemia risk prediction with simultaneous consideration of numerous independent variables and nonlinear effects is statistically ideal in the context of a multifactorial pathology. Creating strict cutoffs of individual variables or combinations of a limited number of variables for clinical guidelines does not take advantage of what can now be reasonably achieved. The USPSTF and ADA guidelines encourage physician judgment in the application of glucose screening but do not provide specific guidance. Simplified classification methods used in cancer have been notoriously poor at discriminating between high- and low-risk patients [25]. Moreover, many existing models for predicting hyperglycemia risk are also likely reducing their prediction accuracy by categorizing continuous variables, which reduces granularity and may miss potentially complex associations between a continuous variable and the outcome. This issue was highlighted by Kattan [26] when he showed that traditional regression techniques that incorporated restricted cubic splines to reduce linearity assumptions were found to produce more accurate risk prediction models when compared with classification methods such as classification and regression trees and artificial neural networks.

Predicting the date of onset of hyperglycemia is difficult due to the lack of symptoms early in the course of the disease and inconsistent testing and/or documentation in clinical practice (particularly in a structured fashion). Previous studies indicate that the onset of type 2 diabetes frequently occurs more than 5 years before diagnosis [27,28]. In contrast, blood measurements of glycated hemoglobin (HbA_{1c}) provide an easy and accurate method for determining current mean glycemia over the previous 8 to 12 weeks without the need for fasting. HbA_{1c} testing is standardized according to specifications defined by the National Glycohemoglobin Standardization Program (NGSP). HbA_{1c} levels are the primary blood marker used for guiding the management of type 2 diabetes, and the ADA has approved HbA_{1c} testing for diabetes screening [24]. The increasing use of HbA_{1c} as a screening tool in patients without prediabetes or diabetes provides data for prediction modeling from EHR records.

The authors strongly believe that the identification of patients with elevated HbA_{1c} is important clinically despite previous studies that have not shown a mortality benefit from screening for diabetes [29]. The early detection of elevated blood sugar can have other significant benefits:

- Behavioral counseling can lead to reductions in cardiovascular disease risk [30].
- Treatment of prediabetes, which affects approximately 35% of the adult population in the United States [31], has been shown to delay progression to diabetes [32].
- Diabetic-specific retinopathy is present in up to 21% of patients with newly diagnosed type 2 diabetes [33], while peripheral neuropathy and nephropathy are present in 21.5% and 26.5%, respectively, of patients with undiagnosed diabetes [34]. Aggressive blood sugar and blood pressure control among patients with diabetes reduces the risk of microvascular complications [35,36].
- Early detection of diabetes allows for the allocation of proven preventive strategies (eg, fundoscopic screening for retinopathy, pneumococcal vaccination, screening for nephropathy, and aggressive prevention of cardiovascular disease) [37].
- Appropriate documentation of elevated blood sugar and diabetes allows health systems and payers to improve the risk stratification of patients and increases the potential pool of patients available for participation in clinical research.

Therefore, an accurate tool for predicting the current probability that a specific patient has an elevation in HbA_{1c} levels would constitute a major advancement in finding patients with the most probable need of screening interventions. To address this gap, we created a calculator for predicting the probability that a given patient with no history of diabetes or elevated blood sugar currently has an elevated HbA_{1c} value (≥5.7%). This cutoff was chosen because it corresponds with the current guidelines published by the ADA that indicate values <5.7%

XSL•FO
RenderX

are considered to be normal. Importantly, the calculator presented in this paper was restricted only to structured variables typically available in EHRs. This focus on common structured variables will enable the tool to be integrated into EHRs for implementation.

Methods

This study was conducted on all adult patients who have undergone HbA_{1c} testing prior to evidence of hyperglycemia (random blood sugar ≥200 mg/dL), any diabetes-related diagnostic code, or prescription for an antihyperglycemic medication. Data were extracted from the Epicare EHR at Wake Forest Baptist Medical Center in Winston-Salem, North Carolina, for the dates between September 2012 and September 2016. The study was approved by the institutional review board and granted a waiver of informed consent. Data were limited to structured data located in these areas of the EHR: encounter diagnoses, problem list, past medical history, procedures, prescriptions, vital signs, demographics, social history, and laboratory values. Candidate predictor variables were chosen based on their theoretical association with hyperglycemia. Textbox 1 shows a list of the candidate predictor variables included in the complete statistical model.

Independent variables were defined on the date of the HbA_{1c} of interest. For missing continuous variables (eg, systolic blood pressure), the most recent prior value was used instead. Patients completely lacking values for independent variables were excluded. The investigators did not impute missing data because it was felt that imputation would not be appropriate at the point of implementation. Comorbidities were considered to be present if the patient had any structured instances of the diagnostic code on or before the date of the first HbA_{1c}. Medications with start dates on or before the date of the HbA_{1c} and end dates on or after the date of HbA_{1c} were considered to be active. Medication order dates were used when the start dates were missing. Medications missing both start and order dates were excluded. Medication categories (eg, antihyperglycemics) were provided by First Databank Inc. Multiple logistic regression was used to model the association between the independent variables and the outcome of HbA_{1c} ≥5.7%. Continuous variables were fit using restricted cubic splines with 3-knots. Due to collinearity, the model could not be fit with the simultaneous inclusion of serum non–high-density lipoprotein and high-density lipoprotein. Therefore, high-density lipoprotein was removed from the complete model. The model was reduced using Harrell's model approximation method [39]. For parsimony, the diagnosis of obesity variable was removed after variable selection. The diagnosis had little impact on the prediction accuracy and was redundant since BMI is also in the model. The reduced model was compared head-to-head with the calculators created by Baan and Griffin [21,22]. The head-to-head comparisons were performed using 10-fold cross-validation in order to calculate the bias-adjusted prediction accuracy of the new model. Prediction model metrics included measures of discrimination (concordance statistic), calibration (calibration curves), and decision curves [40]. Statistical analyses were performed in R version 3.2.5 (R Foundation for Statistical Computing) using the rms (Regression Modeling Strategies) package.

XSL•FO
RenderX

Textbox 1. Candidate variables in the complete model prior to variable selection.

Laboratory measurements:

- Serum triglycerides
- Random blood glucose
- Serum non–high-density lipoprotein
- Serum high-density lipoprotein (dropped due to an inability to fit model)
- Serum total cholesterol
- Estimated glomerular filtration rate (estimated from serum creatinine using the modified Chronic Kidney Disease Epidemiology Collaboration formula [38])

Active prescription medication categories:

- Antihypertensive
- First generation antipsychotic
- Second generation antipsychotic
- 3-hydroxy-3-methyl-glutaryl-*coenzyme A* reductase inhibitor (statin)
- Fibrate
- Valproic acid
- Beta-blocker
- Thiazide diuretic
- Niacin
- Oral glucocorticoid
- Protease inhibitor
- Nucleoside reverse transcriptase inhibitor
- Oral contraceptive
- Injectable medroxyprogesterone acetate
- Cyclosporine
- Sirolimus
- Tacrolimus

Diagnosis codes (see Multimedia Appendix 1):

- Hypertension
- Ischemic heart disease
- Peripheral vascular disease
- Neuropathy
- Obesity
- Hyperlipidemia

Vital signs:

- Systolic blood pressure
- Diastolic blood pressure
- Body mass index

Demographics:

- Race
- Age
- Gender

XSL•FO
RenderX

Family history:

- Number of first degree relatives with diabetes

Social history:

- Smoking status

Results

The record search identified 22,635 patients for model building and validation of which 26% were found to have an elevated HbA_{1c} (≥5.7%). Figure 1 shows the number of patients included and excluded from the model building.

The final model included the following 8 variables ordered from most important to least important: age, BMI, random glucose, race, serum non–high-density lipoprotein, serum total cholesterol, estimated glomerular filtration rate (eGFR), and smoking status. Table 1 shows descriptive statistics for the variables included in the final model by HbA_{1c} results. As expected, patients found to have elevated HbA_{1c} levels were older, had higher BMI, lower eGFR, and higher random glucose values.

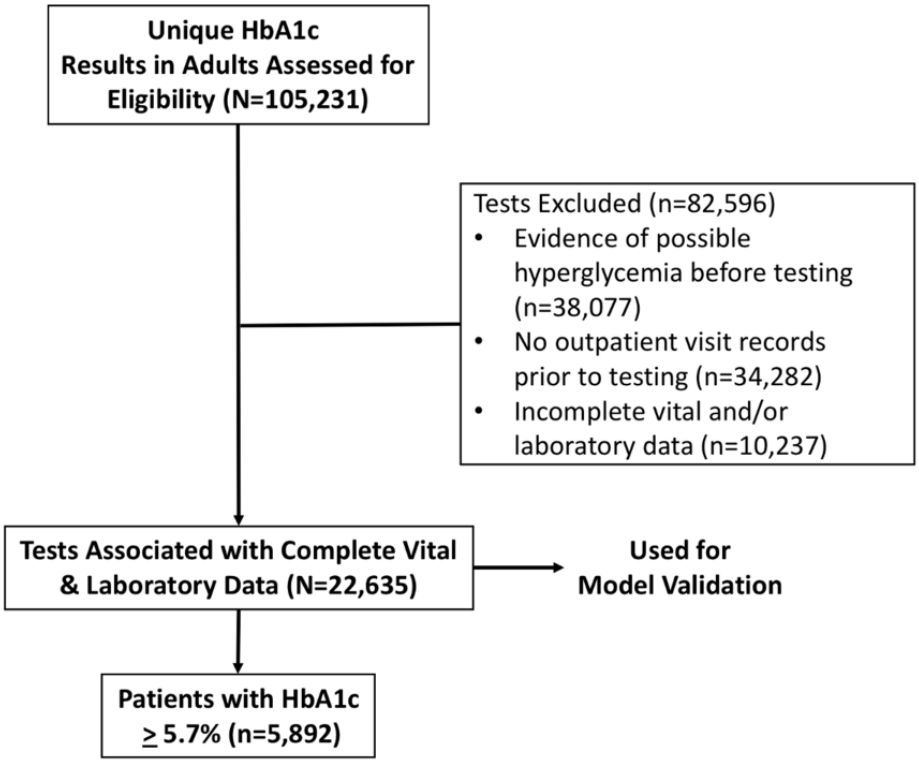
The coefficients along with instructions for calculating the probability of an elevated HbA_{1c} (≥5.7%) and sample calculations for 2 patient scenarios are shown in Multimedia Appendix 2.

The 3 models were compared in their ability to accurately rank patients according to risk as measured by the concordance statistic (c-stat) and bias-adjusted using 10-fold cross-validation. The current model (c-stat 0.765, 95% CI 0.762 to 0.769) demonstrated statistically significant improvements in discrimination when compared to the models created by Baan (c-stat 0.637, 95% CI 0.633 to 0.641) and Griffin (c-stat 0.668, 95% CI 0.665 to 0.672).

The calibration curve shown in Figure 2 reveals that the current model is well calibrated. The predicted probabilities tend to overestimate risk at the right tail of the distribution, but the wide confidence intervals allude to the scarcity of the data at these extreme high levels of risk. Error bars represent the 95% confidence interval around the point estimate.

Decision curves are displayed in Figure 3 and also demonstrate the superiority of this model. Our model shows a net benefit up to a probability of 0.73 for an elevated HbA_{1c} (≥5.7%) without significant net harms above this threshold. This model confers a net benefit that is equal to or greater than the net benefit offered by the other models at all probability thresholds.

Figure 1. Data flowsheet. HbA_{1c}: glycated hemoglobin.



XSL•FO
RenderX

Table 1. Descriptive statistics by glycated hemoglobin outcome.

Characteristics	HbA_{1c}[a] <5.7% (n=16,743)	HbA_{1c} ≥5.7% (n=5892)	*P* value
Age (years) mean (SD)	48.1 (15.4)	54.8 (14.0)	<.001
Race, n (%)			<.001
Black	3692 (22.05)	2183 (37.05)	
Other	1178 (7.00)	487 (8.30)	
White	11873 (70.91)	3222 (54.68)	
BMI[b] (kg/m^2), mean (SD)	30.1 (7.44)	33.0 (8.41)	<.001
Smoking status, n (%)			<.001
Current smoker	2747 (16.41)	1393 (23.64)	
Former smoker	3867 (23.10)	1480 (25.11)	
Never smoker	10129 (60.50)	3019 (51.23)	
eGFR[c] ($mL/min/1.73\ m^2$), mean (SD)	92.0 (33.0)	87.9 (30.8)	<.001
Random blood glucose (mg/dL), mean (SD)	88.4 (12.7)	96.1 (16.0)	<.001
Non-HDL[d] cholesterol (mg/dL), mean (SD)	135 (37.4)	144 (41.7)	<.001
Total cholesterol (mg/dL), mean (SD)	186 (39.4)	192 (43.1)	<.001

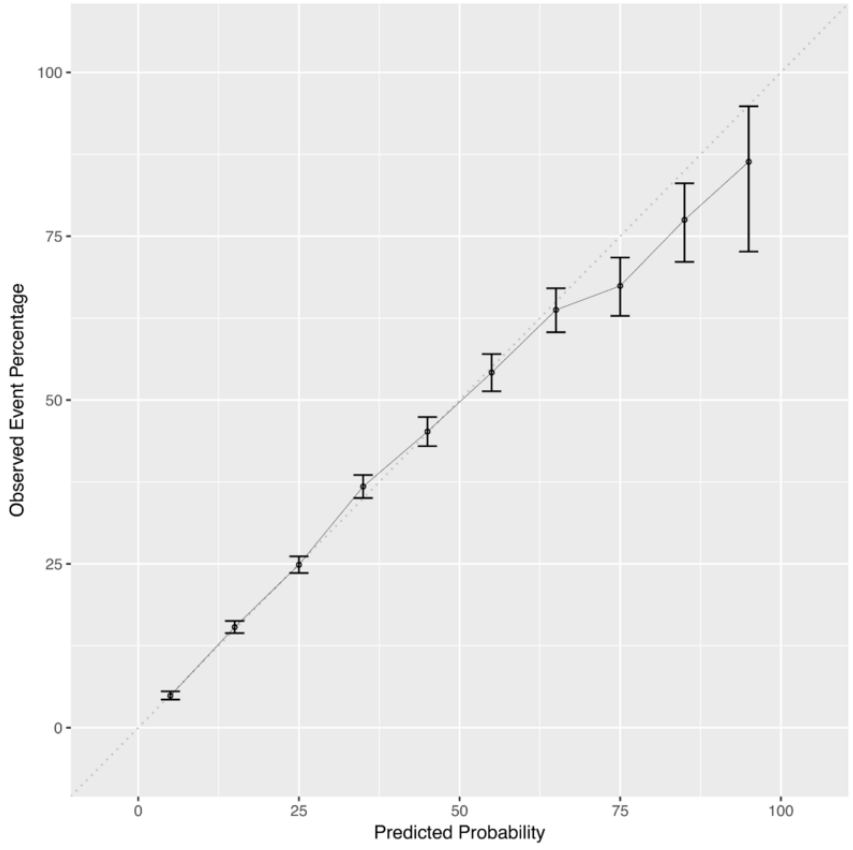
[a]HbA_{1c}: glycated hemoglobin.

[b]BMI: body mass index.

[c]eGFR: estimated glomerular filtration rate, calculated using the Chronic Kidney Disease Epidemiology Collaboration formula (CKD-EPI) [38].

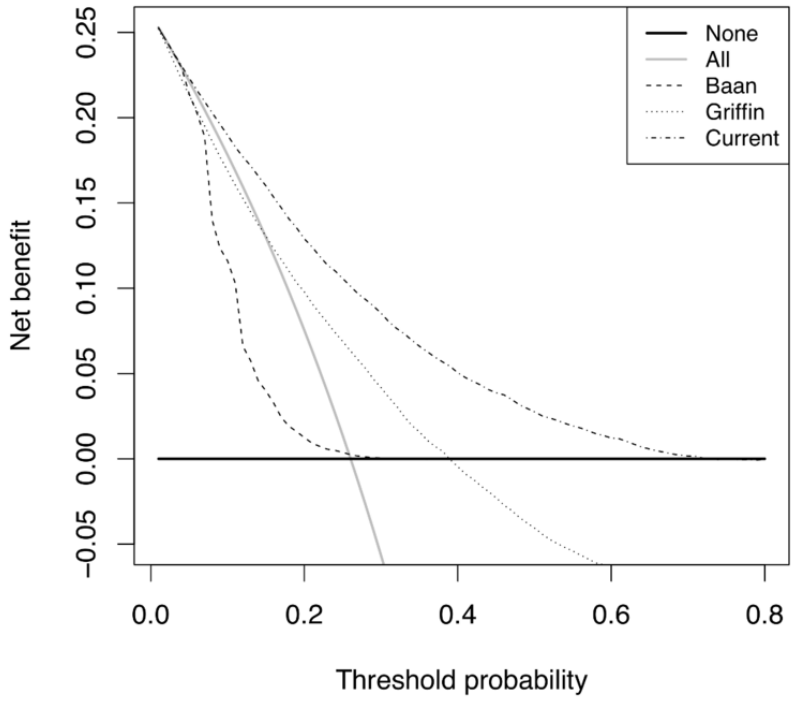
[d]HDL: high-density lipoprotein.

Figure 2. Calibration curve of the new model for predicting glycated hemoglobin ≥5.7%.



XSL•FO
RenderX

Figure 3. Decision curve analysis.



Discussion

Principal Findings

The calculator created for predicting the probability of having an elevated HbA_{1c} significantly outperformed the existing calculators. It should be noted that the calculators created by Baan and Griffin were designed for predicting current glucose tolerance test results and were not specifically calibrated to predict HbA_{1c} values. However, any potential issues with calibration should not impact the ability to discriminate patients according to risk. The authors chose not to develop a simple risk score, which would be easier to calculate without a computer but would be less accurate and would not provide an absolute probability. One of the benefits of using multivariable regression over many machine learning methods is that the mathematical output of this model can be integrated into an EHR using common mathematical operations. In contrast, classification-based methods like random forest, artificial neural networks, and classification and regression trees would increase the complexity of implementation by requiring separate software outside the EHR to calculate probabilities. The movement of data into and out of the EHR also raises concerns about security and privacy.

Limitations and Strengths

Limitations of the study include the lack of external validation. The model was validated internally using resampling and may not reflect the prediction accuracy that would be achieved in a prospective fashion at the current institution or when validated in a different health system. However, the authors used 10-fold cross-validation in which patients in the test data for each fold were not used to build the model. Another limitation pertains to the lack of data from outside health systems. Patients may have additional medication or laboratory results outside of the health system that could alter the predicted risk or change the patient's status in terms of hyperglycemia. In order to ensure that patients have a minimal amount of data to guide the calculator's creation, the investigators required that patients had at least one value for each of the independent variables. Future research and quality improvement projects may need to query patients about health history prior to implementation.

A relatively small proportion of historical HbA_{1c} tests were appropriate for use in model building. Some of the tests were obtained before the installation of a comprehensive EHR and, therefore, accompanying information like vital signs were not available. Many of the tests were obtained in patients who already had evidence of possible hyperglycemia, some of whom were already being treated with antidiabetic medication. These patients would be inappropriate to use for the creation of a model aimed at patients with unknown glucose status. Limiting the model building and validation dataset to patients with complete data further reduced the sample size from 32,872 to 22,635. The authors chose not to impute the missing values given the adequate number of patients with complete data. In addition, the authors are not convinced that imputation would be acceptable to patients and providers when the model is implemented into practice. Patients lacking common variables used in the model such as BMI and blood pressure values are probably very new to the health system or are seeking their usual care elsewhere. Serum creatinine and lipid measurements are routinely obtained in clinical practice, especially among older adults. Patients without any creatinine or lipid measurements are likely to be younger and less likely to be at risk for diabetes. The authors felt it was important to identify a population for model building that matches the future population where the model will be implemented. Despite the restrictions

XSL•FO
RenderX

on data inclusion, the dataset contained >5000 patients with the outcome of interest. The size of this dataset is large compared to most similar studies conducted prior to the adoption of EHRs and is more than adequate for regression modeling. Harrel [39] has proposed that 7 to 10 outcomes for each degree of freedom are necessary to prevent overfitting when building a regression equation. The model created in this study contains 28 degrees of freedom, which could have safely been built from a dataset containing only 196 to 280 outcomes according to the aforementioned heuristic. A feasibility analysis was conducted among patients in the Department of Family and Community Medicine, and it was determined that approximately 20% of the adult patients seen in the past 3 years would be appropriate for application of the tool.

Imprecision in the measurement of HbA_{1c} levels could have negatively impacted the model building and could decrease the prediction accuracy of the model upon implementation. Wake Forest Baptist Medical Center is not a certified member of the NGSP but maintains accreditation by the Clinical Laboratory Improvement Amendments Program (identification number 34D0664386). The Wake Forest Baptist Medical Center's core laboratory performs HbA_{1c} testing using ion exchange high performance liquid chromatography, which is highly precise and constituted the vast majority of HbA_{1c} measurements used to create the data for this study. However, the investigators did not exclude HbA_{1c} measurements obtained using different methods at other locations in the health system (eg, point-of-care testing), which likely introduced variability in the HbA_{1c} measurements. Comorbid conditions such as iron deficiency anemia can lead to HbA_{1c} measurements that do not accurately reflect average blood glucose levels [41]. Despite the potential negative impact of imprecise or inaccurate HbA_{1c} measurements, the prediction model performed very well.

Conclusions

Improving the efficiency of diabetes screening should be of great interest in the United States given the increased use of value-based care contracts. Health systems could use our model for diabetes screening initiatives in a variety of ways. The decision curves suggest that using the new algorithm to guide HbA_{1c} testing would provide a net benefit between probabilities of 0.01 to 0.71. The authors will conduct a targeted screening study in which patients with a predicted risk of an elevated HbA_{1c} ≥50% will be notified directly regarding their elevated risk. Coupled with standing laboratory orders, this direct-to-patient design would enable patients to undergo HbA_{1c} testing prior to a physician visit. This is particularly important for patients with infrequent in-person visits. The hope is that patients with subsequent elevations in HbA_{1c} would be more likely to re-engage with the health system.

In summary, the risk equation created in this study is optimized for integration within an EHR and outperforms other similar models. Future research will attempt to integrate the risk calculator into clinical workflows, examine the ability of the calculator to predict risk in other health systems, and evaluate the potential economic savings of using this model for diabetes screening.

Acknowledgments

We would like to acknowledge the data extraction and statistical assistance of the Wake Forest Clinical and Translational Science Institute, which is supported by the National Center for Advancing Translational Sciences, National Institutes of Health (grant number UL1TR001420).

JFDG is supported by the Postdoctoral Research, Instruction, and Mentoring Experience program. The program is funded by the National Institute of General Medical Sciences as part of the Institutional Research and Career Development Award (grant number 5K12GM10277305).

Authors' Contributions

BJW and KML were responsible for design of the study. BJW, KML, EL, and MWK were responsible for statistical analyses. BJW, KML, JFDG, EL, and KMP collaborated on the writing of the manuscript. JFDG performed the literature review. WF performed data extraction and editing. KMP contributed his clinical expertise.

Conflicts of Interest

MWK conducts research sponsored by Novo Nordisk and Merck that is not directly related to this project. In the past 12 months, KMP has received research support from Merck and Novo Nordisk that is not directly related to this project. In addition, in the past 12 months, KMP has received speaker honoraria from Merck, Novo Nordisk, and Astra Zeneca and consulting honoraria from Merck, Novo Nordisk, Sanofi, and Eli Lilly. The other authors report no potential conflicts of interest.

Multimedia Appendix 1

Diagnostic codes.

[PDF File (Adobe PDF File), 35KB - medinform_v6i4e10780_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2

Instructions for calculating the probability of an elevated glycated hemoglobin and sample calculations.

[PDF File (Adobe PDF File), 70KB - medinform_v6i4e10780_app2.pdf]

References

1. Al-Lawati J, Tuomilehto J. Diabetes risk score in Oman: a tool to identify prevalent type 2 diabetes among Arabs of the Middle East. Diabetes Res Clin Pract 2007 Sep;77(3):438-444. [doi: 10.1016/j.diabres.2007.01.013] [Medline: 17306410]
2. Colagiuri S, Hussain Z, Zimmet P, Cameron A, Shaw J. Screening for type 2 diabetes and impaired glucose metabolism: the Australian experience. Diabetes Care 2004 Feb;27(2):367-371. [Medline: 14747215]
3. Glümer C, Carstensen B, Sandbaek A, Lauritzen T, Jørgensen T, Borch-Johnsen K, inter99 study. A Danish diabetes risk score for targeted screening: the Inter99 study. Diabetes Care 2004 Mar;27(3):727-733. [Medline: 14988293]
4. Herman W, Smith P, Thompson T, Engelgau M, Aubert R. A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes. Diabetes Care 1995 Mar;18(3):382-387. [Medline: 7555482]
5. Hidvegi T, Hetyesi K, Biro L, Jermendy G. Screening for metabolic syndrome in hypertensive and/or obese subjects registered in primary health care in Hungary. Med Sci Monit 2003;9(7):CR328-CR334.
6. Hippisley-Cox J, Coupland C. Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. BMJ 2017:359.
7. Kanaya A, Wassel F, de Rekeneire N, Shorr R, Schwartz A, Goodpaster B, et al. Predicting the development of diabetes in older adults: the derivation and validation of a prediction rule. Diabetes Care 2005 Feb;28(2):404-408. [Medline: 15677800]
8. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care 2003 Mar;26(3):725-731. [Medline: 12610029]
9. Meigs J, Williams K, Sullivan L, Hunt K, Haffner S, Stern M, et al. Using metabolic syndrome traits for efficient detection of impaired glucose tolerance. Diabetes Care 2004;27(6):1417-1426.
10. Mohan V, Deepa R, Deepa M, Somannavar S, Datta M. A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects. J Assoc Physicians India 2005 Sep;53:759-763. [Medline: 16334618]
11. Pearson T, Pronk N, Tan A, Halstenson C. Identifying individuals at risk for the development of type 2 diabetes mellitus. Am J Manag Care 2003 Jan;9(1):57-66 [FREE Full text] [Medline: 12549815]
12. Rauh SP, Heymans MW, Koopman ADM, Nijpels G, Stehouwer CD, Thorand B, et al. Predicting glycated hemoglobin levels in the non-diabetic general population: development and validation of the DIRECT-DETECT prediction model—a DIRECT study. PLoS One 2017 Feb;12(2):e0171816 [FREE Full text] [doi: 10.1371/journal.pone.0171816] [Medline: 28187151]
13. Rolka D, Narayan K, Thompson T, Goldman D, Lindenmayer J, Alich K, et al. Performance of recommended screening tests for undiagnosed diabetes and dysglycemia. Diabetes Care 2001 Nov;24(11):1899-1903. [Medline: 11679454]
14. Ruige J, de Neeling J, Kostense P, Bouter L, Heine R. Performance of an NIDDM screening questionnaire based on symptoms and risk factors. Diabetes Care 1997 Apr;20(4):491-496. [Medline: 9096967]
15. Saydah S, Byrd-Holt D, Harris M. Projected impact of implementing the results of the diabetes prevention program in the US population. Diabetes Care 2002;25(11):1940-1945.
16. Schmidt M, Duncan B, Vigo A, Pankow J, Ballantyne C, Couper D, ARIC Investigators. Detection of undiagnosed diabetes and other hyperglycemia states: the Atherosclerosis Risk in Communities Study. Diabetes Care 2003 May;26(5):1338-1343. [Medline: 12716785]
17. Schulze M, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Möhlig M, et al. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. Diabetes Care 2007 Mar;30(3):510-515. [doi: 10.2337/dc06-2089] [Medline: 17327313]
18. Stern M, Williams K, Haffner S. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? Ann Intern Med 2002 Apr 16;136(8):575-581. [Medline: 11955025]
19. Tabaei B, Engelgau M, Herman W. A multivariate logistic regression equation to screen for dysglycaemia: development and validation. Diabet Med 2005 May;22(5):599-605. [doi: 10.1111/j.1464-5491.2005.01467.x] [Medline: 15842515]
20. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007 May 28;167(10):1068-1074. [doi: 10.1001/archinte.167.10.1068] [Medline: 17533210]
21. Griffin S, Little P, Hales C, Kinmonth A, Wareham N. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. Diabetes Metab Res Rev 2000;16(3):164-171. [Medline: 10867715]
22. Baan CA, Ruige JB, Stolk RP, Witteman JC, Dekker JM, Heine RJ, et al. Performance of a predictive model to identify undiagnosed diabetes in a health care setting. Diabetes Care 1999 Feb;22(2):213-219 [FREE Full text] [Medline: 10333936]
23. Siu AL, US Preventive Services Task Force. Screening for abnormal blood glucose and type 2 diabetes mellitus: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2015 Dec 01;163(11):861-868. [doi: 10.7326/M15-2345] [Medline: 26501513]

XSL•FO
RenderX

24. Professional Practice Committee. Standards of medical care in diabetes—2018. Diabetes Care 2018;41(Supplement 1):S3. [doi: 10.2337/dc18-Sppc01]
25. Kattan M. Nomograms are superior to staging and risk grouping systems for identifying high-risk patients: preoperative application in prostate cancer. Curr Opin Urol 2003 Mar;13(2):111-116. [doi: 10.1097/01.mou.0000058631.64616.54] [Medline: 12584470]
26. Kattan MW. Comparison of Cox regression with other methods for determining prediction models and nomograms. J Urol 2003 Dec;170(6 Pt 2):S6-S9. [doi: 10.1097/01.ju.0000094764.56269.2d] [Medline: 14610404]
27. Harris M, Klein R, Welborn T, Knuiman M. Onset of NIDDM occurs at least 4-7 yr before clinical diagnosis. Diabetes Care 1992 Jul;15(7):815-819. [Medline: 1516497]
28. Porta M, Curletto G, Cipullo D, Rigault de la Longrais R, Trento M, Passera P, et al. Estimating the delay between onset and diagnosis of type 2 diabetes from the time course of retinopathy prevalence. Diabetes Care 2014 Jun;37(6):1668-1674. [doi: 10.2337/dc13-2101] [Medline: 24705614]
29. Selph S, Dana T, Blazina I, Bougatsos C, Patel H, Chou R. Screening for type 2 diabetes mellitus: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2015 Jun 02;162(11):765-776. [doi: 10.7326/M14-2221] [Medline: 25867111]
30. Lin J, O'Connor E, Whitlock E, Beil T. Behavioral counseling to promote physical activity and a healthful diet to prevent cardiovascular disease in adults: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2010 Dec 07;153(11):736-750. [doi: 10.7326/0003-4819-153-11-201012070-00007] [Medline: 21135297]
31. Karve A, Hayward RA. Prevalence, diagnosis, and treatment of impaired fasting glucose and impaired glucose tolerance in nondiabetic U.S. adults. Diabetes Care 2010 Nov;33(11):2355-2359 [FREE Full text] [doi: 10.2337/dc09-1957] [Medline: 20724649]
32. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 7;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
33. Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, et al. Retinopathy in diabetes. Diabetes Care 2004 Jan 01:s84. [doi: 10.2337/diacare.27.2007.S84]
34. Koopman RJ, Mainous AG, Liszka HA, Colwell JA, Slate EH, Carnemolla MA, et al. Evidence of nephropathy and peripheral neuropathy in US adults with undiagnosed diabetes. Ann Fam Med 2006 Sep;4(5):427-432 [FREE Full text] [doi: 10.1370/afm.577] [Medline: 17003143]
35. UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes. BMJ 1998 Sep 12;317(7160):703-713 [FREE Full text] [Medline: 9732337]
36. UK Prospective Diabetes Study Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998 Sep 12;352(9131):837-853. [Medline: 9742976]
37. American Diabetes Association. Standards of medical care in diabetes—2014. Diabetes Care 2014 Jan;37 Suppl 1:S14-S80. [doi: 10.2337/dc14-S014] [Medline: 24357209]
38. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med 2009 May 05;150(9):604-612 [FREE Full text] [Medline: 19414839]
39. Harrell F, Lee K, Mark D. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-387.
40. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006 Sep;26(6):565-574 [FREE Full text] [doi: 10.1177/0272989X06295361] [Medline: 17099194]
41. Sacks DB, John WG. Interpretation of hemoglobin A1c values. JAMA 2014 Jun 11;311(22):2271-2272. [doi: 10.1001/jama.2014.6342] [Medline: 24915255]

Abbreviations

ADA: American Diabetes Association
CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration formula
c-stat: concordance statistic
eGFR: estimated glomerular filtration rate
EHR: electronic health record
HbA $_{1c}$: glycated hemoglobin
NGSP: National Glycohemoglobin Standardization Program
USPSTF: United States Preventive Services Task Force

XSL•FO
RenderX

Edited by G Eysenbach; submitted 16.04.18; peer-reviewed by Y Chu, D Monneret; comments to author 08.07.18; revised version received 18.08.18; accepted 21.09.18; published 22.10.18.

Please cite as:
Wells BJ, Lenoir KM, Diaz-Garelli JF, Futrell W, Lockerman E, Pantalone KM, Kattan MW
Predicting Current Glycated Hemoglobin Values in Adults: Development of an Algorithm From the Electronic Health Record
JMIR Med Inform 2018;6(4):e10780
URL: http://medinform.jmir.org/2018/4/e10780/
doi:10.2196/10780
PMID:30348631

©Brian J Wells, Kristin M Lenoir, Jose-Franck Diaz-Garelli, Wendell Futrell, Elizabeth Lockerman, Kevin M Pantalone, Michael W Kattan. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 22.10.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing the Impacts of Integrated Decision Support Software on Sexual Orientation Recording, Comprehensive Sexual Health Testing, and Detection of Infections Among Gay and Bisexual Men Attending General Practice: Observational Study

Denton Callander[1,2], PhD; Christopher Bourne[1,3,4], MM; Handan Wand[1], PhD; Mark Stoové[5,6], PhD; Jane S Hocking[7], PhD; John de Wit[2], PhD; John M Kaldor[1], PhD; Basil Donovan[1,3], MD; Catherine Pell[8], MM; Robert Finlayson[8], MB; David Baker[9], MB; Bradley Forssman[10], MBBS; BK Tee[11], MBChB; Bill Kefalas[12], MBBS; Tim Duck[13], MPH; Rebecca Guy[1], PhD

[1]The Kirby Institute, University of New South Wales Sydney, Sydney, NSW, Australia
[2]Centre for Social Research in Health, University of New South Wales Sydney, Sydney, NSW, Australia
[3]Sydney Sexual Health Centre, Sydney Hospital, Sydney, NSW, Australia
[4]Sexually Tranmissible Infection Programs Unit, New South Wales Health, Sydney, NSW, Australia
[5]Burnet Institute, Melbourne, VIC, Australia
[6]School of Public Health and Preventative Medicine, Monash University, Melbourne, VIC, Australia
[7]Melbourne School of Population and Global Health, Melbourne University, Melbourne, VIC, Australia
[8]Taylor Square Private Clinic, Sydney, NSW, Australia
[9]East Sydney Doctors, Sydney, NSW, Australia
[10]Fountain Street General Practice, Sydney, NSW, Australia
[11]The Centre Clinic, Victorian AIDS Council, Melbourne, VIC, Australia
[12]University of New South Wales Health Service, University of New South Wales Sydney, Sydney, NSW, Australia
[13]New South Wales Ministry of Health, Sydney, NSW, Australia

Corresponding Author:
Denton Callander, PhD
The Kirby Institute
University of New South Wales Sydney
Wallace Wurth Building
Sydney, NSW, 2052
Australia
Phone: 61 2 9385 9958
Fax: 61 2 9385 0920
Email: d.callander@unsw.edu.au

Abstract

Background: Gay and bisexual men are disproportionately affected by HIV and other sexually transmissible infections (STIs), yet opportunities for sexual health testing of this population are often missed or incomplete in general practice settings. Strategies are needed for improving the uptake and completeness of sexual health testing in this setting.

Objectives: The goal of the research was to evaluate the impact of an intervention centered around integrated decision support software and routine data feedback on the collection of sexual orientation data and sexual health testing among gay and bisexual men attending general practice.

Methods: A study using before/after and intervention/comparison methods was undertaken to assess the intervention's impact in 7 purposively sampled Australian general practice clinics located near the urban centers of Sydney and Melbourne. The software was introduced at staggered points between April and August 2012; it used patient records to prompt clinicians to record sexual orientation and accessed pathology testing history to generate prompts when sexual health testing was overdue or incomplete. The software also had a function for querying patient management system databases in order to generate de-identified data extracts, which were used to report regularly to participating clinicians. We calculated summary rate ratios (SRRs) based on quarterly

XSL•FO
RenderX

trends and used Poisson regression analyses to assess differences between the 12-month preintervention and 24-month intervention periods as well as between the intervention sites and 4 similar comparison sites that did not receive the intervention.

Results: Among 32,276 male patients attending intervention clinics, sexual orientation recording increased 19% (from 3213/6909 [46.50%] to 5136/9110 [56.38%]) during the intervention period (SRR 1.10, 95% CI 1.04-1.11, *P*<.001) while comprehensive sexual health testing increased by 89% (305/1159 [26.32%] to 690/1413 [48.83%]; SRR 1.38, 95% CI 1.28-1.46, *P*<.001). Comprehensive testing increased slightly among the 7290 gay and bisexual men attending comparison sites, but the increase was comparatively greater in clinics that received the intervention (SRR 1.12, 95% CI 1.10-1.14, *P*<.001). In clinics that received the intervention, there was also an increase in detection of chlamydia and gonorrhea that was not observed in the comparison sites.

Conclusions: Integrated decision support software and data feedback were associated with modest increases in sexual orientation recording, comprehensive testing among gay and bisexual men, and the detection of STIs. Tests for and detection of chlamydia and gonorrhea were the most dramatically impacted. Decision support software can be used to enhance the delivery of sexual health care in general practice.

(JMIR Med Inform 2018;6(4):e10808) doi:10.2196/10808

KEYWORDS

men who have sex with men; general practice; sexual health; software; STI

Introduction

In most high-income settings, the prevalence of HIV and other sexually transmissible infections (STIs) is high among gay, bisexual, and other men who have sex with men [1-3]. Combatting this disproportionate burden requires, among other strategies, routine and frequent sexual health testing, particularly among men whose sexual practices place them at risk of infection [4]. For this reason, clinical guidelines in countries like Australia, Canada, and the United States recommend that sexually active gay and bisexual men should receive a comprehensive sexual health screen at least once per year and more frequently as dictated by sexual risk [5-7].

Australian guidelines during this study defined a comprehensive screen for gay and bisexual men as one that involved tests for rectal and urogenital chlamydia, rectal and pharyngeal gonorrhea, infectious syphilis, and (among men not known to be infected), HIV [8]. The importance of comprehensive testing has been underscored by previous research, with one study finding that 60% of gonorrhea infections and 80% of chlamydia in gay and bisexual men would be missed if rectal swabs were not collected [9]. And although approximately three-quarters of gay men in Australia receive some form of sexual health testing annually, far fewer (37%) report receiving a comprehensive screen [10].

In many countries, general practice is responsible for a large amount of sexual health testing. In Australia, just over half of gay and bisexual men report receiving sexual health testing from general practice clinics [11,12] and the regularity with which people attend general practices makes them ideal for routine testing. Some general practitioners, however, are uncomfortable discussing issues of sexuality with patients or simply forget to raise sexual health due to a focus on the primary reason for presentation and other competing demands [13,14]. Further, studies have found that general practitioners rarely take patients' sexual histories or record their sexual orientation [15-17], which is vital information for guiding any approach to sexual health care. Collectively, these factors may challenge the quality and completeness of sexual health care to gay and bisexual men attending general practice clinics. This contention is supported by research that found gaps in the uptake of sexual health testing among gay and bisexual men attending Australian general practice clinics [18].

One way to enhance health care provision is through computerized clinical decision support systems. Systems that prompt clinicians or provide them with tools to make clinical decisions have been shown in diverse fields of health to improve patient outcomes [19]. While a few studies have shown that clinician prompts can improve rates of testing for HIV and other STIs [20-22], nearly all have been based in sexual health clinics and focused on offering sexual health testing to all patients or all members of a particular population. In general practice, patients tend to be seen frequently and for diverse reasons [23], so any decision support system must consider an individual's sexual risk and testing history.

The aim of this paper was to assess if decision support software can improve the delivery of sexual health care in general practice. To that end, we designed and implemented a computerized clinical decision support system that aimed to improve the recording of patient sexual orientation and promote comprehensive sexual health testing among gay and bisexual men. This paper evaluates the clinical impacts of this intervention, known as The eTEST Project.

Methods

Study Design

To assess the intervention's impact, we undertook a quantitative observational study design involving before-after time series analyses at sites that received the intervention and assessment of concurrent trends between intervention and comparison clinics.

Study Sites and Their Patients

Intervention Sites

Purposive sampling was undertaken to recruit 7 general practice clinics with minimum annual caseloads of 50 individual gay and bisexual male patients. All practices were located in Sydney

XSL•FO
RenderX

and Melbourne, urban centers with the largest populations of gay and bisexual men in Australia and where approximately half of men report receiving sexual health care in general practice [11,12]. This study was limited to urban centers for practical reasons, as recruiting and supporting clinics in regional and remote areas would have exceeded available funding. Clinics were identified for recruitment through consultation with organizations representing general practice, sexual health, HIV medicine, and the health of gay and bisexual men. We also located clinics in or around neighborhoods with high concentrations of same-sex partnered households using Australian census data [24] and by posting study advertisements in medical newsletters. Of note, no sites contacted us to participate, suggesting that advertisements were ineffective for recruiting clinics to this kind of intervention research.

Our scoping exercise identified 28 potential sites, which through consultation with research partners in sexual health and general practice was reduced to the 19 most likely to see reasonably sized caseloads of gay and bisexual men. Potential sites were sent an introductory letter or email that outlined the study and proposed an in-person meeting. After introductory information was sent, 12 in-person meetings were undertaken. From those meetings, 3 clinics were found to not have sufficient numbers of gay and bisexual male patients, resulting in 9 sites recruited to participate. Two sites withdrew participation because their practice computers did not meet the minimum requirements for installing and operating the software.

Comparison Sites

In addition to the intervention sites, a convenience sample of comparison sites was created by extracting data from the patient management systems of clinics based in urban areas of Sydney and Melbourne. We selected comparison sites because it was possible to extract data from their patient management systems and they bore similarities to the intervention clinics, each with a minimum of 50 individual gay and bisexual men seen annually and located within 10 kilometers of the intervention sites. Comparison sites were identified as potential intervention clinics but ultimately did not receive the intervention either because they were not interested or their patient management system was incompatible with the intervention software. The number of intervention sites was limited given the general rarity of “gay” general practice clinics of this kind in Sydney and Melbourne.

Study Intervention

We designed a computerized clinical decision support system for sexual health in general practice. The software built upon an existing piece of technology used in general practice known as the PrimaryCare Sidebar, which worked by querying patient databases and using those queries to generate prompts, produce assessments, or trigger patient recalls. One of the 7 study sites used a patient management system incompatible with the study software and, therefore, participated using a modified version of the intervention that involved establishing prompts using existing built-in assessment tools.

A sexual health-specific module was added to the existing Sidebar software. The module included a custom-built sexual history tool that facilitated the assessment and recording of sexual risk practices for gay and bisexual male patients. The tool routed patients into 3 simple categories linked to testing frequency recommendations: high risk (testing every 3 to 6 months: ≥10 sexual partners in the past 6 months or condomless anal sex with a casual partner), medium risk (annual testing: any anal sex in past 6 months), and low risk (test as needed: no sex in past year or in a monogamous relationship). The assessment dialogue also included links to online sexual health testing guidelines [25] and a partner notification resource to which patients could be referred [26].

Additionally, the sexual health module of Sidebar included a system of electronic prompts (Figure 1). These prompts were triggered by opening of the patient records and were dynamic but passive in that they appeared after a patient record was opened and faded shortly thereafter. The fading function was specifically requested by participating doctors in order to reduce unnecessary interference with consultations. Sexual health prompts were activated by the following events:

- Sexual orientation details were not included in a patient’s record.
- A risk level assessment had not been recorded for a gay or bisexual patient.
- Sexual health testing for a gay or bisexual patient was due or incomplete.

Prompts for HIV and STI testing were triggered on the basis of pathology records in the patient management system, which were automatically downloaded from servicing laboratories. Prompts for testing were also generated depending on assessed risk, with clinicians prompted 6 months after previous testing for patients assessed as high risk and after 12 months for medium-risk patients. Of note, prompts were only generated if a patient’s electronic record were opened. Following a 6-month pilot at one clinic, staggered implementation of the software occurred between April and August 2012 and operated in each clinic for a minimum of 24 months with clinics finishing between April 2014 and August 2014. An overview of the prompt system and risk assessment dialogue is provided as a video in Multimedia Appendix 1.

Beyond the user-facing component, the intervention software also facilitated routine extraction of de-identified data from male patients attending each service. To do so, the software would query the patient database to generate a comma-separated values table based on a customized schema for data extraction. The table included line-listed patient data but without any identifiers, such as address or name. These data were used to evaluate the software but also included as part of the intervention itself. Specifically, clinics were provided biannual reports on testing trends, test positivity, and sexual orientation recording at their service. These reports were routinely presented at clinic meetings to allow doctors a chance to discuss the data, ask questions, and share feedback on the software itself. Individually, clinicians also received tailored emails that focused on different indicators specific to their patients and with comparisons within and between clinics. In these ways, clinical data were used like a model of quality improvement.

XSL•FO
RenderX

Intervention Impact

Evaluation Pathway

The intervention's impact was assessed along a clinical pathway of 3 outcome variables: sexual orientation recording, HIV and/or STI testing uptake, and comprehensive sexual health testing. Figure 2 provides an overview of this pathway. As noted, the intervention was introduced at staggered points throughout 2012, which required us to organize the study period into 12 quarters: preintervention period (quarters 1 through 4) and intervention period (quarters 5 through 12). Comparison sites were also organized into quarters, which were established at the mean entry point for the intervention sites, meaning that quarter 1 for comparison sites began in June 2012 and quarter 12 ended in May 2014. Stata version 14 (StataCorp LLC) was used for all analyses.

Sexual Orientation

Sexual orientation recording was calculated as the proportion of attending male patients for whom sexual orientation details were collected. To assess changes over time, we calculated quarterly trends in the proportion of male patients with recorded sexual orientation in the pre- and intervention periods via summary rate ratios (SRRs). SRRs are useful for assessing relative differences in an event occurring over a fixed time frame, which we compared between the pre- and intervention periods using Poisson regression analysis. Unfortunately, differences in how data were extracted between intervention and comparison sites meant that it was not possible to compare changes in sexual orientation between these two clinical groups.

Figure 1. Record-generated, passive prompt dialogues in a sexual health–specific module for decision support software in general practice that encourage clinicians to record patient sexual orientation, collect a sexual risk assessment, and conduct sexual health testing.

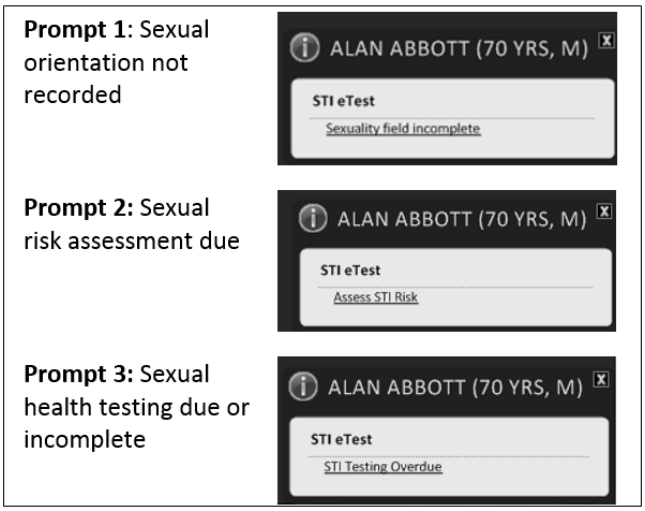
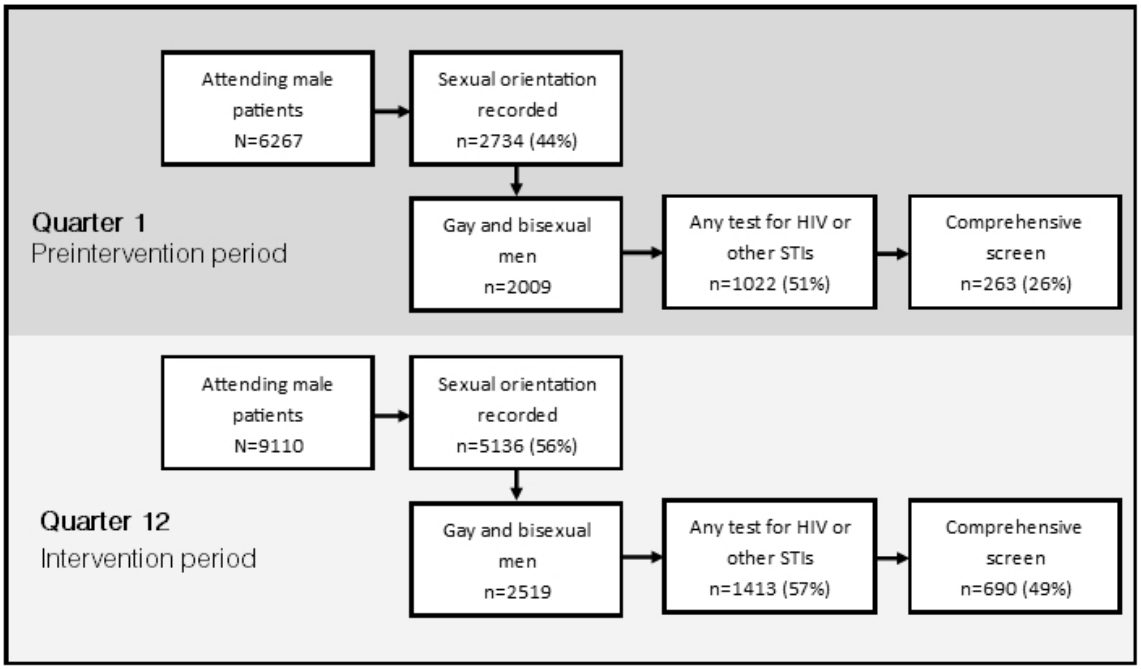


Figure 2. Clinical pathway for assessing intervention impact between the pre- and intervention periods among sites that received the intervention. STI: sexually transmissible infection.



XSL•FO
RenderX

Uptake of Testing for HIV and Sexually Transmissible Infections

The second indicator of the intervention's impact focused on the proportion of men recorded as gay and bisexual men who had any test for HIV or other STI. As with sexual orientation, we calculated this on a quarterly basis and calculated the SRR for the pre- and intervention periods, which were compared using a Poisson regression analysis. Changes over time were also assessed between intervention and comparison sites using a Poisson regression analysis.

In order to facilitate comparisons over time, it was necessary to apply patient sexual orientation across the entire study period. This means that if a patient was later recorded as gay or bisexual we categorized him as such for the purposes of calculating these indicators. It is also important to note that because sexual orientation recording was very low among men attending clinics in the comparison group, to facilitate comparison we used a history of rectal swabs for STI testing to identify patients likely to be gay or bisexual. This has been shown previously to be an effective proximal marker for this population [27], which while not ideal was a necessity for comparative purposes.

Comprehensive Sexual Health Testing

Comprehensive sexual health testing was calculated as the proportion of gay and bisexual men who received any HIV or STI test—chlamydia (rectal or urogenital), gonorrhea (rectal or pharyngeal), syphilis and, among men not known to be infected, HIV. SRRs were calculated and differences over time and between intervention and comparison sites were assessed.

Detection of HIV and Other Sexually Transmissible Infections

Finally, we assessed changes to the detection of HIV and other STIs using positive pathology or, for infectious syphilis, by reviewing historical test results and the interpretative comments provided by labs. In situations where infectious syphilis could not be determined (ie, insufficient information), the result was excluded. SRRs were calculated to assess the mean number of infections diagnosed quarterly during the pre- and intervention periods, which were compared using a Poisson regression analysis. A similar analysis was conducted to assess differences in detection of infections between the intervention and comparison sites.

Data Sources

We collected patient data using a data extraction component of the intervention software. De-identified, line-listed patient data were extracted for the 12-month pre- and 24-month intervention periods. For male patients aged 14 years and older, we extracted the following details per patient visit: unique identifier, age, home postcode, Indigenous status, sexual orientation, HIV status, visit date, visit reason, provider, and HIV/STI testing conducted. Because the results of pathology were downloaded into patient management systems as free text, they could not be extracted directly from participating clinics. Thus, parallel HIV and STI pathology results for all male patients were extracted from the laboratories that serviced participating sites.

Data for comparison sites were extracted from an existing sentinel surveillance network for bloodborne viruses and STIs. The Australian Collaboration for Coordinated Enhanced Sentinel Surveillance (ACCESS) routinely extracts de-identified patient data from a range of clinical sites across Australia and provided the comparison data for sites not participating in the study intervention. Details on this project have been published previously [28].

Ethical Review

Ethical review of this study was provided by the University of New South Wales Human Research Ethics Committee (HC10310). Informed consent was obtained from general practitioners based at clinics that received the intervention.

Results

Study Sites and Their Patients

Intervention Sites

In total, 7 general practices participated in the study intervention, all of which were located in inner urban areas. Participating clinics employed between 3 and 17 general practitioners in full or part-time service. In total, 66 general practitioners participated in the intervention, of which 28 (42%) were female.

During the 3-year observation period, 32,276 individual male patients aged 14 years and older attended intervention clinics with a range of 1905 to 8711 patients per clinic. The median age at baseline was 46 years (interquartile range: 36-56 years), the majority of patients (24,007/32,276, 74.38%) were HIV negative and, reflecting the exclusively urban nature of participating sites, most patients (31,375/32,276, 97.21%) lived in major cities. Only 53.45% (17,251/32,276) of patients had Indigenous status included in their record, with 0.61% (196/32,276) recorded as being of Aboriginal or Torres Strait Islander background.

Comparison Sites

Four general practice clinics were identified as comparison sites, 2 based in Sydney and 2 in Melbourne. In total, 23,712 individual male patients attended comparison sites during the 24-month study period with a range of 952 to 10,279 male patients per clinic. Demographically, the comparison patient group was similar to those attending intervention sites with a median baseline age of 46 years (interquartile range: 36-56 years). In total, 75.41% (17,881/23,712) of male patients attending these clinics were HIV negative, 94.67% (22,448/23,712) lived in an urban area, and 1.67% (396/23,712) were of Aboriginal or Torres Strait Islander background.

Intervention Impact

Sexual Orientation Recording

Figure 2 outlines the clinical pathway we used to evaluate this intervention, comparing the first quarter of the prestudy period (Q1) with the last quarter of the intervention period (Q12). In the first quarter, 43.63% (2734/6267) of attending male patients had details about their sexual orientation recorded, which remained stable across the prestudy period (2961/6639 [44.60%] in Q4). During the intervention period, however, the proportion

XSL•FO
RenderX

of male patients with sexual orientation details increased from 46.50% (3213/6909) in Q5 to 56.38% (5136/9110) in Q12 (P<.001) with an SRR of the average trend between the before and intervention periods of 1.10 (95% CI 1.04-1.11, P<.001). In intervention sites, increases in recording of sexual orientation were observed across age groups, with the lowest baseline proportion but the greatest change in patients less than 30 years old, increasing by 71% during the intervention period from 17.06% (252/1477) recorded in Q5 to 29.45% (494/1680) in Q12 (P<.001).

Multimedia Appendix 2 provides an overview of the SRRs of sexual orientation recording between the before and intervention periods. As noted, it was not possible to calculate this variable among sites in the comparison group, noting that only 4.26% (1010/22,702) of men attending these clinics had sexual orientation included formally in their record. Using rectal swab details, however, it was possible to identify 30.74% of men attending these sites as either gay or bisexual (7290/23,712 of male patients).

Uptake of Testing for HIV or Sexually Transmissible Infections

As detailed in Figure 2, sexual orientation recording was only the first step in our clinical pathway. Among men recorded as gay or bisexual, we also assessed the proportion who in a quarter had any test for HIV or other STIs. In Q1, 50.87% (1022/2009) of attending gay and bisexual men received some form of sexual health testing, which was stable during the preintervention period (1115/2185 [51.03%] in Q4) and also during the intervention period (1159/2290 [50.61%] in Q5 to 1413/2519 [56.09%] in Q12, P=.9). There did not appear to be a difference in sexual health testing uptake between the pre- and intervention periods (SRR 0.97, 95% CI 0.94-1.00, P=.2) nor was a change observed in comparison sites (SRR 1.00, 95% CI 0.97-1.02, P=.9).

Comprehensive Sexual Health Testing

Among men who received a test for HIV or other STI, only 25.73% (263/1022) in Q1 went on to receive the full complement of tests recommended by guidelines. This proportion remained stable during the preintervention period (306/1115 [27.44%] in Q4, P=.5) but increased during the intervention period (305/1159 [26.32%] in Q5 to 690/1413 [48.83%] in Q12), representing an 88% relative increase in comprehensive testing (P<.001). The SRR comparing the quarterly before and intervention trends of comprehensive testing was 1.37 (95% CI 1.28-1.43, P<.001). Multimedia Appendix 2 provides an overview of the comparative increases in comprehensive sexual health testing among gay and bisexual men by HIV status and age. Increases in comprehensive testing were observed across age groups, including a relative increase of 84% increase among men 30 years and younger (17/46 [37%] to 93/137 [67.9%], P<.001), a 79% increase among men aged 30 to 49 years (174/596 [29.2%] to 417/796 [52.4%], P<.001), and a twofold increase among men aged 50 years and older (72/380 [19.0%] in Q5 to 180/480 [37.5%] in Q12, P<.001). Comprehensive testing doubled also among patients living with HIV (94/560 [16.8%] to 240/665 [36.1%], P<.001).

Tests for HIV and syphilis were, by far, the most common component of testing events among gay and bisexual patients. Prior to the intervention, 83.55% (12,016/14,282) of testing events included syphilis and HIV, which increased to 89.46% (29,934/33,461) for syphilis during the intervention period (P<.001) but remained stable for HIV (P=.1). The overall proportions of tests involving chlamydia and gonorrhea were much lower in the before period: 19.90% (2842/14,282) of testing events included rectal swabs for chlamydia and gonorrhea, which increased to 34.51% (11,547/33,461) during the intervention (P<.001), while urine testing for chlamydia increased from 22.00% (3142/14,282) to 37.67% (12,605/33,461) (P<.001). Pharyngeal swabs for gonorrhea also increased, from 17.13% (2447/14,282) before to 34.71% (11,614/33,461) during the intervention (P<.001).

In the 4 comparison sites, comprehensive testing uptake was lower overall than in the intervention sites but increased over time (Figure 3). During the 2-year period that the intervention was active in other sites, comprehensive testing for HIV and STIs increased from 19.25% (319/1657) to 24.26% (464/1913) among gay and bisexual men (P<.001). Compared with the prior 12 months, the SRR for these sites was 1.18 (95% CI 1.11-1.26, P<.001). Overall, although comprehensive testing for HIV and STIs increased across both intervention and comparison sites, the increase and difference between periods was greater for sites that received the intervention than for those that did not (SRR 1.12, 95% CI 1.10-1.14, P<.001).

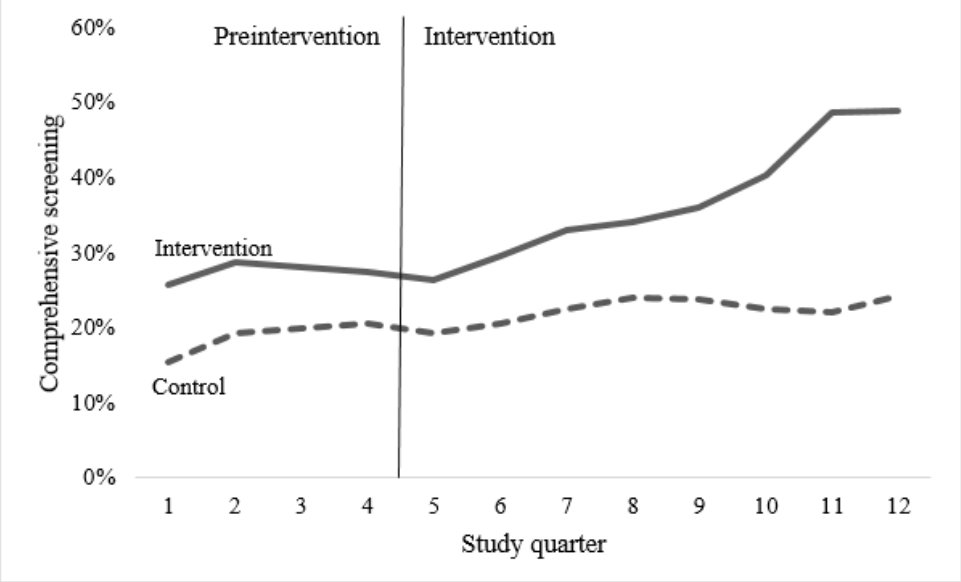
Detection of HIV and Other Sexually Transmissible Infections

Finally, we explored changes in the detection of HIV and STIs between study and comparison sites. While there was a 46% increase in the detection of rectal chlamydia during the intervention, from a mean of 40.0 infections per quarter in the preperiod to 58.5 per quarter of the intervention (SRR 1.28, 95% CI 1.07-1.53, P=.007), there was a nonsignificant increase of 27% in comparison sites (SRR 1.17, 95% CI 0.99-1.37, P=.06). Similarly, while detection of urogenital chlamydia in clinics with the intervention was 44% higher during the intervention period than before (26.5 to 28.0 average per quarter, SRR 1.26, 95% CI 1.01-1.57, P=.04), there was no similar increase among comparison clinics (43.5 to 40.3 average per quarter, SRR 0.85, 95% CI 0.71-1.02, P=.09).

For gonorrhea, detection of rectal infections increased 45% from a mean of 25.4 to 35.0 per quarter during the intervention (SRR 1.27, 95% CI 1.01-1.61, P=.04) but was stable among comparison sites (SRR 1.14, 95% CI 0.97-1.34, P=.1). Pharyngeal diagnoses of gonorrhea were the one infection to increase between both study and comparison sites, rising from 23.8 to 46.0 per quarter in intervention sites (SRR 1.69, 95% CI 1.35-2.11, P<.001) and from 9.0 to 20.0 in comparison sites (SRR 2.04, 95% CI 1.42-2.94, P<.001). In the intervention sites there were no differences in diagnoses of infectious syphilis during the intervention (42.8 to 51.0, SRR 1.05, 95% CI 0.88-1.26, P=.6) or in HIV (13.5 to 13.0, SRR 0.75, 95% CI 0.52-1.05, P=.09), which was the same for comparison sites.

XSL•FO
RenderX

Figure 3. Comprehensive sexual health testing uptake among gay and bisexual men attending intervention and comparison sites.



Of note, in both intervention and comparison sites, only 0.29% of syphilis tests (170/57,993) did not have sufficient information for classification as a new or previously treated infection.

Discussion

Principal Findings

The findings of this study suggest that integrated decision support software in general practice can improve sexual health care for gay and bisexual men. The implementation of software that facilitated routine feedback and prompted general practitioners to collect details on patient sexuality and offer comprehensive sexual health testing was associated with increased recording of sexual orientation among male patients and, among gay and bisexual men, increases in comprehensive sexual health testing in line with clinical guidelines. The observed increases in testing led to increased detection of infections with chlamydia and gonorrhea.

It is worth noting that testing for STIs increased across general practice settings that did and did not receive the intervention. Australia's epidemiology of these infections has documented rising rates for years [2] and the governments of New South Wales and Victoria—the states in which this study took place—have both implemented numerous strategies aimed at increasing sexual health testing [29]. Nevertheless, the intervention appears to have contributed to higher rates of testing and diagnoses than would have otherwise taken place, suggesting a cumulative effective with other initiatives. Although promising, the observed increases were moderate, with the intervention demanding consistent energy to produce data reports and ensure that the software remained operational. Future analyses of integrated decision support software and quality improvement reports should consider the balance between costs and gains.

In spite of the increases in sexual orientation recording among male patients, at the end of the intervention period this variable remained unrecorded for the majority of men. Given the value of knowing sexual orientation for providing care beyond just sexual health [30], additional effort may be required to encourage the collection of this variable among clinicians in general practice. It is possible, however, that for some patients these details were recorded somewhere other than the "official" location in their file, which would not have been captured by our analysis.

The intervention appears to have encouraged greater completeness of testing among those engaged in sexual health care. The prompts, however, did not impact the overall uptake among gay and bisexual men, demonstrated by the stable proportion of men who received any form of sexual health testing. While testing may not have been required for some of the men who received no sexual health testing (ie, sexually inactive men, men who received sexual health care elsewhere), it would seem that the software was useful for capitalizing on existing testing opportunities but not necessarily for creating new ones. Different strategies to improve the overall offer of testing may be warranted, particularly among those not already engaged in sexual health care at a clinic.

While the intervention increased comprehensive testing, this was largely due to more samples being collected for chlamydia and gonorrhea; it had a lesser impact on testing for syphilis and no impact on HIV testing. This finding echoes earlier work assessing the effects of health promotion, which was associated with increased testing for chlamydia and gonorrhea but not syphilis or HIV [29]. It may be that some doctors are unaware of all the different samples required to effectively test for chlamydia and gonorrhea, with our findings echoing earlier work that found anal and throat swabs are among the most commonly missed [10]. It is also possible that patients are more likely to request a test for HIV than other STIs. Thus, the intervention's impact in this domain highlights its usefulness for capitalizing on clinical encounters.

The observed increases in sexual orientation recording and comprehensive testing were gradual, due likely to the time it took for clinicians to become familiar with the software. Further,

XSL•FO
RenderX

the time required to properly calibrate the software to each clinic's technical infrastructure may also have hampered its usefulness in the intervention's earlier days. These factors underscore the need for careful attention and routine follow-up to ensure that newly designed systems are functioning as expected.

Limitations

It is possible that the changes in sexual orientation recording and comprehensive testing were due to some factor unrelated to this study's intervention. This influence of external forces, however, was likely limited by the staggered intervention introduction, the fact that no significant trends were identified before the intervention, and the use of comparison sites. Further, we are unaware of any new clinical activities that occurred before or during the intervention period. It is also worth noting that study recruitment specifically targeted clinics providing care to gay and bisexual men based in major urban centers. As such, it is not possible to generalize these findings to other clinic or patient types, including those in rural or regional areas. Additional research is required to evaluate if this style of decision support software could similarly influence the delivery of sexual health care among other groups of patients.

A limitation of our analysis was the reliance on rectal swabs to identify gay and bisexual men in comparison sites, which while necessary may have actually diluted the intervention's impact. This study was ecological in nature and, as such, required a body of analyses with nonoverlapping limitations. No one analysis proves the intervention's impact but taken together they paint a complementary picture. Finally, noting the careful management required to ensure the intervention's uptake in its early days, more research is required to assess the cost effectiveness of this kind of software and its sustainability over time.

Conclusions

This study provides evidence that computerized clinical decision support systems can be effectively used in general practice to moderately improve sexual health clinical practice among gay and bisexual men. Further, as detecting these infections reduces the likelihood of onward transmission to sexual partners, these systems may have a part to play in reducing community prevalence of HIV and STIs.

Acknowledgments

Funding for this project was provided by the New South Wales Ministry of Health and University of New South Wales (UNSW) Sydney. The authors acknowledge the following people and organizations for their support and guidance: Anna Roberts (Australasian Society for HIV, Viral Hepatitis, and Sexual Health Medicine); Craig Cooper (Positive Life NSW), Geoff Honnor (AIDS Council of New South Wales); Sonny Williams (Positive Life NSW); Marianne Gale (NSW Ministry of Health); Vijay Ramanathan (Royal Australian College of General Practitioners); and Liza Doyle, Larissa Lewis, and Fraser Drummond (Kirby Institute, UNSW Australia). The authors also acknowledge the developers and managers of the PrimaryCare Sidebar sexual health module, in particular Christine Chidgey, Paul Matthews, and Monica Schlesinger (PEN Computer Systems). Data from pathology laboratories were extracted with the help of the following people: Lisa Crawford (Douglas Hanly Moir), Stella Pendle and Adrian Yap (Australian Clinical Labs), Lisa Katon and Joymarie Armstrong (SydPath Central Laboratory), Rob Warren (Victorian Infectious Diseases Reference Laboratory), and Gareth Halbert (Laverty Pathology). And finally, the authors acknowledge the tremendous contribution of practice managers and nursing staff at participating clinics, namely Shakira Watts, Esme Sinnott, Jo Toomey, Marie Stoal, Marion Solomon, Elisha Smith, Claire Johnson, Erin McDonald, Alastair Colgrave, Lucy Kalangi, Cheryl Walker, Danielle Collins, and Simon Powell. For the provision of additional data, the authors acknowledge additional ACCESS investigators who are not authors on this paper, Carol El-Hayek and Margaret Hellard.

Conflicts of Interest

None to declare.

Multimedia Appendix 1

Video guide of sexual health prompts and risk assessment dialogue introduced in general practice clinics via the PrimaryCare Sidebar.

[MP4 File (MP4 Video), 4MB - medinform_v6i4e10808_app1.mp4]

Multimedia Appendix 2

Average annual trends and summary rate ratios in the quarterly proportions of male patients who had sexual orientation recorded, any HIV or other sexually transmissible infection tests, and comprehensive sexual health testing in the pre- and intervention periods.

[PDF File (Adobe PDF File), 26KB - medinform_v6i4e10808_app2.pdf]

References

XSL•FO
RenderX

1. HIV Surveillance Report 2011. Atlanta: Centers for Disease Control and Prevention; 2013. URL: https://www.cdc.gov/hiv/pdf/statistics_2011_HIV_Surveillance_Report_vol_23.pdf [accessed 2018-10-19] [WebCite Cache ID 73IBC82Q8]
2. Australian Annual Surveillance Report of HIV Viral Hepatitis and Sexually Transmissible Infections. Sydney: The Kirby Institute, UNSW; 2017.
3. Sexually Transmitted Disease Surveillance 2012. Atlanta: Centers for Disease Control and Prevention; 2013. URL: https://www.cdc.gov/std/stats12/default.htm [accessed 2018-10-19] [WebCite Cache ID 73IBMt6BI]
4. Steen R, Wi TE, Kamali A, Ndowa F. Control of sexually transmitted infections and prevention of HIV transmission: mending a fractured paradigm. Bull World Health Organ 2009 Nov;87(11):858-865 [FREE Full text] [Medline: 20072772]
5. Templeton DJ, Read P, Varma R, Bourne C. Australian sexually transmissible infection and HIV testing guidelines for asymptomatic men who have sex with men 2014: a review of the evidence. Sex Health 2014 Jul;11(3):217-229. [doi: 10.1071/SH14003] [Medline: 24690473]
6. HIV Screening and Testing Guide. Ottawa: Centre for Communicable Diseases and Infection Control; 2012.
7. Branson BM, Handsfield HH, Lampe MA, Janssen RS, Taylor AW, Lyss SB, et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR Recomm Rep 2006 Sep 22;55(RR-14):1-17 [FREE Full text] [Medline: 16988643]
8. Sexually Transmitted Infection Testing Guidelines for Men who have Sex with Men. Sydney: Sexually Transmissible Infections in Gay Men's Action Group (STIGMA); 2010.
9. Bachmann LH, Johnson RE, Cheng H, Markowitz L, Papp JR, Palella FJ, et al. Nucleic acid amplification tests for diagnosis of Neisseria gonorrhoeae and Chlamydia trachomatis rectal infections. J Clin Microbiol 2010 May;48(5):1827-1832 [FREE Full text] [doi: 10.1128/JCM.02398-09] [Medline: 20335410]
10. Holt M, Hull P, Lea T, Guy R, Bourne C, Prestage G, et al. Comprehensive testing for, and diagnosis of, sexually transmissible infections among Australian gay and bisexual men: findings from repeated, cross-sectional behavioural surveillance, 2003-2012. Sex Transm Infect 2014 May;90(3):208-215 [FREE Full text] [doi: 10.1136/sextrans-2013-051294] [Medline: 24234070]
11. Hull P, Mao L, Kolstee J. Gay Community Periodic Survey: Sydney 2015. Sydney: Centre for Social Research in Health, UNSW; 2015. URL: https://csrh.arts.unsw.edu.au/media/CSRHFile/CSRH_Report__GCPS_Sydney_2015.pdf [accessed 2018-10-19] [WebCite Cache ID 73IBUgS75]
12. Lee E, Mao L, McKenzie T. Gay Community Periodic Survey: Melbourne 2016. Sydney: Centre for Social Research in Health, UNSW; 2018. URL: https://csrh.arts.unsw.edu.au/media/CSRHFile/GCPS_Melbourne_2016.pdf [accessed 2018-10-19] [WebCite Cache ID 73IBZeUG2]
13. Khan A, Plummer D, Hussain R, Minichiello V. Sexual risk assessment in general practice: evidence from a New South Wales survey. Sex Health 2007 Mar;4(1):1-8. [Medline: 17382030]
14. Barber B, Hellard M, Jenkinson R, Spelman T, Stoove M. Sexual history taking and sexually transmissible infection screening practices among men who have sex with men: a survey of Victorian general practitioners. Sex Health 2011 Sep;8(3):349-354. [doi: 10.1071/SH10079] [Medline: 21851775]
15. Callander D, Bourne C, Pell C, Finlayson R, Forssman B, Baker D, et al. Recording the sexual orientation of male patients attending general practice. Fam Pract 2015 Feb;32(1):35-40. [doi: 10.1093/fampra/cmu061] [Medline: 25236604]
16. Wimberly YH, Hogben M, Moore-Ruffin J, Moore SE, Fry-Johnson Y. Sexual history-taking among primary care physicians. J Natl Med Assoc 2006 Dec;98(12):1924-1929. [Medline: 17225835]
17. Meystre-Agustoni G, Jeannin A, de Heller K, Pécoud A, Bodenmann P, Dubois-Arber F. Talking about sexuality with the physician: are patients receiving what they wish? Swiss Med Wkly 2011 Mar 08;141:w13178 [FREE Full text] [doi: 10.4414/smw.2011.13178] [Medline: 21384286]
18. Guy R, Goller JL, Spelman T, El-Hayek C, Gold J, Lim M, et al. Does the frequency of HIV and STI testing among men who have sex with men in primary care adhere with Australian guidelines? Sex Transm Infect 2010 Oct;86(5):371-376. [doi: 10.1136/sti.2009.040972] [Medline: 20460263]
19. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
20. Lister NA, Smith A, Fairley CK. Introduction of screening guidelines for men who have sex with men at an STD clinic, the Melbourne Sexual Health Centre, Australia. Sex Health 2005;2(4):241-244. [Medline: 16402672]
21. Bissessor M, Fairley CK, Leslie D, Chen MY. Use of a computer alert increases detection of early, asymptomatic syphilis among higher-risk men who have sex with men. Clin Infect Dis 2011 Jul 01;53(1):57-58. [doi: 10.1093/cid/cir271] [Medline: 21653303]
22. Zou H, Fairley CK, Guy R, Bilardi J, Bradshaw CS, Garland SM, et al. Automated, computer generated reminders and increased detection of gonorrhoea, chlamydia and syphilis in men who have sex with men. PLoS One 2013;8(4):e61972 [FREE Full text] [doi: 10.1371/journal.pone.0061972] [Medline: 23613989]
23. Britt H, Miller GC, Henderson J, Charles J, Valenti L, Harrison C, et al. General Practice Activity in Australia 2011-12: BEACH, Bettering the Evaluation And Care of Health. Sydney: Sydney University; 2012.

XSL•FO
RenderX

24. Same-Sex Couples. Canberra: Australian Bureau of Statistics; 2013. Australian Social Trends 4102 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4102.0Main+Features10July+2013 [accessed 2018-10-19] [WebCite Cache ID 73IBn4hUj]
25. NSW Government. STI in Gay Men's Action Group sexual health testing guidelines. 2013. URL: https://stipu.nsw.gov.au/stigma/ [accessed 2018-04-17] [WebCite Cache ID 6yl1zl4TA]
26. The Drama Downunder. Sydney: Australian Federation of AIDS Organisations; 2013. Let him know by SMS URL: http://www.thedramadownunder.info/notify/sms/ [accessed 2018-04-17] [WebCite Cache ID 6yl1ncBrW]
27. Ampt FH, El-Hakey C, Agius PA, Bowring AL, Bartnik N, Fairley CK, et al. Anorectal swabs as a marker of male-to-male sexual exposure in STI surveillance systems. Epidemiol Infect 2017 Dec;145(12):2530-2535. [doi: 10.1017/S095026881700098X] [Medline: 28528588]
28. Guy RJ, Kong F, Goller J, Franklin N, Bergeri I, Dimech W, et al. A new national Chlamydia Sentinel Surveillance System in Australia: evaluation of the first stage of implementation. Commun Dis Intell Q Rep 2010 Sep;34(3):319-328. [Medline: 21090187]
29. Wilkinson AL, Pedrana AE, El-Hayek C, Vella AM, Asselin J, Batrouney C, et al. The impact of a social marketing campaign on HIV and sexually transmissible infection testing among men who have sex with men in Australia. Sex Transm Dis 2016 Jan;43(1):49-56. [doi: 10.1097/OLQ.0000000000000380] [Medline: 26650997]
30. Makadon HJ, Mayer KH, Garofalo R. Optimizing primary care for men who have sex with men. JAMA 2006 Nov 15;296(19):2362-2365. [doi: 10.1001/jama.296.19.2362] [Medline: 17105799]

Abbreviations

ACCESS: Australian Collaboration for Coordinated Enhanced Sentinel Surveillance
STI: sexually transmissible infections
SRR: summary rate ratio
UNSW: University of New South Wales

Edited by G Eysenbach; submitted 17.04.18; peer-reviewed by J Constantin, J Bilardi, TW Chien, MJ Rotheram-Borus; comments to author 21.06.18; revised version received 11.07.18; accepted 16.08.18; published 06.11.18.

Please cite as:
Callander D, Bourne C, Wand H, Stoové M, Hocking JS, de Wit J, Kaldor JM, Donovan B, Pell C, Finlayson R, Baker D, Forssman B, Tee BK, Kefalas B, Duck T, Guy R
Assessing the Impacts of Integrated Decision Support Software on Sexual Orientation Recording, Comprehensive Sexual Health Testing, and Detection of Infections Among Gay and Bisexual Men Attending General Practice: Observational Study
JMIR Med Inform 2018;6(4):e10808
URL: http://medinform.jmir.org/2018/4/e10808/
doi:10.2196/10808
PMID:30401672

©Denton Callander, Christopher Bourne, Handan Wand, Mark Stoové, Jane S Hocking, John de Wit, John M Kaldor, Basil Donovan, Catherine Pell, Robert Finlayson, David Baker, Bradley Forssman, BK Tee, Bill Kefalas, Tim Duck, Rebecca Guy. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 06.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of Electronic Versus Paper Vital Sign Observations on Length of Stay in Trauma Patients: Stepped-Wedge, Cluster Randomized Controlled Trial

David CW Wong[1], MEng, DPhil; Julia Knight[2], BSc, MPH; Jacqueline Birks[3], MA, MSc; Lionel Tarassenko[4], MA, DPhil; Peter J Watkinson[2], BM, BCh, MD

[1]Leeds Institute of Health Sciences, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

[2]Kadoorie Centre for Critical Care Research and Education, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

[3]Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom

[4]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom

Corresponding Author:
David CW Wong, MEng, DPhil
Leeds Institute of Health Sciences
Faculty of Medicine and Health
University of Leeds
Worsley Building, Level 10
Leeds, LS29JT
United Kingdom
Phone: 44 113 3439671
Email: d.c.wong@leeds.ac.uk

Abstract

Background: Electronic recording of vital sign observations (e-Obs) has become increasingly prevalent in hospital care. The evidence of clinical impact for these systems is mixed.

Objective: The objective of our study was to assess the effect of e-Obs versus paper documentation (paper) on length of stay (time between trauma unit admission and "fit to discharge") for trauma patients.

Methods: A single-center, randomized stepped-wedge study of e-Obs against paper was conducted in two 26-bed trauma wards at a medium-sized UK teaching hospital. Randomization of the phased intervention order to 12 study areas was computer generated. The primary outcome was length of stay.

Results: A total of 1232 patient episodes were randomized (paper: 628, e-Obs: 604). There were 37 deaths in hospital: 21 in the paper arm and 16 in the e-Obs arm. For discharged patients, the median length of stay was 5.4 (range: 0.2-79.0) days on the paper arm and 5.6 (range: 0.1-236.7) days on the e-Obs arm. Competing risks regression analysis for time to discharge showed no difference between the treatment arms (subhazard ratio: 1.05; 95% CI 0.82-1.35; *P*=.68). A greater proportion of patient episodes contained an Early Warning Score (EWS) ≥3 using the e-Obs system than using paper (subhazard ratio: 1.63; 95% CI 1.28-2.09; *P*<.001). However, there was no difference in the time to the subsequent observation, "escalation time" (hazard ratio 1.05; 95% CI 0.80-1.38; *P*=.70).

Conclusions: The phased introduction of an e-Obs documentation system was not associated with a change in length of stay. A greater proportion of patient episodes contained an EWS≥3 using the e-Obs system, but this was not associated with a change in "escalation time."

Trial Registration: ISRCTN Registry ISRCTN91040762; http://www.isrctn.com/ISRCTN91040762 (Archived by WebCite at http://www.webcitation.org/72prakGTU)

(JMIR Med Inform 2018;6(4):e10221) doi:10.2196/10221

KEYWORDS

vital signs; medical records system, computerized; length of stay; evaluation studies; trauma

XSL•FO
RenderX

Introduction

Background

Physiological vital signs are regularly measured in hospitalized patients. Deterioration in vital signs often precedes adverse outcomes [1,2]. However, vital sign alterations can go unrecognized, resulting in treatment delay that may worsen outcomes [3-5]. One method for identifying deterioration is the Early Warning Score (EWS), in which a score is given to each set of recorded vital signs. The overall score, the EWS, is the aggregate of scores assigned to each individual vital sign, depending on its level of abnormality. Higher scores indicate greater physiological abnormality [6].

Until recently, vital signs and EWS have been recorded on paper observation charts. The shortcomings of paper charts include incorrect score assignment to the vital signs and incorrect calculation of EWS [7-9], difficulty in interpretation [10], and poor compliance with clinical escalation protocols [11]. Electronic methods for recording vital sign observations and EWS, known as e-Obs, are becoming increasingly common [12-14]. e-Obs systems may circumvent many of these highlighted issues by automatically assigning EWS and prompting appropriate clinical response. The facility to display recent observations and scores on a central station may also improve the ability of clinical staff to recognize patient deterioration. However, the clinical impact of such systems is currently unclear, with studies reporting conflicting results [15,16].

Objective

In this study, we prospectively assessed whether the deployment of the VitalPAC e-Obs system (VitalPAC; The Learning Clinic), compared with the paper-based system, changed patients' hospital length of stay (since the conclusion of this study, The Learning Clinic has been acquired by System C, Maidstone, Kent). Our null hypothesis was that length of stay remained unaltered. A cluster-randomized design was not appropriate as the VitalPAC e-Obs system was to be introduced gradually in the trauma unit. We, therefore, evaluated the intervention as it was being introduced using a stepped-wedge study design.

Methods

Study Design

We conducted a randomized stepped-wedge interventional study in the two adult inpatient wards of the trauma unit at the John Radcliffe Hospital, Oxford University Hospitals National Health Service Trust. Each ward had 26 inpatient beds. A stepped-wedge study is one in which the intervention is phased into the study population within clusters across successive time periods, with the time being determined by randomization [17]. The full study protocol is available online [18]. Approval for this study was obtained from the National Health Service Research Ethics Committee (REC #11/H0308/11), and the study was registered with ISRCTN (ISRCTN91040762). Informed written consent was sought from all eligible participants after hospital discharge.

Control and Intervention

Initially, nursing staff measured vital signs (blood pressure, pulse, oxygen saturation, temperature, respiratory rate, and consciousness level) using spot-check monitors and documented the result on paper. Nurses manually calculated the EWS and recorded the score on paper. The EWS used was the previously published centile-EWS [19], in which a score of 3 or more (EWS≥3) requires urgent intervention.

The intervention was the VitalPAC e-Obs system, which allowed vital signs to be documented on a hand-held device. The EWS was then automatically calculated, and relevant hospital guidance for escalation was displayed. EWS were also displayed at a central patient ward list that adjusted colors and symbols to prompt nurses to record timely observations according to the hospital protocol.

Prior to study commencement, VitalPAC was installed and tested and staff were trained to use the system. Refresher teaching was also provided for the paper system, and staff were reminded of local clinical escalation policy. During the training period, the paper system was used throughout the trauma unit, but staff had the opportunity to enter data into a test installation of VitalPAC. Study staff attended the ward daily during the week and once at weekends, providing top-up training throughout the duration of the study.

Trial Design

The e-Obs intervention was phased in 12 clusters. Each cluster was a physical zone within the trauma unit consisting of a 4-bed bay or a collection of six 1-bed side rooms. The clusters received the electronic intervention sequentially. A new zone switched from the control to the intervention every 3 weeks on Tuesday at 2 pm.

Participants

All episodes from patients aged ≥16 years admitted to the trauma unit during the study period were considered for eligibility. Episodes were excluded from analysis if the treatment plan for the patient was palliative at admission to the ward.

Study Data

Clinical and demographic data on study participants were obtained from an electronic patient record (*Casenotes*) and from paper medical records. Research nurses collated the following: age, sex, ethnicity, American Society of Anesthesiologists score, reason for admission, and admission method (emergency, elective, between wards) for each study participant.

We also recorded initial ward and zone on the trauma unit; date and time (hh:mm) of admission into the ward; date, time (hh:mm), and clinician-recorded EWS for the first vital sign observations recorded on the trauma unit; total number of vital sign observations on the unit recorded electronically and on paper; "fit to discharge" date; actual hospital discharge date and time (hh:mm); in-hospital mortality and 30-day mortality following ward admission; and unplanned admission to the intensive care unit (ICU) or a cardiac arrest for each episode.

Two other electronic sources, *Cerner Millenium* (Cerner, Kansas City, MO) and *Bluespier* (Bluespier International,

XSL•FO
RenderX

Worcestershire), were used to validate the information. Mortality status at 30 days after admission was verified using data from the National Health Service personal demographic service.

Outcomes

The primary endpoint was length of stay (the time from admission to the trauma unit until "fit to discharge"). "Fit to discharge" was defined as the first of the following: discharged from the ward to home or alternative care or accepted by social services as a "delayed discharge." This outcome measure was chosen because some trauma patients were known to have extended stays while waiting for suitable support mechanisms to be put in place outside hospital. Secondary endpoints were mortality (in-hospital and 30-day following ward admission), whether a patient experienced a cardiac arrest or an unplanned ICU admission; the time between observations (length of stay/total number of observations); the time until a patient first scored EWS≥3; and the time between the first observation that scored EWS≥3 and the subsequent observation ("escalation time").

Sample Size

The number of participants was determined by the speed at which zones transitioned from paper to e-Obs. Slower transitions would include more participants for each step, thereby increasing the power of the study. However, clinical staff wished to minimize the concurrent use of multiple systems. Therefore, a clinically accepted transition rate of 1 zone every 3 weeks was chosen, and no sample size calculation was undertaken.

Randomization

A random sequence generated using MATLAB (function *randperm* [20]) determined the order in which zones received e-Obs.

All zones were recruited and enrolled at baseline and followed for the entire duration of the study. Research nurses administered the sequential assignment to the e-Obs intervention. The nurses visited the ward during each transition to facilitate adherence to the change from paper to e-Obs.

Patients were allocated to either paper or electronic recording of observations based on the zone to which they were allocated on arrival to the trauma unit. Patients remained with the same recording method even if subsequently moved. Allocation to the initial zone was determined by normal ward practices, which remained unchanged during the study period. Therefore, the allocation ratio could not be determined a priori.

Blinding

The randomization sequence was concealed from patients, clinical staff, and all researchers involved in data collection until the day a new zone was due to receive e-Obs. It was not possible to conceal the intervention.

Statistical Methods

Episode characteristics, including patient age and length of stay, were summarized by study arm. A time-to-event analysis, Cox proportional hazards regression, was undertaken for length of stay with the competing risk of death in hospital, for death in hospital with discharge from hospital as a competing risk, and for time to first EWS≥3. The intervention arm, step, age, and sex of patients were included as covariates. Subhazard ratios were calculated with respect to paper charts, age>80 years, and male sex. The step was included as a continuous measure of time and as a factor with 13 levels. The study design has two levels, patient and study zone. The SEs of the coefficients in the Cox proportional hazards model were adjusted for the cluster variable, study zone. A similar competing risks Cox regression analysis was undertaken for "escalation time," censored at 200 hours.

The binary outcome, death within 30 days of admission, was analyzed using a logistic regression model and the SEs of the coefficients were adjusted for the cluster variable, study zone. The numbers of cardiac arrests and admissions to ICU were reported.

All analyses were performed on an intention-to-treat basis, with patients analyzed according to the randomization intervention. Furthermore, all analyses were completed for hospital episodes. A patient may have had multiple episodes consisting of distinct admissions to the study wards for unrelated reasons. We assumed a priori that the number of such admissions was small such that each episode may be treated as an independent event. Post hoc, we repeated the analyses on the per-protocol populations.

All statistical analyses were performed using Stata (Stata Statistical Software: Release 14. College Station, TX; StataCorp LP) [21].

Results

A total of 1518 admissions to the trauma unit were recorded between August 31, 2011 and May 31, 2012. After excluding 286 episodes that did not meet the study criteria or for which patients had declined consent, 1232 episodes (from 1199 patients) were included for analysis (Figure 1).

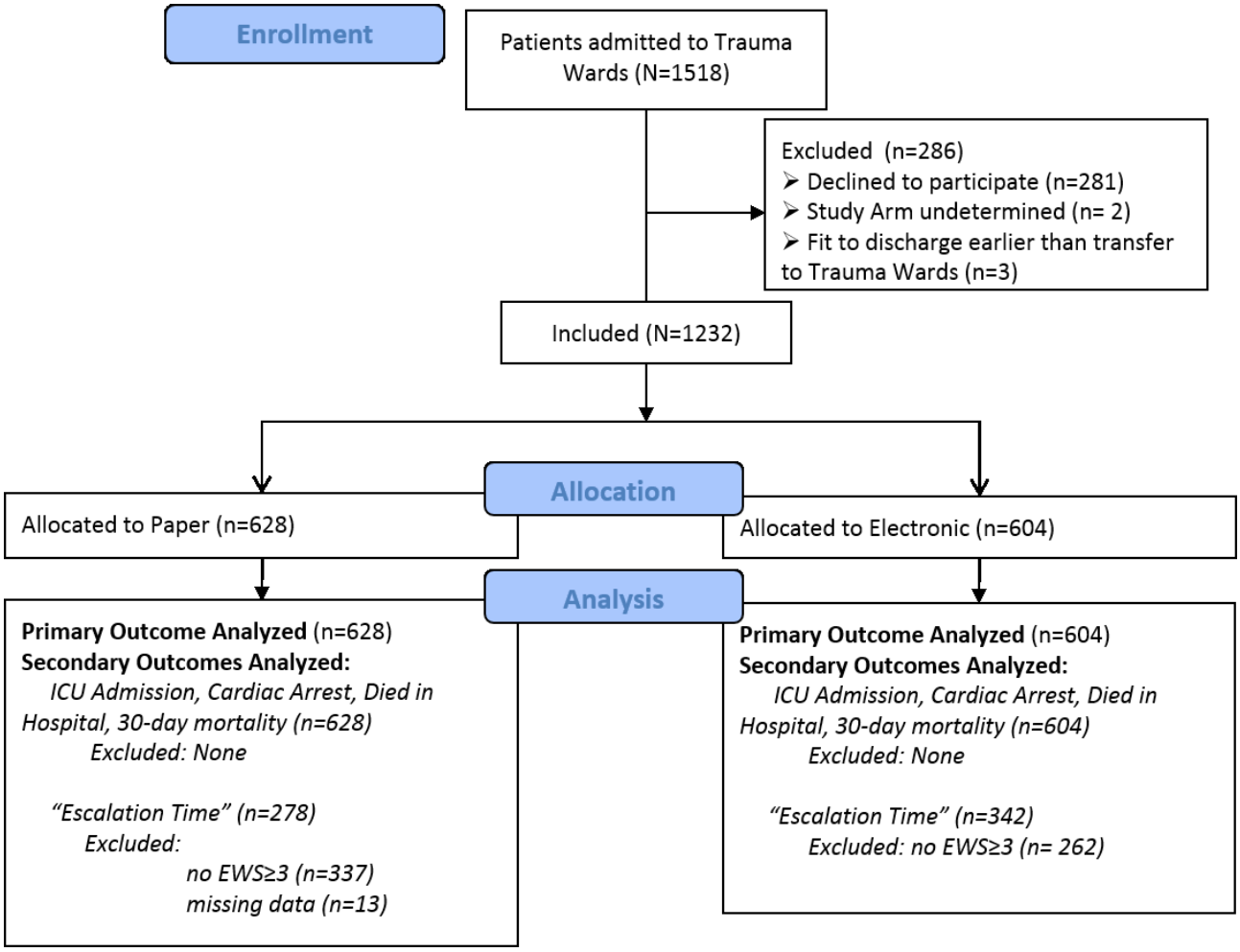
Of the included episodes, 628 were randomized to paper and 604 to e-Obs. Moreover, 873 episodes (paper: 558, e-Obs: 315) had vital sign observations that were fully consistent with the randomized intervention. A further 194 episodes (paper: 32, e-Obs: 162) had over 80% of observations on the allocated intervention. Vital sign observation charts were absent from the paper notes for 13 episodes.

Baseline characteristics of the study patients are shown in Table 1. Allocation between the study arms was almost equal (paper: 628/1232, 50.97%). There were no significant differences in any of the measured characteristics. In all analyses, the intraclass correlation coefficient for patients within a study area was not significantly different from zero.

There were 37 deaths in hospital: 21 in the paper arm and 16 in the e-Obs arm. For patients who were discharged, the median length of stay (time from admission to "fit to discharge") was 5.4 (range: 0.2-79.0) days on the paper arm (607 patients) and 5.6 (range: 0.1-236.7) days on the e-Obs arm (588 patients). Longer time to discharge was associated with greater age, but there was no difference between the treatment arms (Table 2).

XSL•FO
RenderX

Figure 1. Flow chart for patients' enrollment in the study. "Escalation Time" is the time between the first observation that scored Early Warning Score (EWS)≥3 and the subsequent observation. ICU: intensive care unit.



Results from the competing risks Cox regression analysis for time to death in hospital with discharge from hospital as a competing risk are reported in Table 2. There was no difference between the treatment arms.

There were 45 deaths within 30 days from admission to hospital: 23 in the paper arm and 22 in the e-Obs arm. The results from the logistic regression analysis of the number of deaths within 30 days of admission are reported in Table 2. There was no difference between the treatment arms.

There were 8 admissions to the ICU from the paper arm and 7 from the e-Obs arm, and there were 2 cardiac arrests in each arm. The median time between observations for those without a delayed discharge was 7.1 (interquartile range [IQR]: 5.0-9.8) hours on paper and 7.0 (IQR: 5.3-9.5) hours on e-Obs.

Figure 2 shows the Kaplan-Meier estimates of time to first EWS≥3. At least one EWS≥3 was recorded for 278 and 342 episodes in the paper and e-Obs arms, respectively. The number at risk indicates the episodes that had not had an EWS≥3 after {0, 200, 400, 600, 800, and 1000} hours. Numbers in parentheses are those who had an EWS≥3 before the next time point. Moreover, 7 episodes in the e-Obs arm had no recorded observations, while 13 episodes in the paper arm had missing data (observation chart missing from notes). The subhazard ratio for e-Obs with respect to paper from a competing risks Cox regression analysis for time to first EWS ≥3 with death in hospital as a competing risk including age>80 years, step, and sex was 1.63 (95% CI 1.28-2.09; *P*<.001).

Figure 3 shows the Kaplan-Meier estimates of *escalation time*. Inclusion in the analysis was conditional on reaching an EWS of ≥3. The number at risk indicates the episodes that had not yet had an observation after {0, 25, 50,..., 200} hours. Numbers in parentheses are those who had the next set of observations before the next time point. Patients could be censored at any time due to discharge from hospital or death in hospital. Three episodes had the next observation after more than 50 hours, 18 were discharged from hospital, and 3 died in the hospital before a further set of observations. The hazard ratio for e-Obs with respect to paper from a competing risk Cox regression analysis for time to escalation including age>80 years, step, and sex was 1.05 (95% CI 0.80-1.38; *P*=.70).

All results presented here are for intention-to-treat analysis. The results of the per-protocol analysis are shown in Multimedia Appendix 1. On per-protocol analysis, we found no difference in "escalation time" or length of stay. Per-protocol in-hospital time to mortality analysis suggested a mortality benefit in favor of e-Obs, but this was not sustained at 30 days.

XSL•FO
RenderX

Table 1. Patient baseline characteristics (N=1232 episodes).

Characteristic	Paper (n=628)	e-Obs[a] (n=604)
Age (years) mean (SD)	58.1 (23.4)	60.4 (23.3)
Males, n (%)	308 (49.0)	316 (52.3)
Ward 1:Ward 2	312:316	280:324
Ethnicity, n (%)		
White British	471 (74.9)	446 (73.6)
Not stated	110 (17.5)	129 (21.3)
Other	48 (7.6)	31 (5.1)
Reason for admission: injury type, n (%)		
Lower limb (excluding neck of femur)	179 (28.5)	188 (31.1)
Neck of femur	160 (25.5)	169 (28.0)
Upper limb (excluding wrist)	59 (9.4)	63 (10.4)
Polytrauma (excluding head)	48 (7.6)	45 (7.5)
Wrist	38 (6.1)	21 (3.5)
Spinal trauma	33 (5.3)	30 (5.0)
Polytrauma + head	29 (4.6)	23 (3.8)
Nontrauma	29 (4.6)	15 (2.5)
Other	53 (8.4)	50 (8.3)
Primary specialty, n (%)		
Trauma	581 (92.5)	574 (95.0)
Other	47 (7.5)	30 (4.9)
Admission method, n (%)		
Emergency department (not via intensive care unit)	323 (51.4)	342 (56.3)
Trauma clinic within Oxford University Hospitals Trust	108 (17.2)	107 (17.6)
Other ward	89 (14.3)	57 (9.7)
Transfer from other United Kingdom hospital	24 (3.8)	24 (4.0)
Emergency department via emergency admissions unit	50 (7.9)	50 (8.2)
Other	34 (5.4)	24 (4.0)
American Society of Anesthesiologists score, n (%)		
1	101(16.1)	97 (16.1)
2	97 (15.4)	107 (17.7)
3	92 (14.6)	86 (14.2)
4	18 (2.9)	22 (3.7)
5	3 (0.5)	1 (0.2)
1E-5E	11 (1.8)	12(2.0)
Unrecorded	180 (28.7)	169 (28.0)
Not applicable	126 (20.0)	110 (18.2)

[a]e-Obs: electronic recording of vital sign observations.

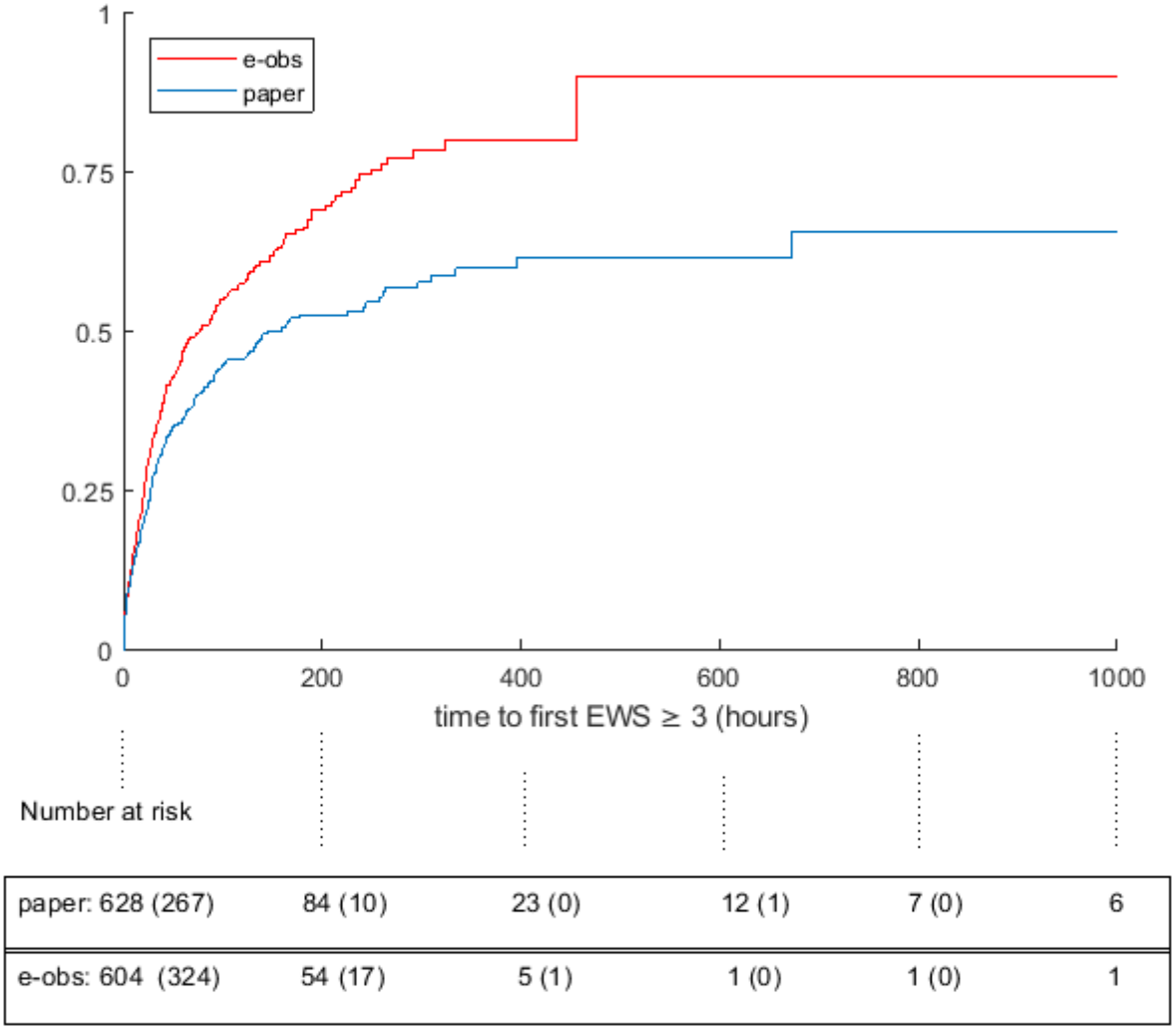
XSL•FO
RenderX

Table 2. Results from the competing risks regression analysis for time to discharge from hospital with death in hospital as a competing risk, from the competing risks regression analysis for time to death in hospital with discharge from hospital as a competing risk, and the logistic regression for number of deaths within 30 days from admission.

Variable	Time to discharge from hospital with death in hospital as a competing risk		Time to death in hospital with discharge from hospital as a competing risk		Number of deaths within 30 days from admission	
	Subhazard ratio (95% CI)	*P* value	Subhazard ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
e-Obs[a]	1.05 (0.82-1.35)	.68	0.77 (0.42-1.40)	.39	0.82 (0.46-1.47)	.51
Step	1.00 (0.97-1.03)	.99	0.98 (0.88-1.10)	.72	1.01 (0.92-1.12)	.80
Age ≥80 years	0.62 (0.57-0.67)	<.001	5.69 (3.48-9.30)	<.001	8.93 (5.69-14.01)	<.001
Female	1.09 (0.96-1.25)	.19	0.63 (0.35-1.12)	.11	0.60 (0.39-0.94)	.02

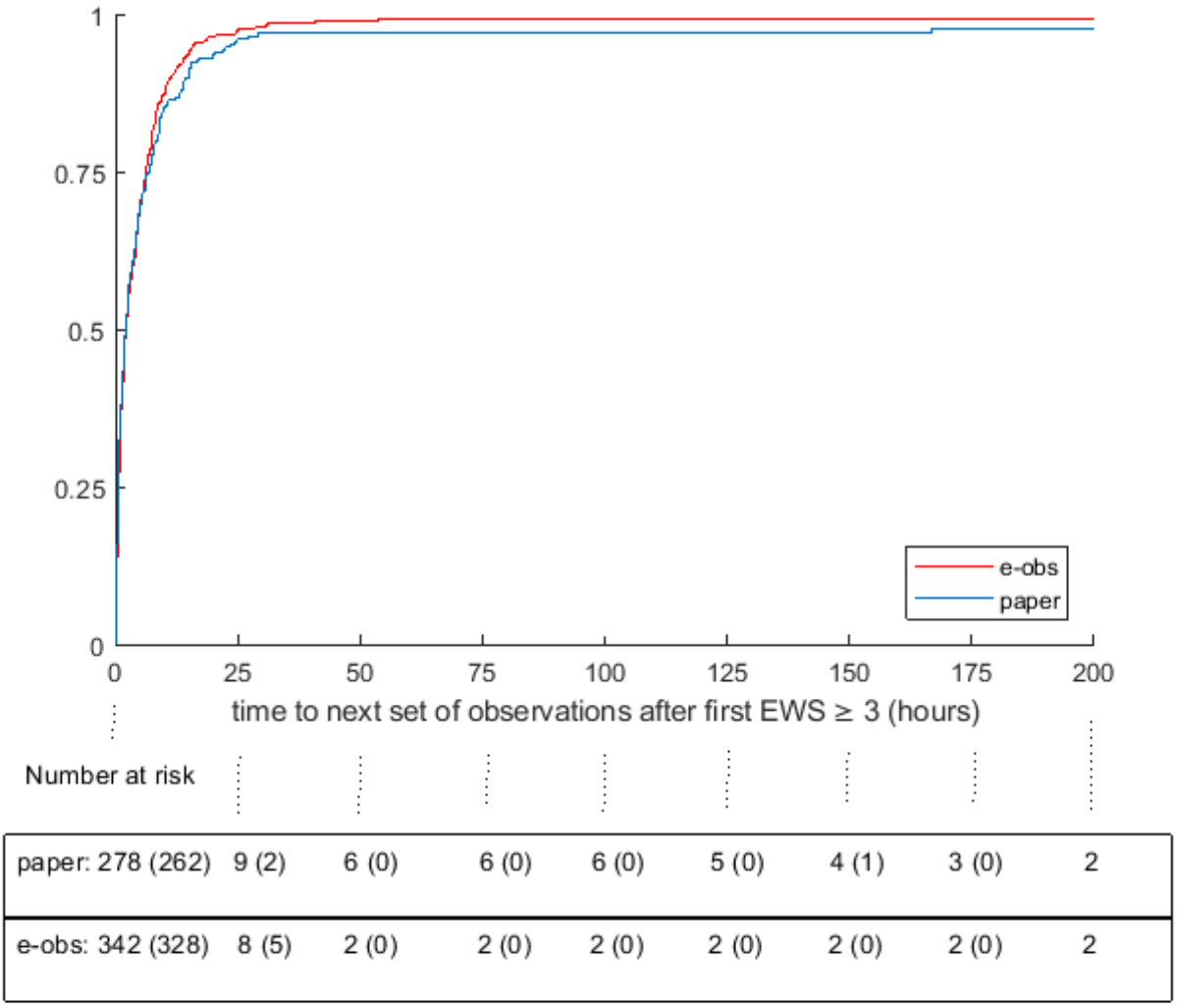
[a]e-Obs: electronic recording of vital sign observations.

Figure 2. Kaplan-Meier failure estimates for time (hours) from admission until first Early Warning Score (EWS)≥3. e-Obs: electronic recording of vital sign observations.



XSL•FO
RenderX

Figure 3. Kaplan-Meier failure estimates for time (hours) to next set of observations after the first Early Warning Score (EWS)≥3. e-Obs: electronic recording of vital sign observations.



Discussion

Principal Findings

To our knowledge, this is the first stepped-wedge evaluation of an e-Obs intervention. We found no difference in length of stay (the primary outcome) between paper and e-Obs. A significantly greater proportion of patient episodes contained an EWS≥3 using the e-Obs system than using paper. However, we found no difference in "escalation time." Per-protocol analysis showed improved in-hospital survival with e-Obs, but this was not shown either on intention-to-treat analysis or for outcome at 30 days. The absence of intraclass correlation indicates that patient groups were well matched at each individual step level.

Limitations

Although 86.61% (1067/1232) of patients had over 80% of their observations in the correct format, cross over was unequal. Only 79.0% (477/604) of patients in the electronic observations arm had more than 80% of their observations recorded electronically. The presence of multiple documentation systems in the wards may have led to confusion and suboptimal system use. To avoid this issue, future studies might ensure clear physical separation between zones to prevent study arm contamination. Our follow-up study over 4 hospitals instead considers whole wards as a study zone [22].

Length of stay is a complex outcome, risking confounding by competing interventions. However, it is an important outcome for patients and their families [23], as well as clinicians and managers, and has been recommended as a potential outcome measure within the National Institute for Health and Care Excellence recommendations for recognizing patient deterioration in hospital [24]. We used it as our primary outcome as delayed detection of deterioration might be expected to increase length of stay even where the delay or deterioration

XSL•FO
RenderX

was not sufficient to warrant ICU admission or cause death. Indeed, improved recognition of clinical deterioration is associated with earlier discharge [25] and has previously been used in observational studies assessing e-Obs systems [13,15,26]. To minimize the effect of competing interventions, we used "fit to discharge" rather than actual discharge time and took account of the competing risk of death in our analysis. We included other outcome measures that would be expected to be associated with more extreme delays in recognizing deterioration (death and ICU admission) as secondary outcome measures.

Interpretation

In this study, more patient episodes contained an EWS≥3 using the e-Obs system than paper. We have previously shown paper documentation errors to be biased toward values lower than the true EWS, particularly when a patient first develops physiological instability [27]. As the groups were otherwise well balanced, it is likely that the result is explained by a bias in paper documentation. As the clinical behavior underlying this bias appears to be related to the actual instability of the patient, it is unclear whether removing the bias will affect patient outcomes [27].

Despite differences in documented EWS, there was no difference in the timeliness of observations between the two cohorts. This supports the results reported by Hands et al, in which an e-Obs system was introduced to all adult inpatient areas of one hospital [11]. The time of vital sign observations was recorded. They found that observations were more frequently recorded at particular hours in the day, rather than simply responding to escalation algorithms.

There was also no difference in the primary outcome, patient length of stay from admission to the trauma unit until "fit to discharge" after accounting for potential confounding variables. Previous evaluations of e-Obs systems have focused on process improvements such as data accuracy and speed of documentation [7,28-30]. More recently, longitudinal data have been used to assess the impact of e-Obs on patient mortality via before-and-after analysis, with conflicting results [15,16]. Dawes et al found a 2-day reduction in average length of stay after introduction of the VitalPAC e-Obs system when comparing results in 2010 to those in 2005 [15]. However, further exploration suggested this finding was due to a decreased severity of admission rather than a change in in-hospital care. Jones et al reported reductions in length of stay using the Patientrack e-Obs system [13]. Subsequent correspondence suggested that this may have resulted from changes in discharge processes rather than changes in care resulting from the e-Obs system [31,32].

A more recent study suggested substantial reductions in hospital mortality in two hospitals that implemented VitalPAC [16]. Although our per-protocol analysis also showed a survival benefit with e-Obs, this was a post hoc analysis of a secondary outcome and was not found in either intention-to-treat analysis or 30-day mortality (analyzing either intention-to-treat or per-protocol groups). We, therefore, do not think great weight should be attached to this outcome, particularly as we did not find any change in observation frequency in those becoming unstable to support the hypothesis that e-Obs caused a change in care.

The previous before-and-after studies are inherently limited in their ability to account for temporal changes in covariates [33]. Our stepped-wedge design and relatively short total study time reduced the risk of other major changes in practice affecting our outcomes. The stepped-wedge methodology employed here is a practical choice for phased interventions that allows for control of temporal covariates because both control and intervention are active over the whole study period [34]. One disadvantage of the stepped-wedge design is that there is no established consensus on the most appropriate methods of modeling the data [35]. However, if properly analyzed, the quality of evidence is better than that of before-and-after studies and approaches that of randomized controlled trials [36].

Although e-Obs had no effect on patient outcomes in our study, there are some positive findings. First, the frequency of observations before and after e-Obs remained stable. Furthermore, although the timeliness of observations when patients were physiologically unstable did not improve, they also did not worsen, matching previous reports of e-Obs introduction [13]. These findings suggest that it is possible to introduce an e-Obs system without adversely affecting these ward staff functions. Without an adverse effect, the availability of the vital signs electronically brings the possibility of benefits outside the patients studied, or in the future that may make e-Obs worthwhile.

Generalizability

The results here are specific to the VitalPAC e-Obs system. This system contains some key features that are available in alternative e-Obs solutions; these include automatic EWS calculation and real-time ward lists [12]. The results are also location specific. While trauma was chosen as a representative specialty that contained a wide range of care, results may not be true in other hospital contexts. Although these factors may reduce the generalizability of results, the results are robust due to the large number of study participants and the stepped-wedge study design.

Conclusions

The introduction of an electronic system for recording vital sign observations was not associated with reduction in time from admission to the trauma unit until "fit to discharge." More patient episodes contained an EWS≥3 using the e-Obs system, but this was not associated with a change in "escalation time."

Acknowledgments

We thank the Learning Clinic (System C) for provision of the VitalPAC system and full training for staff during the study. We also thank Jane Cottrell for assistance in developing the study protocol and Jane Boutflower and Gloria Noreen for collecting

XSL•FO
RenderX

data during the study. This work was supported by the Oxford National Institute for Health Research Biomedical Research Centre (DFRWAO00).

Authors' Contributions

DCWW assisted in collecting the data, managed and refined the data, produced overview statistics, and prepared the manuscript draft. JK collected the data and assisted in preparing the manuscript. JB developed the statistical methodology, computed the statistical analysis of the data, and assisted in preparing the manuscript. LT and PJW designed the research study, provided overall supervision, and revised the manuscript. All authors have read and approved the manuscript.

Conflicts of Interest

PJW, DCWW, and LT developed an alternative e-Obs system (SEND) after study completion and prior to paper submission. In February 2017, this system was licensed by the University of Oxford and the Oxford University Hospitals National Health Service Trust to Drayson Technologies. Under the license agreement, PJW, DCWW, and LT may receive future royalties. DCWW and LT have undertaken consultancy for Drayson Technologies. JK and JB declare no conflict of interest. The Learning Clinic had no access to the data or influence on the analysis of the study.

Multimedia Appendix 1

Per-protocol analysis.

[PDF File (Adobe PDF File), 296KB - medinform_v6i4e10221_app1.pdf]

Multimedia Appendix 2

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 136KB - medinform_v6i4e10221_app2.pdf]

References

1. Cretikos M, Chen J, Hillman K, Bellomo R, Finfer S, Flabouris A. The objective medical emergency team activation criteria: a case-control study. Resuscitation 2007 Apr;73(1):62-72. [doi: 10.1016/j.resuscitation.2006.08.020] [Medline: 17241732]
2. Watkinson PJ, Barber VS, Price JD, Hann A, Tarassenko L, Young JD. A randomised controlled trial of the effect of continuous electronic physiological monitoring on the adverse event rate in high risk medical and surgical patients. Anaesthesia 2006 Nov;61(11):1031-1039 [FREE Full text] [doi: 10.1111/j.1365-2044.2006.04818.x] [Medline: 17042839]
3. McQuillan P, Pilkington S, Allan A, Taylor B, Short A, Morgan G, et al. Confidential inquiry into quality of care before admission to intensive care. BMJ 1998 Jun 20;316(7148):1853-1858 [FREE Full text] [Medline: 9632403]
4. Garnacho-Montero J, Aldabo-Pallas T, Garnacho-Montero C, Cayuela A, Jiménez R, Barroso S, et al. Timing of adequate antibiotic therapy is a greater determinant of outcome than are TNF and IL-10 polymorphisms in patients with sepsis. Crit Care 2006;10(4):R111 [FREE Full text] [doi: 10.1186/cc4995] [Medline: 16859504]
5. Chalfin DB, Trzeciak S, Likourezos A, Baumann BM, Dellinger RP, DELAY-ED study group. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. Crit Care Med 2007 Jun;35(6):1477-1483. [doi: 10.1097/01.CCM.0000266585.74905.5A] [Medline: 17440421]
6. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted 'track and trigger' systems. Resuscitation 2008 May;77(2):170-179. [doi: 10.1016/j.resuscitation.2007.12.004] [Medline: 18249483]
7. Prytherch DR, Smith GB, Schmidt P, Featherstone PI, Stewart K, Knight D, et al. Calculating early warning scores-a classroom comparison of pen and paper and hand-held computer methods. Resuscitation 2006 Aug;70(2):173-178. [doi: 10.1016/j.resuscitation.2005.12.002] [Medline: 16806641]
8. Smith AF, Oakey RJ. Incidence and significance of errors in a patient 'track and trigger' system during an epidemic of Legionnaires' disease: retrospective casenote analysis. Anaesthesia 2006 Mar;61(3):222-228 [FREE Full text] [doi: 10.1111/j.1365-2044.2005.04513.x] [Medline: 16480345]
9. Subbe CP, Gao H, Harrison DA. Reproducibility of physiological track-and-trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007 Apr;33(4):619-624. [doi: 10.1007/s00134-006-0516-8] [Medline: 17235508]
10. Wilson SJ, Wong D, Clifton D, Fleming S, Way R, Pullinger R, et al. Track and trigger in an emergency department: an observational evaluation study. Emerg Med J 2013 Mar;30(3):186-191. [doi: 10.1136/emermed-2011-200499] [Medline: 22440880]
11. Hands C, Reid E, Meredith P, Smith GB, Prytherch DR, Schmidt PE, et al. Patterns in the recording of vital signs and early warning scores: compliance with a clinical escalation protocol. BMJ Qual Saf 2013 Sep;22(9):719-726. [doi: 10.1136/bmjqs-2013-001954] [Medline: 23603474]

XSL•FO
RenderX

12. Wong D, Bonnici T, Knight J, Morgan L, Coombes P, Watkinson P. SEND: a system for electronic notification and documentation of vital sign observations. BMC Med Inform Decis Mak 2015 Aug 13;15:68 [FREE Full text] [doi: 10.1186/s12911-015-0186-y] [Medline: 26268349]
13. Jones S, Mullally M, Ingleby S, Buist M, Bailey M, Eddleston JM. Bedside electronic capture of clinical observations and automated clinical alerts to improve compliance with an Early Warning Score protocol. Crit Care Resusc 2011 Jun;13(2):83-88. [Medline: 21627575]
14. Smith GB, Prytherch DR, Schmidt P, Featherstone PI, Knight D, Clements G, et al. Hospital-wide physiological surveillance-a new approach to the early identification and management of the sick patient. Resuscitation 2006 Oct;71(1):19-28. [doi: 10.1016/j.resuscitation.2006.03.008] [Medline: 16945465]
15. Dawes TR, Cheek E, Bewick V, Dennis M, Duckitt RW, Walker J, et al. Introduction of an electronic physiological early warning system: effects on mortality and length of stay. Br J Anaesth 2014 Oct;113(4):603-609 [FREE Full text] [doi: 10.1093/bja/aeu107] [Medline: 24878563]
16. Schmidt PE, Meredith P, Prytherch DR, Watson D, Watson V, Killen RM, et al. Impact of introducing an electronic physiological surveillance system on hospital mortality. BMJ Qual Saf 2015 Jan;24(1):10-20. [doi: 10.1136/bmjqs-2014-003073] [Medline: 25249636]
17. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 2015 Feb 06;350:h391 [FREE Full text] [Medline: 25662947]
18. Watkinson P, Tarassenko L, Price J, Webb S. nuffield department of clinical medicine. 2011. OTEST: Oxford Track-&-Trigger Electronic System for Trauma Protocol URL: https://www.ndcn.ox.ac.uk/publications/718248 [accessed 2018-02-21] [WebCite Cache ID 6xOegqqq8]
19. Tarassenko L, Clifton DA, Pinsky MR, Hravnak MT, Woods JR, Watkinson PJ. Centile-based early warning scores derived from statistical distributions of vital signs. Resuscitation 2011 Aug;82(8):1013-1018. [doi: 10.1016/j.resuscitation.2011.03.006] [Medline: 21482011]
20. The Mathworks, I. MATLAB and statistics toolbox release 2014a. Natick, Massachusetts, United States: The Mathworks, Inc; 2014. URL: https://www.mathworks.com/products/matlab.html [accessed 2018-10-02] [WebCite Cache ID 72rzyjw86]
21. Banks J. STATA. The Economic Journal 1992 Nov;102(415):1581. [doi: 10.2307/2234838]
22. Bonnici T, Gerry S, Wong D, Knight J, Watkinson P. Evaluation of the effects of implementing an electronic early warning score system: protocol for a stepped wedge study. BMC Med Inform Decis Mak 2016 Feb 09;16:19 [FREE Full text] [doi: 10.1186/s12911-016-0257-8] [Medline: 26860362]
23. Clarke A. Why are we trying to reduce length of stay? Evaluation of the costs and benefits of reducing time in hospital must start from the objectives that govern change. Qual Health Care 1996 Sep;5(3):172-179 [FREE Full text] [Medline: 10161532]
24. National Institute for Health and Clinical Excellence. Acutely ill patients in hospital : recognition of and response to acute illness in adults in hospital. (NICE clinical guideline No 50.) 2007. [Medline: 21204323]
25. Torsvik M, Gustad LT, Mehl A, Bangstad IL, Vinje LJ, Damås JK, et al. Early identification of sepsis in hospital inpatients by ward nurses increases 30-day survival. Crit Care 2016 Dec 05;20(1):244 [FREE Full text] [doi: 10.1186/s13054-016-1423-1] [Medline: 27492089]
26. Smith MEB, Chiovaro JC, O'Neil M, Kansagara D, Quiñones AR, Freeman M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. Ann Am Thorac Soc 2014 Nov;11(9):1454-1465. [doi: 10.1513/AnnalsATS.201403-102OC] [Medline: 25296111]
27. Clifton DA, Clifton L, Sandu D, Smith GB, Tarassenko L, Vollam SA, et al. 'Errors' and omissions in paper-based early warning scores: the association with changes in vital signs--a database analysis. BMJ Open 2015 Jul 03;5(7):e007376 [FREE Full text] [doi: 10.1136/bmjopen-2014-007376] [Medline: 26141302]
28. Wager KA, Schaffner MJ, Foulois B, Swanson KA, Parker C, Walo H. Comparison of the quality and timeliness of vital signs data using three different data-entry devices. Comput Inform Nurs 2010;28(4):205-212. [doi: 10.1097/NCN.0b013e3181e1df19] [Medline: 20571371]
29. Mohammed M, Hayton R, Clements G, Smith G, Prytherch D. Improving accuracy and efficiency of early warning scores in acute care. Br J Nurs 2009;18(1):18-24. [doi: 10.12968/bjon.2009.18.1.32072] [Medline: 19127227]
30. Wong D, Bonnici T, Knight J, Gerry S, Turton J, Watkinson P. A ward-based time study of paper and electronic documentation for recording vital sign observations. J Am Med Inform Assoc 2017 Feb 11. [doi: 10.1093/jamia/ocw186] [Medline: 28339626]
31. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI, Meredith P. Bedside electronic capture - can it influence length of stay? Crit Care Resusc 2011 Dec;13(4):281-2; author reply 283. [Medline: 22129292]
32. Subbe CP. Bedside electronic capture - methodological concerns. Crit Care Resusc 2011 Dec;13(4):283; author reply 283-283; author reply 284. [Medline: 22129293]
33. Goodacre S. Uncontrolled before-after studies: discouraged by Cochrane and the EMJ. Emerg Med J 2015 Jul;32(7):507-508. [doi: 10.1136/emermed-2015-204761] [Medline: 25820301]
34. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007 Feb;28(2):182-191. [doi: 10.1016/j.cct.2006.05.007] [Medline: 16829207]

XSL•FO
RenderX

35. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials 2015;16:352 [FREE Full text] [doi: 10.1186/s13063-015-0842-7] [Medline: 26279154]
36. Davey C, Hargreaves J, Thompson JA, Copas AJ, Beard E, Lewis JJ, et al. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. Trials 2015;16:358 [FREE Full text] [doi: 10.1186/s13063-015-0838-3] [Medline: 26278667]

Abbreviations

e-Obs: electronic recording of vital sign observations
EWS: Early Warning Score
ICU: intensive care unit

Edited by G Eysenbach; submitted 26.02.18; peer-reviewed by P Scott, A Jayatilleke, P Pasquier; comments to author 02.08.18; revised version received 10.08.18; accepted 10.08.18; published 31.10.18.

Please cite as:
Wong DCW, Knight J, Birks J, Tarassenko L, Watkinson PJ
Impact of Electronic Versus Paper Vital Sign Observations on Length of Stay in Trauma Patients: Stepped-Wedge, Cluster Randomized Controlled Trial
JMIR Med Inform 2018;6(4):e10221
URL: http://medinform.jmir.org/2018/4/e10221/
doi:10.2196/10221
PMID:30381284

©David CW Wong, Julia Knight, Jacqueline Birks, Lionel Tarassenko, Peter J Watkinson. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 31.10.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Processing of Electronic Medical Records for Health Services Research in an Academic Medical Center: Methods and Validation

Nabilah Rahman[1*], BSc; Debby D Wang[1*], BSc, PhD; Sheryl Hui-Xian Ng[1], BSc, MPH; Sravan Ramachandran[1], BTech, MTech; Srinath Sridharan[1], BTech, MTech, PhD; Astrid Khoo[2], BSc, MPH, CIP; Chuen Seng Tan[1,3], BSc, MSc, PhD; Wei-Ping Goh[4], MBBS, MBA; Xin Quan Tan[2,3], MBBS, MPH

[1]Centre for Health Services and Policy Research, Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
[2]Regional Health System Planning Office, National University Health System, Singapore, Singapore
[3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
[4]University Medicine Cluster, National University Hospital, Singapore, Singapore
[*]these authors contributed equally

Corresponding Author:
Xin Quan Tan, MBBS, MPH
Regional Health System Planning Office
National University Health System
NUHS Tower Block, Level 13
1E Kent Ridge Road
Singapore, 119228
Singapore
Phone: 65 97888319
Email: kyletanxq@gmail.com

Abstract

Background: Electronic medical records (EMRs) contain a wealth of information that can support data-driven decision making in health care policy design and service planning. Although research using EMRs has become increasingly prevalent, challenges such as coding inconsistency, data validity, and lack of suitable measures in important domains still hinder the progress.

Objective: The objective of this study was to design a structured way to process records in administrative EMR systems for health services research and assess validity in selected areas.

Methods: On the basis of a local hospital EMR system in Singapore, we developed a structured framework for EMR data processing, including standardization and phenotyping of diagnosis codes, construction of cohort with multilevel views, and generation of variables and proxy measures to supplement primary data. Disease complexity was estimated by Charlson Comorbidity Index (CCI) and Polypharmacy Score (PPS), whereas socioeconomic status (SES) was estimated by housing type. Validity of modified diagnosis codes and derived measures were investigated.

Results: Visit-level (N=7,778,761) and patient-level records (n=549,109) were generated. The International Classification of Diseases, Tenth Revision, Australian Modification (ICD-10-AM) codes were standardized to the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) with a mapping rate of 87.1%. In all, 97.4% of the ICD-9-CM codes were phenotyped successfully using Clinical Classification Software by Agency for Healthcare Research and Quality. Diagnosis codes that underwent modification (truncation or zero addition) in standardization and phenotyping procedures had the modification validated by physicians, with validity rates of more than 90%. Disease complexity measures (CCI and PPS) and SES were found to be valid and robust after a correlation analysis and a multivariate regression analysis. CCI and PPS were correlated with each other and positively correlated with health care utilization measures. Larger housing type was associated with lower government subsidies received, suggesting association with higher SES. Profile of constructed cohorts showed differences in disease prevalence, disease complexity, and health care utilization in those aged above 65 years and those aged 65 years or younger.

Conclusions: The framework proposed in this study would be useful for other researchers working with EMR data for health services research. Further analyses would be needed to better understand differences observed in the cohorts.

(JMIR Med Inform 2018;6(4):e10933) doi:10.2196/10933

XSL•FO
RenderX

KEYWORDS

health services; electronic medical records; data curation; validation studies

Introduction

Secondary use of electronic medical records (EMRs) data by clinicians, researchers, data analysts, and computer scientists has led to promising findings in population health research such as patient-utilization stratification [1], treatment-effectiveness evaluation [2], early detection of diseases [3], and predictive modeling [4]. However, dealing with EMR data is often labor intensive [5] and challenging because of the lack of standardization in data entry, changes in coding procedures over time, and the impact of missing information [6,7]. Processing EMR data for analysis is a critical step in health services research requiring significant time and effort.

Different research teams have described EMR data processing methods [6,8-18]. However, most have focused only on partial aspects of data processing [11-13,15-18] or processing related to a specific disease [6,11,13]. Designing an efficient and structured way to standardize records, process features, link data, and select cohorts for analysis is urgently needed, given the increasing emphasis on big data and analytics to improve patient care and reduce health care expenditure [5,19].

Although the standardization of diagnosis codes of different nosologies or different versions of the same nosology has been reported previously [20,21], the completeness and validity of such mapping is rarely reported. This lack of transparent sharing of code set definitions, construction process, and validity is a barrier to rapid scaling of health services research [22], given its importance and widespread relevance. With the change in coding procedures over time, standardization is hence necessary for longitudinal analyses and cross-period comparisons.

Measures of patient complexity, disease severity, and socioeconomic status (SES) are not readily available in most datasets [23] but have been shown to be useful in population health [24-27] and disease progression studies [28]. Although some studies have used the Charlson Comorbidity Index (CCI) [26,29-31] and drug burden [32,33] to estimate patient complexity, validity of these measures as an estimate for patient complexity has rarely been established in Asia. In the absence of income data, SES is typically derived from area-based income level from census data [34,35], insurance status [36], or property value [37,38]. However, these proxies require additional data as well, which are often not readily available in health care administrative datasets or EMRs.

This study has attempted to address some of these challenges common to the use of EMR data for health services research by detailing a structured framework for EMR data processing. Furthermore, the study proposed and validated methods for standardization of diagnosis codes and construction of disease phenotypes and also proposed and tested derived measures of disease complexity and SES, which could be applicable to other datasets with similar data fields.

Methods

Local Electronic Medical Records System and Architecture

The National University Hospital (NUH) is a 1000-bed Academic Medical Center (AMC) in Singapore [39]. Being 1 of only 2 AMCs in Singapore, its EMR offers an important view of the local patient population, particularly those who have sought care in a tertiary setting. The Patient Affordability Simulation System (PASS) dataset, which this study is premised on, originated from the NUH's EMR system [40,41]. Specifically, PASS captures information of all patients who visited NUH since 2004, and for this work, we examined data from 2005 to 2013. PASS information is organized in 6 tables: (1) demographic, (2) movement, (3) billing, (4) pharmacy, (5) diagnosis, and (6) diagnosis-related group (DRG) as depicted in Figure 1.

The cascade architecture of PASS is patient → visit → record as shown in Figure 2 where record is the basic row element for (2) to (5) before aggregation. Five PASS tables were used in our study. The DRG table was not used, as the information captured is a subset of the more comprehensive International Classification of Diseases (ICD) codes found in the diagnosis table.

Patient ID is common in each table and Visit ID is available across (2) to (5). These IDs were used to link features across tables.

Standardizing and Phenotyping of Diagnosis Codes With Quality Validation

The National University Hospital EMR system adopted International Classification of Diseases, Ninth Revision (ICD-9), Clinical Modification (CM) codes before 2010 and then migrated to the more updated the International Classification of Diseases, Tenth Revision (ICD-10), Australian Modification (AM) codes afterward. To standardize the ICD codes, we transformed ICD-10-AM codes to ICD-9-CM format using Australia Consortium for Classification Development (ACCD) backward mapping tables [42]. ICD-10 is more precise than ICD-9 (ie, there could be multiple ICD-10 codes for each ICD-9, providing greater granularity such as distinguishing the site [left vs right] of pathology). Due to the added granularity, majority of ICD-10 codes cannot be represented by forward mapping of ICD-9 codes [43]. As forward mapping from ICD-9 to ICD-10 and backward mapping from ICD-10 to ICD-9 differ in terms of scope and coverage, both approaches run the risk of ambiguous mappings and loss of information [44,45]. ICD-10 codes also form a significantly smaller portion of diagnosis codes in our database. In this regard, backward mapping of ICD-10 codes to ICD-9 would minimize the impact of above-mentioned risks.

XSL•FO
RenderX

Figure 1. Components of PASS (Patient Affordability Simulation System) database before aggregation, which consist of demographic table (each row is a patient), movement table (each row is a record), billing table (each row is a record or transaction), pharmacy table (each row is a record or transaction), diagnosis table (each row is a record), and diagnosis-related group (DRG) table (not used).

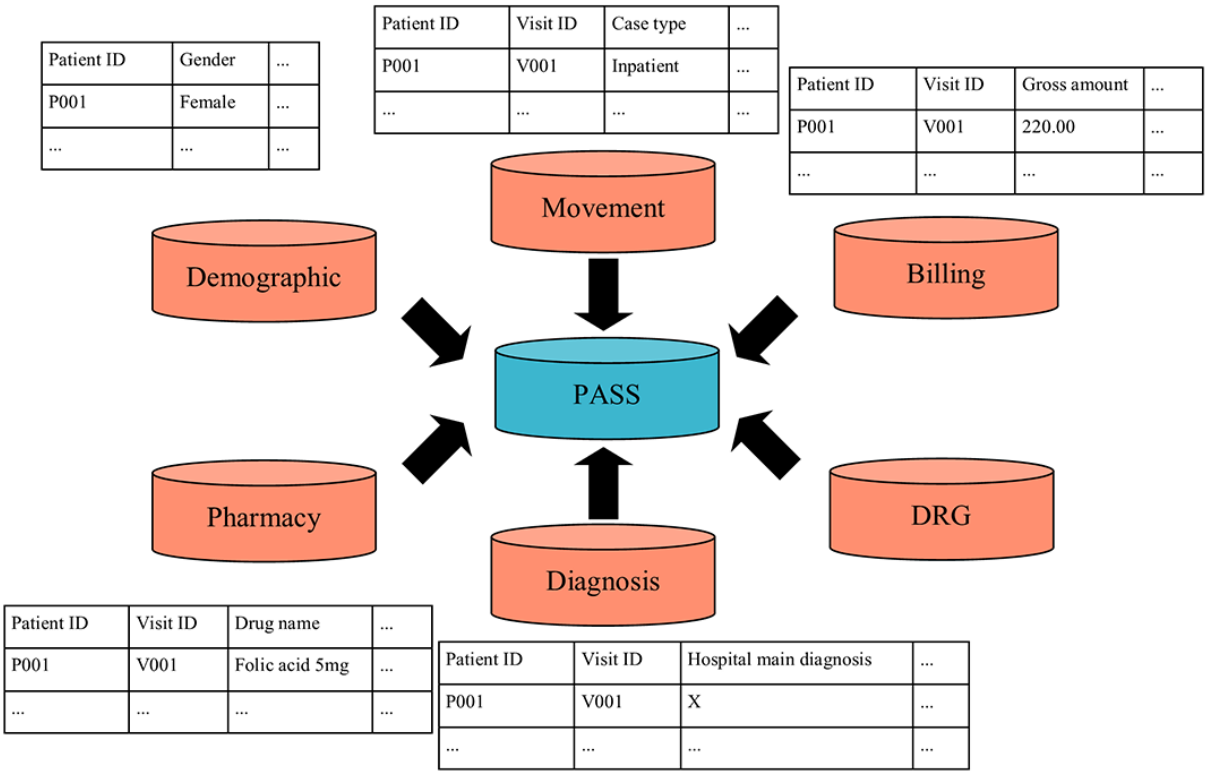
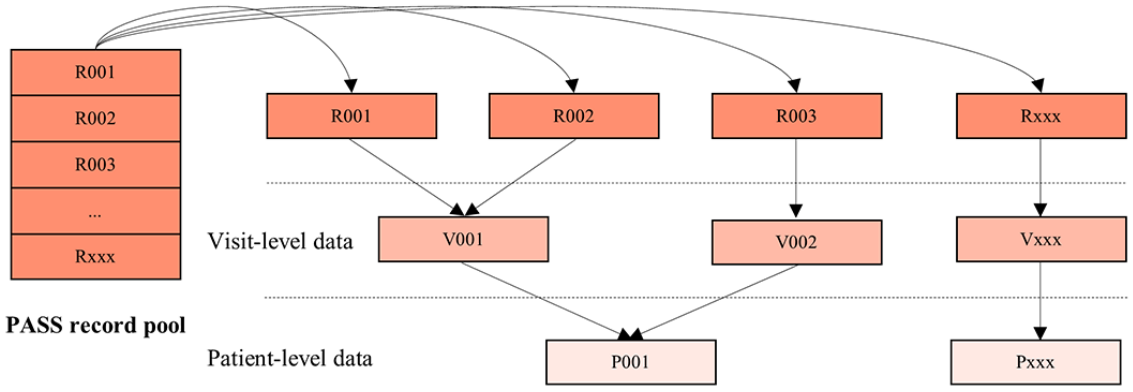


Figure 2. Flow of aggregation from records (before aggregation) to visits and then to patients after aggregations. PASS: Patient Affordability Simulation System.



In transformation to the ICD-9-CM, if an ICD-10-AM code could not be directly mapped to ICD-9-CM using the ACCD backward mapping tables, it will undergo truncation (ie, truncation down to 3 heading digits) or zero addition (ie, addition of up to 2 trailing *0* digits to the ICD code). Mapping will be performed again thereafter. ICD-10-AM codes that were left unmapped after code modification were excluded from further analyses. A diagnosis will be classified as primary diagnosis (PD) if it is indicated as the hospital's main diagnosis (may be referred as principal diagnosis in other systems [46]). Otherwise, it will be classified as a secondary diagnosis (SD). All PD and SD codes were standardized to ICD-9-CM format. Figure 3 describes the code standardization approach in detail.

As the ACCD backward mapping table is well established and defined, we regard the mapping from original ICD-10-AM codes (no truncations or zero additions) to ICD-9-CM codes as valid [42]. Therefore, to determine the quality of mapping, only those with truncations and zero additions during mapping were examined. We sampled 151 unique ICD-10-AM codes that underwent truncation or zero addition (modified) during the conversion. These 151 codes comprised 23.1% of total 653

XSL•FO
RenderX

unique ICD-10-AM codes that were modified. Thereafter, 2 physicians independently reviewed and rated the validity of the mapping from ICD-10-AM to ICD-9-CM for these sampled codes. List of disagreements in terms of validity of the mapping was generated at the end of the rating exercise and shared between the 2 physicians to reconcile differences through discussions. In the event where disagreement could not be resolved, a third physician would then be brought in. In our study, the 2 physicians managed to reconcile differences without the involvement of the third physician. The ratio of valid mappings after reconciling rating differences by the 2 physicians was then calculated to validate our code standardization approach. Similar method of validating diagnosis codes has been documented in other studies [47-49].

ICD codes have good utility for clinical research where the researcher needs the granularity for identification and attribution of pathology at an individual level [20,50]. However, for health services research, broader classification and coding methods such as the Clinical Classification Software (CCS) by Agency for Healthcare Research and Quality [51] demonstrate utility as there is sufficient granularity at a population level, yet reduced sparsity [52-54].

To extend the utility of our dataset to support health services research, we sought ways to phenotype the more than 10,000 ICD-9-CM codes (both PD and SD) into meaningful groups. To that end, we grouped the ICD-9-CM codes (including those converted from ICD-10-AM codes) using CCS to 283 mutually exclusive disease categories (eg, *essential hypertension* and *cancer of breast*). For ICD codes that could not be classified directly using CCS, the approach outlined in Figure 3 was adopted as well. Validation of the ICD-9-CM codes that underwent truncation or zero addition in the phenotyping was conducted using the same methodology as described above. In total, 361 (20.7%) unique ICD-9-CM codes of the total 1747 unique ICD-9-CM codes that were modified during the phenotyping were sampled for this purpose.

Cohort Generation and Feature Processing

Generating Visit and Patient-Level Records

The PASS EMR had captured data at the record level. For meaningful analysis to be performed, the database had to be processed to generate visit-level and patient-level records. Visit-level records capture information related to a single encounter with NUH. Patient-level records capture information on the patient himself as well as information related to the visits accumulated over the study period.

The 2 types of unique identifiers used for record linkage are Patient ID and Visit ID. To generate visit-level records, we used Visit ID to aggregate records within each table (eg, all bills for a visit) and then fully join the data for each visit by drawing on data across tables (ie, linking the movement, billing, pharmacy, and diagnosis information to provide more complete utilization and clinical details for each visit). Age and date exclusion criteria were applied to 3 of the tables (2-4) before the join. The diagnosis table was then also filtered using Visit ID from the other tables (2-4) to filter out diagnoses not related to visits within our cohort after applying the earlier exclusion criteria. The tables (2-5) were fully joined thereafter. The joined data were further linked to demographics through Patient ID. Patient-level records were then generated by aggregating visits by Patient ID. Patient-related exclusion criteria were then applied after obtaining patient-level records.

Aggregation and analysis were undertaken using *R* version 3.2.0 [55]. *R* package *multidplyr* [56] was used for efficient parallel aggregation.

Figure 3. Pseudocode for converting International Classification of Diseases, Tenth Revision, Australian Modification (ICD-10-AM) codes to International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes.

Input: All ICD-10-AM codes Ω from diagnosis table, backward mapping table $M=(\Omega^{ICD-10} \mapsto \Omega^{ICD-9})$ for ICD-10-AM to ICD-9-CM
Output: Converted ICD-9-CM codes $\tilde{\Omega}$
Procedure
1: Remove invalid codes $\hat{\Omega}$ (exceeding five characters):
$\Omega \leftarrow \Omega - \hat{\Omega}$
Initialization of outputs and quality indicators: $\tilde{\Omega} \leftarrow \emptyset$
2: Loop for ω in Ω
If ω is found in M, then pick out the ICD-9 version code $\tilde{\omega}$
Else if the heading 4 characters of ω is found in M, then pick out the ICD-9 code $\tilde{\omega}$
Else if the heading 3 characters of ω is found in M, then pick out the ICD-9 code $\tilde{\omega}$
Else if ω with a trailing *0* digit added is found in M, then pick out the ICD-9 code $\tilde{\omega}$
Else if ω with two trailing *0* digits added is found in M, then pick out the ICD-9 code $\tilde{\omega}$
Else if ω is a incorrectly-labeled ICD-9 code, then $\tilde{\omega} \leftarrow \omega$
Else $\tilde{\omega} \leftarrow NA$
$\tilde{\Omega} \leftarrow \tilde{\Omega} \cup \tilde{\omega}$ and
3: Return $\tilde{\Omega}$

XSL•FO
RenderX

Exclusion Criteria

The following exclusion criteria were applied to streamline the data for subsequent analysis:

- Any inpatient visit with admission date before 2005 or discharge date after 2013 was dropped. This ensured that the entire period of each inpatient visit was captured.
- Visits when patients were aged less than 21 years were excluded in this study as subsequent analysis is focused on adult patients.
- Patients with no PD were excluded.
- Patients with birth years 1900 or earlier were excluded (patients without birth date information were assigned a default 1900 as the birth year; hence, they were excluded from the study).
- Patients with invalid diagnoses (eg, male patients with diagnoses of pregnancies and female infertility) were removed.

The final cohort analyzed was an adult cohort aged 21 years and above, with valid age and at least one PD record.

Preparing Primary and Secondary Variables

The main source variables in the database had to be extracted and processed to generate secondary variables useful for cohort profiling and other health services research. In addition, we attempted to generate proxies for clinical and socioeconomic indicators unavailable in the dataset, namely, disease complexity and SES. Summary and details of all the extracted variables can be found in the Multimedia Appendices 1-3.

On the basis of the source variables from the 5 PASS tables, we generated a series of secondary variables falling in categories of (1) demographics (including SES), (2) health care utilization, (3) disease indicators, and (4) disease complexity. For categorical source variables, we created dummy variables for visits, such as whether a visit is an emergency department (ED) visit or whether it has a specific CCS disease, and then we aggregated them to patient-level by adding new categories, summation, or logic operation. For numerical source variables, such as inpatient length-of-stay (LOS), hospital charges, and the components, a simple summation over all visits led to the features at patient-level. Hospital charges (full cost of care before subsidy) was adjusted using Monetary Authority of Singapore Web-based inflation calculator [57] for health goods and services to 2015 levels before the aggregation to patient-level.

Estimating Disease Complexity and Validation of Measures

As clinical indicators and investigation results [23] that provide information on disease severity and patient complexity were not available in the dataset, we introduced 2 measures to estimate disease complexity—CCI [58] and Polypharmacy Score (PPS) [59]. CCI and PPS have both been shown to be good measures of patient comorbidity and complexity in many studies [26,29-33].

In our dataset, the Charlson comorbidities were identified using ICD-9-CM codes [58], and both PD and SD codes were considered for each patient. *R* package *icd* [59] was used to calculate CCI [28]. PPS quantifies drug burden, and high drug burden is usually reflective of more severe disease or greater comorbidity [60]. PPS at the visit-level was defined as the number of unique drugs dispensed in a visit, and PPS at patient-level was defined as the maximum PPS value at visit-level for that patient across all visits. When computing the PPS, nonprescription drugs and devices were excluded.

Validity of CCI and PPS were assessed to ensure that these measures were consistent with theoretical understanding and literature. To assess convergent validity, Spearman rank correlation between PPS and CCI was computed. This measures the degree to which PPS and CCI, that should be measuring disease complexity, are in fact related. To assess criterion validity, Spearman rank correlations between health care utilization (number of inpatient, specialist outpatient clinic [SOC] and ED visits) and PPS and CCI were computed. This measures the extent to which higher CCI and PPS is associated with higher health care utilization, under the assumption that clinically complex patients require more health care utilization [61]. The 95% CIs of the correlations were adjusted for multiple comparisons using Holm method. The health care utilization measures were also regressed on CCI and PPS separately, controlling for demographic variables and observed period to further ascertain its criterion validities. Log-linked negative binomial generalized linear models were used to perform the regression analyses. Missing values are removed pair-wise for the regression analyses in this study. Our methods to assess validity of these proxy measures are similar to methods used in numerous other studies [62-64].

Estimating Socioeconomic Status and Validation of Measure

To estimate the SES of PASS patients, we used housing type as a proxy, given the lack of a direct indicator of SES in the dataset. We then validated the use of housing type as a proxy for SES as part of the study.

Each residential block and house in Singapore has a postal code assigned. Using the postal code data of each patient, we were able to determine the block and, consequently, housing type for each patient. The latest postal codes captured in PASS EMR were used, as patients' past addresses were not available. For all Housing Development Board (HDB) blocks (public housing), we obtained information of flat types by postal codes collected using OneMap Singapore [65] from the official HDB website [66]. The full HDB flat type list includes rental flats, studios, 1- to 5-room flats, and other executive flats. We then grouped the flat types by size as follows: rental to 2-room, 3-room, 4-room, and 5-room to executive flats. If a housing block comprised multiple flat types, it was assigned to the flat type with the largest proportion in that block. Residents living in private condominiums or landed properties were classified as private housing and were identified based on a postal code list of private housing provided by a collaborative research team. Blocks with postal codes not belonging to either lists were defined as nonresidential. Patients with postal codes of nonresidential buildings or with no valid postal codes were assigned with a missing value.

XSL•FO
RenderX

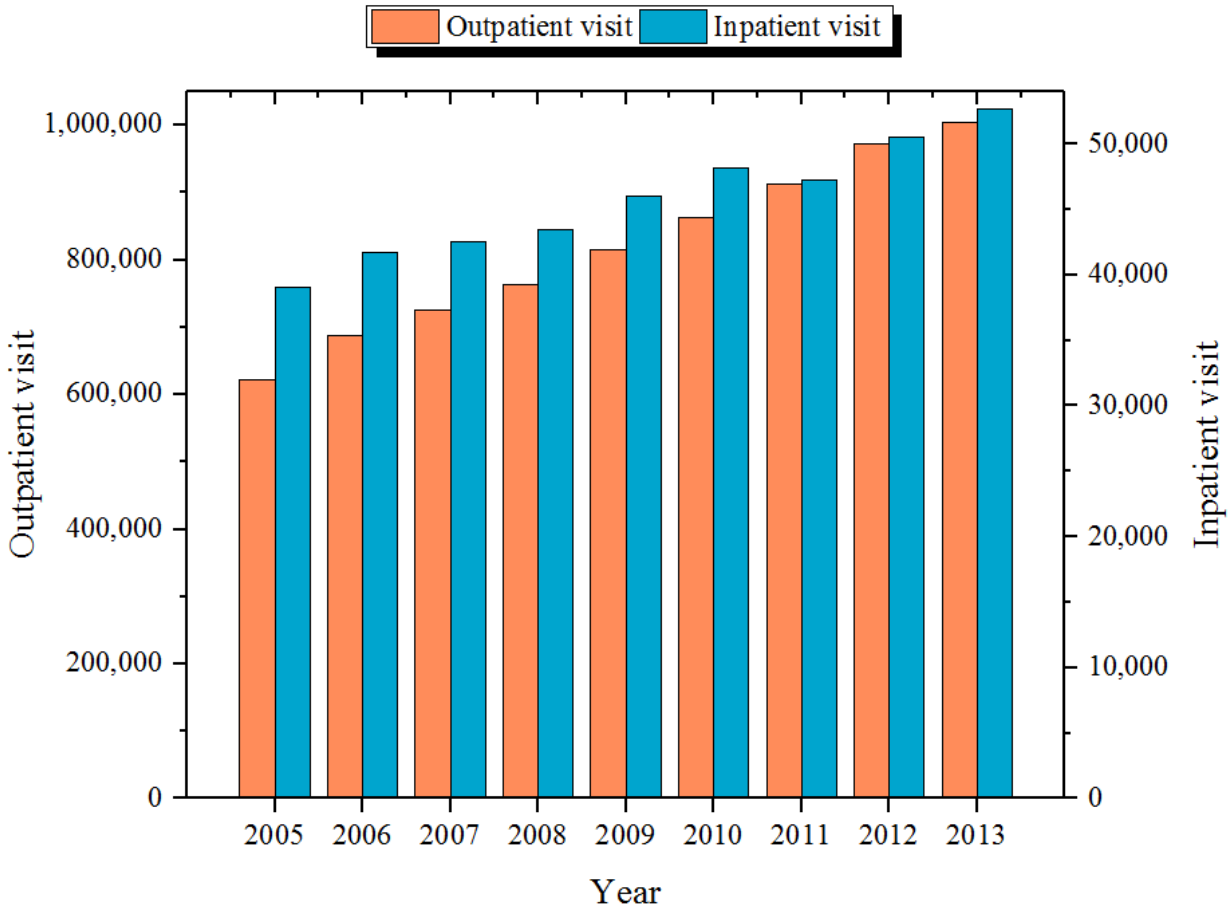
Criterion validity of housing type as a proxy for SES was assessed through studying the relationship between housing type and 2 measures: (1) subsidy status and (2) relative subsidy received (RSR). Subsidy status indicates whether a patient received government subsidized care or nonsubsidized (ie, private) care where nonsubsidized care is costlier and involves higher out-of-pocket payments. Typically, the lower the income level of an individual, the more likely one is to opt for subsidized care given the lower cost [67]. RSR indicates the proportion of the cumulative hospital charges that were paid for with government subsidies. In Singapore, the amount of subsidy one is eligible for and receives is dependent on the income level of the individual [68]. The lower the income level, the more subsidies one is eligible for and a higher percentage of bill will be subsidized. Both subsidy status and RSR were used to validate our SES proxy using Pearson chi-square (χ^2) and Kruskal-Wallis rank-sum test [69,70], respectively, assuming that lower income groups are more likely to opt for subsidized care and that RSR increases with decreasing income. Subsidy status and RSR were also regressed on SES to further ascertain its criterion validity while controlling for nationality (Singaporean vs non-Singaporean). Multinomial logistic and linear regression models were used to perform the regression analyses.

Results

Overview of Electronic Medical Records Aggregation in Patient Affordability Simulation System

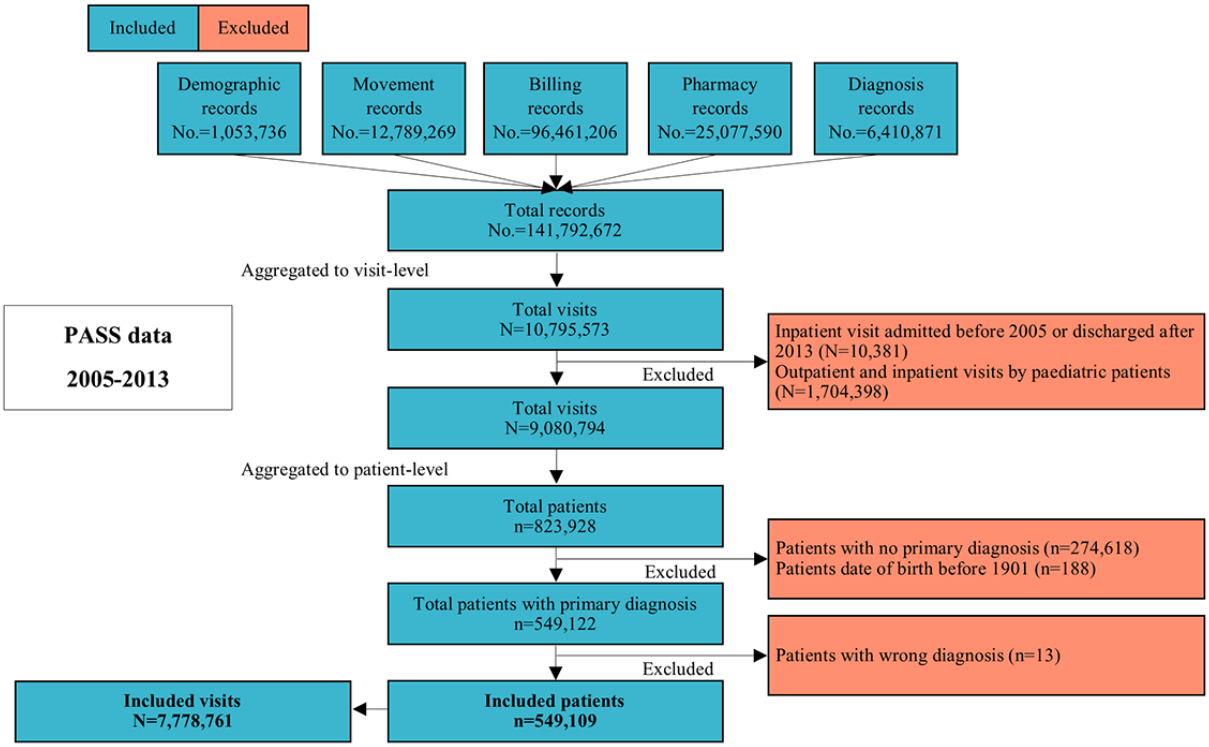
Among 10,795,573 visits during the study period, 7,778,761 satisfied our inclusion criteria and constitute our visit-level data. The visit-level data comprised 7,367,495 outpatient visits and 411,266 inpatient visits. An increasing trend was observed in the number of visits from 2005 to 2013 (Figure 4). The visit-level data were subsequently aggregated to the patient-level data, resulting in a cohort of 549,109 adult patients. The flowchart of EMR processing and cohort generation is depicted in Figure 5.

Figure 4. Annual frequency of outpatient and inpatient visits in the cohort.



XSL•FO
RenderX

Figure 5. Records, visits, and patients in Patient Affordability Simulation System (PASS) Electronic Medical Records (EMR) aggregation. No.: number of records; N: number of visits; n: number of patients.



Mapping Rates and Validation for International Classification of Diseases-10-Australian Modification to International Classification of Diseases-9-Clinical Modification Conversion

There was a total of 4,842,705 diagnoses belonging to our patient cohort after visit-level aggregation, of which 19.2% was coded in ICD-10-AM, with the remainder in ICD-9-CM. The ICD-10-AM codes in our cohort were standardized to ICD-9-CM codes with a mapping rate of 90.3% for PD codes, 78.2% for SD codes, and 81.4% overall using ACCD backward mapping tables. This resulted in 4,670,111 ICD-9-CM codes in the cohort, with 16.2% converted directly from ICD-10-AM by the ACCD backward mapping table. As mentioned in the Methods section, the ACCD backward mapping table has been validated previously; hence, the team regarded these 16.2% of codes that were mapped directly through ACCD as valid. Detailed statistics for code mapping rates are presented in Table 1.

In addition, there were 172,594 codes that could not be mapped through ACCD. Of these, 23,800 (13.79%) ICD-10-AM codes were converted after truncation and 29,005 (16.80%) converted after zero addition (Table 2). These 52,805 ICD-10-AM codes that underwent code modification translated to 653 unique ICD-10-AM codes or 8.9% of the 7373 unique codes that were converted in total. The 52,805 codes accounted for only 6.5% of the 810,459 ICD-10-AM codes that were converted. Validation on a sample of the 653 codes that underwent code modification as part of mapping process was performed. Out of the 151 sampled unique codes, 137 (90.7%) were rated to have valid mappings by the physicians (Table 3).

In total 810,459 (87.1%) of the total ICD-10-AM codes were successfully converted to ICD-9-CM codes (97.2% of PD and 83.5% of SD). These converted codes and the original ICD-9-CM codes form a pool of 4,722,916 (4,722,916/4,842,705, 97.5%) ICD-9-CM codes in our cohort. The unmapped codes, which consisted of 119,789 (12.9%) of the total ICD-10-AM codes, or 471 unique codes were excluded.

XSL•FO
RenderX

Table 1. International Classification of Diseases (ICD) and Clinical Classification Software (CCS) codes mapping rates.

Diagnosis	Primary	Secondary	Primary and secondary
Total diagnosis codes[a], n (% of total codes)			
Total	1,718,049 (100.00)	3,124,656 (100.00)	4,842,705 (100.00)
ICD[b]-9-CM[c]	1,470,473 (85.59)	2,441,984 (78.15)	3,912,457 (80.79)
ICD-10-AM[d]	247,576 (14.41)	682,672 (21.85)	930,248 (19.21)
Total ICD-9-CM codes after ACCD[e] backward mapping, n (%)	1,693,940 (98.60)	2,976,171 (95.25)	4,670,111 (96.44)
Total ICD-9-CM codes after ACCD backward mapping and code modification, n (%)	1,711,180 (99.60)	3,011,736 (96.39)	4,722,916 (97.53)
Total CCS[f] codes after phenotyping, n (% of total ICD-9-CM codes after conversion)	1,402,931 (81.99)	2,775,931 (92.17)	4,178,862 (88.48)
Total CCS codes after phenotyping and code modification, n (% of total ICD-9-CM codes after conversion)	1,696,963 (99.17)	2,901,525 (96.34)	4,598,488 (97.37)

[a]Total number of diagnosis codes from cohort (nonunique codes).

[b]ICD: International Classification of Diseases.

[c]CM: clinical modification.

[d]AM: Australian modification.

[e]ACCD: Australian Consortium for Classification Development.

[f]CCS: Clinical Classification Software.

Table 2. Proportion of International Classification of Diseases (ICD) codes that underwent truncation or zero addition.

Diagnosis	No modification, n (% of total mapped)	Modified		Total mapped, n (% of total codes)	Total codes, n (% of total codes)
		Truncated, n (% of total mapped)	Zero added, n (% of total mapped)		
ICD[a]-10-AM[b] codes converted to ICD-9-CM[c] (unique codes)	6720 (91.14)	195 (2.64)	458 (6.21)	7373 (94.00)	7844 (100.00)
ICD-10-AM codes converted to ICD-9-CM	757,654 (93.48)	23,800 (2.94)	29,005 (3.58)	810,459 (87.12)	930,248 (100.00)
ICD-9-CM[d] converted to CCS[e] codes (unique codes)	9220 (84.07)	246 (2.24)	1501 (13.69)	10,967 (88.26)	12,426 (100.00)
ICD-9-CM[d] converted to CCS Codes	4,178,862 (90.87)	27,240 (0.59)	392,386 (8.53)	4,598,488 (97.37)	4,722,916 (100.00)

[a]ICD: International Classification of Diseases.

[b]AM: Australian modification.

[c]CM: clinical modification.

[d]After conversion from ICD-10-AM to ICD-9-CM using Australian Consortium for Classification Development (ACCD) backward mapping tables and code modification.

[e]CCS: Clinical Classification Software.

Table 3. Validity rate of International Classification of Diseases (ICD) codes, which underwent truncation or zero addition during standardization and phenotyping.

Diagnosis	Valid, n (%)	Invalid, n (%)	Total sample, n (%)
ICD-9-CM[a] codes from modified ICD-10-AM[b] codes	137 (90.7)	14 (9.3)	151 (100.0)
CCS[c] codes from modified ICD-9-CM	332 (92.0)	29 (8.0)	361 (100.0)

[a]ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification.

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

[c]CCS: Clinical Classification Software.

XSL•FO
RenderX

Mapping Rates and Validation for Phenotyping of International Classification of Diseases-9-Clinical Modification Codes

The ICD-9-CM codes were then phenotyped to CCS codes, which resulted in 282 mutually exclusive groups. Out of the 4,722,916 ICD-9-CM codes, 4,178,862 (4,178,862/4,722,916, 88.48%) were converted to CCS codes directly through the CCS. These 4,178,862 (88.48%) are regarded to be valid conversions, given the previous validation done on the CCS [71]. Detailed statistics for code-mapping rates are presented in Table 1.

In addition, 27,240 (27,240/4,722,916, 0.58%) ICD-9-CM codes were converted after truncation and 392,386 (392,386/4,722,916, 8.31%) converted after zero addition (Table 2) through our proposed methodology. These 419,626 ICD-9-CM codes that underwent code modification translated to 1747 unique codes or 15.9% of 10,967 unique codes that were collapsed to CCS codes. Moreover, 332 (332/361, 92.0%) of the 361 sampled unique codes were rated as valid mappings by the physicians (Table 3).

In total, 4,598,488 (97.4%) of the ICD-9-CM codes in our cohort were successfully converted to CCS codes (99.2% of PD and 96.3% of SD; Table 1). The 419,626 codes that underwent code modification accounted for only 9.1% of the 4,598,488 ICD-9-CM codes that were collapsed. The unmapped codes, which consisted of 124,428 (124,428/4,722,916, 2.63%) of the total valid ICD-9-CM codes, or 1459 unique codes were excluded.

Validation of Proxy Measures

CCI was found to be positively correlated with health care utilization measures, including number of inpatient visits (ρ=.54; CI 0.54-0.54; *P*<.001), number of SOC visits (ρ=.30; CI 0.29-0.30; *P*<.001), and number of ED visits (ρ=.21; CI 0.21-0.21; *P*<.001); PPS was found to have an even stronger correlation with health care utilization measures, with exception being number of ED visits: number of inpatient visits (ρ=.74; CI 0.74-0.74; *P*<.001), number of SOC visits (ρ=.53; CI 0.53-0.54; *P*<.001), and number of ED visits (ρ=.19; CI 0.19-0.19; *P*<.001). CCI and PPS were also found to be positively correlated (ρ=.47; CI 0.46-0.47; *P*<.001). On the basis of multivariate regression analysis, which adjusted for gender, race, age, and observed period, health care utilization was expected to increase when there was a unit increase in CCI (Table 4). Number of inpatient visits, SOC visits, and ED visits were expected to change by a factor of 1.46 (*P*<.001), 1.32 (*P*<.001), and 1.23 (*P*<.001), respectively. Health care utilization was also expected to increase when there was a unit increase in PPS; the number of inpatient visits, SOC visits, and ED visits were expected to change by a factor of 1.10 (*P*<.001), 1.08 (*P*<.001), and 1.03 (*P*<.001), respectively.

For all the patients with valid housing type data, the proportions by subsidy status categories within each housing type are presented in Figure 6. As housing size decreased, an increase in proportion of subsidized patients was observed—only 43.8% of patients staying in private housing were subsidized compared with 84.9% of patients staying in 2-room or smaller HDB flats. The Pearson chi-square test showed that subsidy status was not independent of housing type (χ^2_8=23602, *P*<.001), further confirming the observation. The median and mean RSR of patients by housing type were plotted in Figure 7. Patients who lived in larger housing types tended to have a lower percentage of their bill subsidized (eg, those in 2-room or smaller HDB flats had a median RSR of 57.0% compared with those in private housing with a median RSR of 33.9%). Statistically significant differences in median RSR were observed using Kruskal-Wallis rank-sum test (χ^2_4=245232, *P*<.001). The mixed group is a composite group and, hence, it was difficult to interpret the results for this group. On the basis of multivariate regression analysis, which adjusted for nationality, the odds of receiving subsidized care only rather than nonsubsidized care only were higher in patients who lived in smaller housing types when compared with patients who lived in private housing (Table 4). Patients who stayed in 2-room or smaller flats had the highest odds ratio (OR) of 14.43 (*P*<.001), and those who stayed in 5-room flats or executive housing had the lowest OR of 2.97 (*P*<.001). A relatively smaller effect size, but the same trend, was observed when mixed group was compared with nonsubsidized group. Patients who stayed in 2-room flats or smaller and 5-room flats or executive housing had the highest and lowest ORs of receiving both subsidized and nonsubsidized (mixed) care rather than only nonsubsidized care, respectively. The ORs were 3.29 (*P*<.001) and 1.65 (*P*<.001), respectively. RSR was also expected to be higher for patients who stayed in smaller housing after adjusting for nationality (Table 4). Patients who stayed in 2-room or smaller flats were expected to receive 19.0% (*P*<.001) more relative subsidy than those who stayed in private housing, and patients who stayed in 5-room flats or executive housing were expected to receive 9.8% (*P*<.001) more relative subsidy than those who stayed in private housing.

Profile of Cohort

The detailed demographic, medical, and utilization characteristics of the cohort are shown in Table 5. Overall, most of the 549,109 patients were male, Chinese, aged 30 to 39 years, and lived in a 4-room HDB flat. Of the total patient cohort, 62.0% received only subsidized care in NUH. The total inflated-adjusted hospital charges incurred by the cohort during the 9 years were more than SG $5 billion.

The patients older than 65 years had a greater prevalence of chronic diseases and disease complexity scores as compared with those younger than or at 65 years. They also had almost 7 times the median hospital charges, median LOS that was 3 days longer, and 3 times the median SOC visits during the study period compared with those younger than or at 65 years.

XSL•FO
RenderX

Table 4. Multivariate log-linked negative binomial regression on health care utilization, multinomial logistic regression on subsidy status, and linear regression on relative subsidy received (RSR).

Multivariate regression model[a]	Effect (95% CI)	*P* value
Number of inpatient visits between 2005-2013[b]		
CCI[c]	1.47 (1.46-1.47)	<.001
PPS[d]	1.10 (1.10-1.10)	<.001
Number of SOC[e] visits between 2005-2013[b]		
CCI	1.32 (1.31-1.32)	<.001
PPS	1.08 (1.08-1.08)	<.001
Number of ED[f] visits between 2005-2013[b]		
CCI	1.23 (1.23-1.24)	<.001
PPS	1.03 (1.03-1.03)	<.001
Subsidy status between 2005-2013[g]		
Rental, studios, 1-room, and 2-room vs private		
Subsidized vs nonsubsidized	14.43 (12.73-16.36)	<.001
Mixed vs nonsubsidized	3.29 (2.89-3.76)	<.001
3-room vs private		
Subsidized vs nonsubsidized	4.98 (4.81-5.17)	<.001
Mixed vs nonsubsidized	1.99 (1.92-2.07)	<.001
4-room vs private		
Subsidized vs nonsubsidized	4.14 (4.01-4.27)	<.001
Mixed vs nonsubsidized	1.79 (1.74-1.85)	<.001
5-room and executive vs private		
Subsidized vs nonsubsidized	2.97 (2.88-3.07)	<.001
Mixed vs nonsubsidized	1.65 (1.60-1.71)	<.001
Relative subsidy received between 2005-2013[h]		
Rental, studios, 1-room, and 2-room vs private	18.95 (18.57-19.33)	<.001
3-room vs private	14.43 (14.21-14.64)	<.001
4-room vs private	12.72 (12.52-12.92)	<.001
5-room and executive vs private	9.79 (9.59-10.00)	<.001

[a]Eight different models in total.

[b]Effects are exp(β), which can also be interpreted as multiplicative effect.

[c]CCI: Charlson Comorbidity Index.

[d]PPS: Polypharmacy Score.

[e]SOC: specialist outpatient clinic.

[f]ED: emergency department.

[g]Effects are exp(β), which can also be interpreted as odds ratio.

[h]Effects are β.

XSL•FO
RenderX

Figure 6. Proportion of subsidy status categories within each housing type.

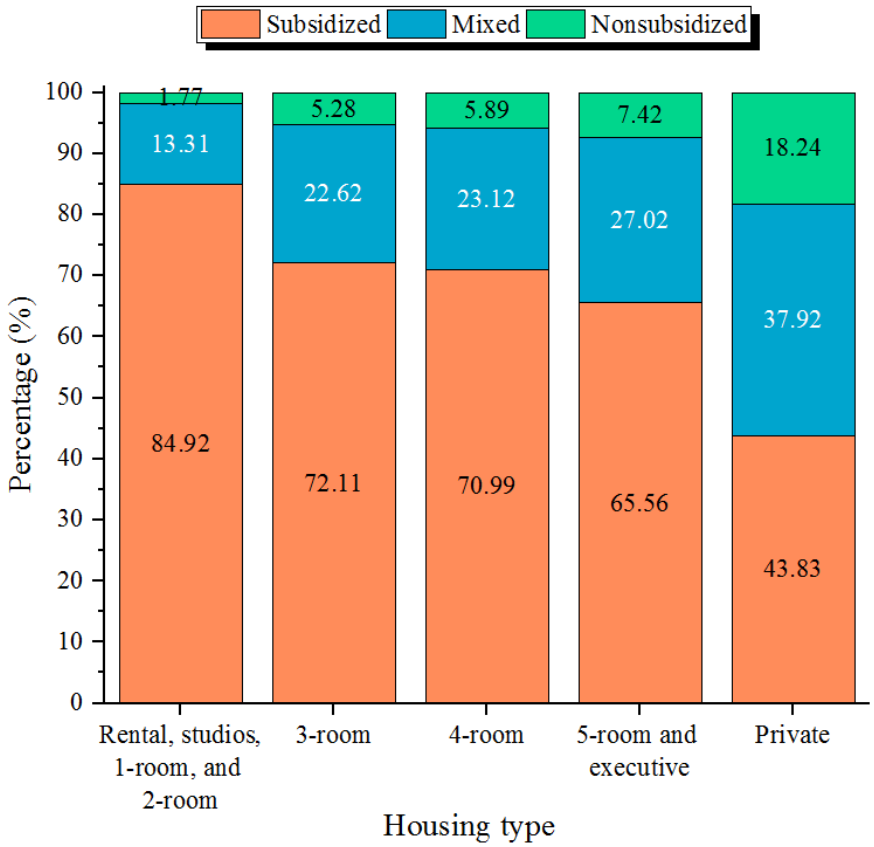
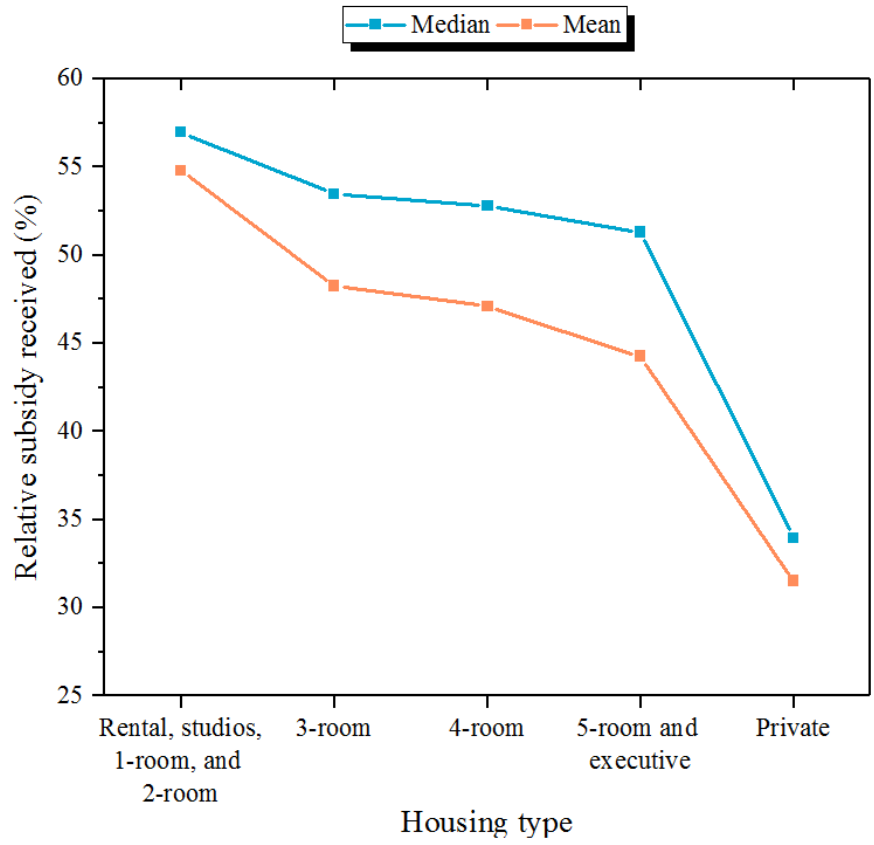


Figure 7. Mean and median relative subsidy received (RSR) within each housing type.



XSL•FO
RenderX

Table 5. Characteristics of patient cohort.

Variables	Patient cohort		
	Total (n=549,109)	21 to 65 years (n=458,069)	>65 years (n=91,040)
Male, n (%)[a]	311,650 (56.76)	267,061 (58.30)	44,589 (48.98)
Age as at 2013 in years[b], n (%)[a]			
21-29	96,856 (17.64)	96,856 (21.14)	—[c]
30-39	132,758 (24.18)	132,758 (28.98)	—
40-49	98,124 (17.87)	98,124 (21.42)	—
50-59	85,438 (15.56)	85,438 (18.65)	—
60-69	67,013 (12.20)	44,893 (9.80)	22,120 (24.30)
70-79	42,327 (7.71)	—	42,327 (46.49)
≥80	26,593 (4.84)	—	26,593 (29.21)
Race, n (%)[a]			
Chinese	329,544 (60.01)	260,656 (56.90)	68,888 (75.67)
Indian	70,151 (12.78)	64,483 (14.08)	5,668 (6.23)
Malay	63,660 (11.59)	54,670 (11.93)	8,990 (9.87)
Others	85,754 (15.62)	78,261 (17.08)	7,493 (8.23)
Singaporean, n (%)[a]	357,009 (65.02)	279,729 (61.07)	77,280 (84.89)
Subsidy status, n (%)[a]			
Subsidized	340,384 (61.99)	280,810 (61.30)	59,574 (65.44)
Mixed	154,895 (28.21)	131,943 (28.80)	22,952 (25.21)
Nonsubsidized	53,830 (9.80)	45,316 (9.89)	8,514 (9.35)
Housing types, n (%)[a]			
Rental, studios, 1-room, and 2-room	14,618 (2.66)	10,881 (2.38)	3,737 (4.10)
3-room	92,137 (16.78)	71,619 (15.63)	20,518 (22.54)
4-room	141,637 (25.79)	116,861 (25.51)	24,776 (27.21)
5-room and executive	119,845 (21.83)	100,116 (21.86)	19,729 (21.67)
Private	67,152 (12.23)	54,850 (11.97)	12,302 (13.51)
Missing	113,720 (20.71)	103,742 (22.65)	9,978 (10.96)
CCS[d] chronic conditions (primary and secondary), n (%)[a]			
Essential hypertension	67,611 (12.31)	29,705 (6.48)	37,906 (41.64)
Disorders of lipid metabolism	46,060 (8.39)	21,860 (4.77)	24,200 (26.58)
Diabetes mellitus	43,267 (7.88)	20,725 (4.52)	22,542 (24.76)
Acute cerebrovascular disease	17,731 (3.23)	7344 (1.60)	10,387 (11.41)
Asthma	10,177 (1.85)	7588 (1.66)	2589 (2.84)
Chronic obstructive pulmonary disease and bronchiectasis	8195 (1.49)	3342 (0.73)	4853 (5.33)
Numerical–total (median; interquartile range)			
Inflation-adjusted hospital charges (SG $)	5,177,231,809 (1846; 419-7696)	3,203,726,321 (1363; 352-5366)	1,973,505,488 (9186; 2595-26,100)
Inpatient visits	411,266 (0; 0-1)	251,891 (0; 0-1)	159,375 (1; 0-2)
Length-of-stay (days)	2,470,759 (0; 0-3)	1,330,125 (0; 0-2)	1,140,634 (3; 0-14)
Outpatient visits	7,367,495 (4; 1-13)	5,136,750 (4; 1-11)	2,230,745 (10; 3-29)

XSL•FO
RenderX

Variables	Patient cohort		
	Total (n=549,109)	21 to 65 years (n=458,069)	>65 years (n=91,040)
Specialist outpatient clinic visits	4,122,156 (2; 0-8)	2,821,298 (2; 0-6)	1,300,858 (6; 1-18)
Emergency department visits	834,192 (1; 1-2)	644,230 (1; 1-2)	189,962 (1; 1-2)
CCI[e]	−(0; 0-0)	−(0; 0-0)	−(1;0-2)
PPS[f]	−(3; 1-9)	−(3; 1-7)	−(9; 3-17)

[a]As a percentage of respective patient cohorts.

[b]Age of death if patient died before 2013.

[c]Not applicable.

[d]CCS: Clinical Classification Software.

[e]CCI: Charlson Comorbidity Index.

[f]PPS: Polypharmacy Score.

Discussion

Principal Findings and Generalizability

Conversion to ICD-10 codes from other codes such as ICD-9 or International Classification of Primary Care is commonly applied in medical or health care studies to increase granularity for identification and attribution of pathology at an individual level. However, given that our dataset is prepped for future health services studies, our primary objective in code standardization was to balance code sparsity with granularity. In this regard, backward mapping from ICD-10 to ICD-9 codes was a more suitable method for this system, and the phenotyping of these standardized codes using broader CCS codes provided different levels of granularity in line with our objectives. Our study showed that standardization of diagnosis codes to ICD-9-CM codes from ICD-10-AM and phenotyping to broader CCS groups through open source-mapping tables could achieve high mapping rates of more than 81% and 88%, respectively. The mapping rates could be further improved through code modification to rates in excess of 97% for ICD PD. Code modification through truncation or zero addition as applied in our study was a robust way of improving the mapping rates as shown by high validity when assessed by independent physicians. Overall, we also showed that bias resulting from code modification was small in our dataset, given that modified codes only constituted less than 2% of total ICD codes and 9% of total CCS codes and that high validity was observed even with these modifications. Given the frequent shifts in ICD codes, these results assure health services researchers that the use of open source-mapping tables together with code modification can rapidly standardize diagnosis coding with low biases and high validity to facilitate retrospective longitudinal analyses. However, we advise caution to researchers who wish to use the ICD-9-CM codes (original and mapped together) directly, without collapsing to CCS codes for their studies as there were ICD-9-CM codes that were unmapped to. Further details on this can be found in Multimedia Appendix 4.

CCI and PPS were introduced as proxy measures of disease complexity. CCI and PPS demonstrated positive correlation with health care utilization measures in keeping with theoretical understanding that patients with more complex disease consume more health care [72]. CCI and PPS were also moderately correlated, which is expected given that both are measures of disease complexity. These associations held true after multivariate regression analysis, demonstrating criterion validity of the measures as proxies for disease complexity. Although other studies found a similar association and effect size between CCI and LOS [73] and between CCI and PPS [74], our study was the first to find such an association between CCI and PPS with health care utilization measures such as inpatient admissions, SOC, and ED visits. These findings support the use of CCI and PPS as measures to stratify patients by complexity and possibly as an aggregate measure of health care utilization, given their correlation with all health care utilization metrics. This finding could be useful in works on profiling, risk stratification, and predictive modeling.

SES is a key determinant of health outcomes and health care utilization [75]. Neither direct measures through individual or household income nor alternate measures of SES such as area-based income were available in our dataset. Hence, we proposed an alternative method of estimating SES using housing type and size because of data availability and the housing landscape in Singapore. In Singapore, the proportion of bill that is subsidized is determined after a rigorous financial assessment and pegged to the income level of the patient (with lower income patients receiving greater levels of subsidy); hence, we hypothesize that lower SES groups would have a greater proportion of their bills subsidized. Given that subsidized care is lower in cost compared with nonsubsidized care, we also expect lower SES groups to opt for subsidized care. In our study, we showed that with decreasing size of housing, the proportion of the hospital bill subsidized increased and the proportion of patients who opted for subsidized care increased. This observation is consistent with our hypothesis that patients who stay in smaller housing types had a greater proportion of their bills subsidized and tended to opt for subsidized care. We have thus shown that in Singapore, housing type and size derived through postal code data are good proxy for income level and SES. Although other studies have shown that staying in rental housing is associated with an increased risk of frequent admissions [76] and readmission [75], as far as the authors are aware, there have not been studies in the Singapore context that have demonstrated the use of housing type as a proxy for SES. Although a missing rate of 20.7% was observed for the housing

XSL•FO
RenderX

type variable, this was attributed to foreign patients who registered nonresidential or overseas addresses (86.6% of those with missing housing data are nonresidents). Hence, the missing data are unlikely to bias the findings described above. We were also not able to account for any changes in housing type during the study as the EMR only captured the last postal code of the patient. Public resale data from HDB showed that only 1.6% of public housing units had a change in ownership in 2013 [77]. Hence, we believe it is reasonable to assume that the housing information is static.

Our finding on the suitability of housing type as a proxy for SES is useful, given that most clinical and administrative databases collect addresses but not direct SES information or other proxies. Our method of estimating SES would serve well in countries where methods of estimating SES such as area-based estimates [34,35], insurance status [36], and property value [37,38] are not suitable because of contextual reasons and unavailability of data. For example, area-based estimates may not be suitable in countries where spatial segregation level is low such as in many densely built cities in Asia. In such densely built cities, area-based estimations in effect would need to go down to blocks, which would be similar to using postal codes or addresses. Insurance status is better applied to countries that have high health insurance coverage, which is not the case in most of Asia. Finally, in countries where the real estate market is volatile, property value may be difficult to interpret as a proxy of SES, as the measure would reflect supply and demand dynamics at the point of estimate and numerous extrinsic factors unrelated to SES.

Unlike in countries where zip codes are area-based, the postal codes in Singapore are assigned to each individual building; hence, they serve almost like an address. Housing in Singapore can be divided into 3 main classes: private housing, public housing, and public rental housing. The private housing caters mainly to the upper-middle to upper income groups, whereas the public housing caters to the middle-class population, with 80% of the permanent population living in public housing as owner-occupiers [78]. Eligibility for public housing schemes and new units of certain public housing types depends on household incomes. Moreover, 6% of the public housing stocks are rental units, which serve as social housing for the underprivileged (households with income not exceeding SG $1500) [79]. The housing estates in Singapore were carefully designed to prevent the formation of social enclaves. In the absence of social enclaves (where there is a high concentration of either low or high value housing in an area) [80], area-based estimates are likely to be less valid. With more countries and cities adopting public housing policies and town planning measures to reduce the formation of urban ghettos and sharp sociospatial divisions [81,82] and higher proportions of the population living in tiered public housing [83,84], we do see the applicability of our proposed approach outside of Singapore. Hong Kong is an example of a city with similar ecology where the proposed approach to estimate SES could be used. Although the details may vary, the principle of stratification by type of housing tenure (eg, rental [low-income social housing], public housing, and private housing) first followed by unit size within each tenure type can still be adopted. In countries where urban social residential enclaves exist, 2-stage estimation of SES may be worth exploring by incorporating area-based indices with housing type approach proposed in this study to alleviate the problem of ecological fallacy from solely using area-based indices [85].

Finally, our cohort was found to be similar in profile with the Singapore national population. Comparison with National Census data in 2010 [86] found a similar trend in demographics and housing type, with the exception that our cohort skewed older, which is not unexpected given that health care utilization has been shown to increase with age [87,88]. Patients without PD are excluded from our cohort. These patients exist in our database because it was not mandatory for doctors to key in PD codes for outpatient visits. This would underestimate the number of patients who solely received outpatient care. As such, results from future analysis using the cohort would need to be interpreted with this limitation in mind. Within our cohort, differences in disease profile, disease complexity, and health care utilization could be observed when divided by age.

Conclusions

With increasing digitization of medical records, use of wearables and Internet-of-Things–connected devices in health care, the amount of data generated by health care systems is growing at a tremendous rate [89,90]. Being able to quickly process and analyze the data generated is key to health care transformation that is needed for sustainability [91]. In this study, we demonstrated how an EMR system in an AMC was processed for health services research. The approach (in whole or part) could be generalized to other EMR systems structured in a similar fashion to support research efforts. In addition, further analyses to better understand differences in the cohorts [1,92] would allow us to better segment the population and eventually predict cost and utilization drivers [4,93]. This is key as we seek to transform care and reduce utilization through targeted interventions and system redesign. The processed database with its multilevel views across time, as well as primary and secondary variables would be integral in achieving these goals.

Acknowledgments

The study is cofunded by the National University Health System and National University of Singapore (NUS) and approved by the Domain Specific Review Board (DSRB), National Healthcare Group, Singapore (2016/01011), and the data approved as a DSRB Standing Database: NUS-SSHSPH/2015-00032. The database is stored in secured micro-access laboratories protected with 2-factor authentication for entry, locked down workstations, and 24-hour video and electronic surveillance. All identifiers in Patient Affordability Simulation System (PASS) have also been masked to avoid leaking sensitive patient data [94]. The authors would like to thank Associate Professor Alex Cook (Saw Swee Hock School of Public Health [SSHSPH], NUS) for providing

XSL•FO
RenderX

the Property Guru data (in 2015) to identify private housing type in PASS. The authors would also like to thank the advisory panel for this study, consisting of Professor Teo Yik Ying (SSHSPH, NUS), Associate Professor Joanne Yoong (SSHSPH, NUS), Assistant Professor Tan Chuen Seng (SSHSPH, NUS), Assistant Professor Mornin Feng (SSHSPH, NUS), and Assistant Professor Sue-Anne Toh (Yong Loo Lin School of Medicine, NUS) for their inputs and guidance.

Authors' Contributions

XQT, DDW, and NR conceived this manuscript. DDW and NR drafted the manuscript and interpreted the data under the supervision of XQT. NR performed statistical analyses. XQT and WPG validated the diagnosis codes. XQT and AK acquired funding for the project. SHXN, SR, SS, DDW, and NR designed and curated the data. CST provided resources and software for this project. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Basic row elements and column variables in each table in Patient Affordability Simulation System (PASS).

[PDF File (Adobe PDF File), 26KB - medinform_v6i4e10933_app1.pdf]

Multimedia Appendix 2

Summary of primary and secondary variables.

[XLSX File (Microsoft Excel File), 9KB - medinform_v6i4e10933_app2.xlsx]

Multimedia Appendix 3

Description of extracted primary and secondary variables.

[XLSX File (Microsoft Excel File), 15KB - medinform_v6i4e10933_app3.xlsx]

Multimedia Appendix 4

International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes that were not mapped back from International Classification of Diseases, Tenth Revision, Australian Modification (ICD-10-AM) because of absence of equivalence map in Australian Consortium for Classification Development (ACCD) mapping table.

[XLSX File (Microsoft Excel File), 14KB - medinform_v6i4e10933_app4.xlsx]

References

1. Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. Can Med Assoc J 2016 Feb 16;188(3):182-188 [FREE Full text] [doi: 10.1503/cmaj.150064] [Medline: 26755672]
2. Markatou M, Don PK, Hu J, Wang F, Sun J, Sorrentino R, et al. Case-based reasoning in comparative effectiveness research. IBM J Res Dev 2012 Sep;56(5):4:1-412. [doi: 10.1147/JRD.2012.2198311]
3. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Med Care 2010 Jun;48(6 Suppl):S106-S113. [doi: 10.1097/MLR.0b013e3181de9e17] [Medline: 20473190]
4. Chechulin Y, Nazerian A, Rais S, Malikov K. Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada). Healthc Policy 2014 Feb;9(3):68-79 [FREE Full text] [doi: 10.12927/hcpol.2014.23710] [Medline: 24726075]
5. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a health knowledge graph from electronic medical records. Sci Rep 2017 Jul 20;7(1):5994 [FREE Full text] [doi: 10.1038/s41598-017-05778-z] [Medline: 28729710]
6. Abrahão MT, Nobre MR, Gutierrez MA. A method for cohort selection of cardiovascular disease records from an electronic health record system. Int J Med Inform 2017 Dec;102:138-149. [doi: 10.1016/j.ijmedinf.2017.03.015] [Medline: 28495342]
7. Batra S, Sachdeva S. Organizing standardized electronic healthcare records data for mining. Health Policy and Technology 2016 Sep;5(3):226-242. [doi: 10.1016/j.hlpt.2016.03.006]
8. Tran T, Luo W, Phung D, Gupta S, Rana S, Kennedy RL, et al. A framework for feature extraction from hospital medical data with applications in risk prediction. BMC Bioinformatics 2014 Dec 30;15:425 [FREE Full text] [doi: 10.1186/s12859-014-0425-8] [Medline: 25547173]
9. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. Annu Rev Public Health 2016;37:61-81. [doi: 10.1146/annurev-publhealth-032315-021353] [Medline: 26667605]

XSL•FO
RenderX

10. Hota B, Jones RC, Schwartz DN. Informatics and infectious diseases: what is the connection and efficacy of information technology tools for therapy and health care epidemiology? Am J Infect Control 2008 Apr;36(3):S47-S56. [doi: 10.1016/j.ajic.2007.07.005]
11. Ford JB, Roberts CL, Taylor LK. Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge data. Paediatr Perinat Epidemiol 2006 Jul;20(4):329-337. [doi: 10.1111/j.1365-3016.2006.00715.x] [Medline: 16879505]
12. Randall SM, Ferrante AM, Boyd JH, Semmens JB. The effect of data cleaning on record linkage quality. BMC Med Inform Decis Mak 2013 Jun 05;13:64 [FREE Full text] [doi: 10.1186/1472-6947-13-64] [Medline: 23739011]
13. Byrd JB, Vigen R, Plomondon ME, Rumsfeld JS, Box TL, Fihn SD, et al. Data quality of an electronic health record tool to support VA cardiac catheterization laboratory quality improvement: the VA Clinical Assessment, Reporting, and Tracking System for Cath Labs (CART) program. Am Heart J 2013 Mar;165(3):434-440. [doi: 10.1016/j.ahj.2012.12.009] [Medline: 23453115]
14. Bentley JP, Ford JB, Taylor LK, Irvine KA, Roberts CL. Investigating linkage rates among probabilistically linked birth and hospitalization records. BMC Med Res Methodol 2012 Sep 25;12:149 [FREE Full text] [doi: 10.1186/1471-2288-12-149] [Medline: 23009079]
15. 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]
16. Gunapal PP, Kannapiran P, Teow KL, Zhu Z, Xiaobin You A, Saxena N, et al. Setting up a regional health system database for seamless population health management in Singapore. Proc Singapore Healthcare 2015 Oct 16;25(1):27-34. [doi: 10.1177/2010105815611440]
17. Richesson RL. An informatics framework for the standardized collection and analysis of medication data in networked research. J Biomed Inform 2014 Dec;52:4-10 [FREE Full text] [doi: 10.1016/j.jbi.2014.01.002] [Medline: 24434192]
18. Sutherland SM, Kaelber DC, Downing NL, Goel VV, Longhurst CA. Electronic health record-enabled research in children using the electronic health record for clinical discovery. Pediatr Clin North Am 2016 Apr;63(2):251-268. [doi: 10.1016/j.pcl.2015.12.002] [Medline: 27017033]
19. Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2010 Aug;62(8):1120-1127 [FREE Full text] [doi: 10.1002/acr.20184] [Medline: 20235204]
20. Wei WQ, Bastarache LA, Carroll RJ, Marlo JE, Osterman TJ, Gamazon ER, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. PLoS One 2017;12(7):e0175508 [FREE Full text] [doi: 10.1371/journal.pone.0175508] [Medline: 28686612]
21. Williams R, Kontopantelis E, Buchan I, Peek N. Clinical code set engineering for reusing EHR data for research: a review. J Biomed Inform 2017 Dec;70:1-13 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.010] [Medline: 28442434]
22. Ainsworth J, Buchan I. Combining health data uses to ignite health system learning. Methods Inf Med 2015;54(6):479-487. [doi: 10.3414/ME15-01-0064] [Medline: 26395036]
23. Bello A, Hemmelgarn B, Manns B, Tonelli M, Alberta Kidney Disease Network. Use of administrative databases for health-care planning in CKD. Nephrol Dial Transplant 2012 Oct;27(Suppl 3):iii12-iii18. [doi: 10.1093/ndt/gfs163] [Medline: 22734112]
24. Abrahamsen B, Eiken P, Eastell R. Subtrochanteric and diaphyseal femur fractures in patients treated with alendronate: a register-based national cohort study. J Bone Miner Res 2009 Jun;24(6):1095-1102 [FREE Full text] [doi: 10.1359/jbmr.081247] [Medline: 19113931]
25. Williams RR, Horm JW. Association of cancer sites with tobacco and alcohol consumption and socioeconomic status of patients: interview study from the Third National Cancer Survey. J Natl Cancer Inst 1977 Mar;58(3):525-547. [Medline: 557114]
26. Birim O, Maat AP, Kappetein AP, van Meerbeeck JP, Damhuis RA, Bogers AJ. Validation of the Charlson comorbidity index in patients with operated primary non-small cell lung cancer. Eur J Cardiothorac Surg 2003 Jan;23(1):30-34. [Medline: 12493500]
27. Alter DA, Austin PC, Naylor CD, Tu JV. Factoring socioeconomic status into cardiac performance profiling for hospitals: does it matter? Med Care 2002 Jan;40(1):60-67. [doi: 10.1097/00005650-200201000-00008] [Medline: 11748427]
28. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
29. Lorimer JW, Doumit G. Comorbidity is a major determinant of severity in acute diverticulitis. Am J Surg 2007 Jun;193(6):681-685. [doi: 10.1016/j.amjsurg.2006.10.019] [Medline: 17512276]
30. Marya SK, Amit P, Singh C. Impact of Charlson indices and comorbid conditions on complication risk in bilateral simultaneous total knee arthroplasty. Knee 2016 Dec;23(6):955-959. [doi: 10.1016/j.knee.2016.05.013] [Medline: 27802921]

XSL•FO
RenderX

31. Hennus PM, Kroeze SG, Bosch JL, Jans JJ. Impact of comorbidity on complications after nephrectomy: use of the Clavien Classification of Surgical Complications. BJU Int 2012 Sep;110(5):682-687 [FREE Full text] [doi: 10.1111/j.1464-410X.2011.10889.x] [Medline: 22432906]
32. Higdon R, Stewart E, Roach JC, Dombrowski C, Stanberry L, Clifton H, et al. Predictive Analytics In Healthcare: Medications as a Predictor of Medical Complexity. Big Data 2013 Dec;1(4):237-244. [doi: 10.1089/big.2013.0024] [Medline: 27447256]
33. Salvi F, Rossi L, Lattanzio F, Cherubini A. Is polypharmacy an independent risk factor for adverse outcomes after an emergency department visit? Intern Emerg Med 2017 Mar;12(2):213-220. [doi: 10.1007/s11739-016-1451-5] [Medline: 27075646]
34. Australian Bureau of Statistics. Socio-Economic Indexes for Areas 2013. URL: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa [WebCite Cache ID 6yyvuQc1A]
35. Berkman LF, Macintyre S. The measurement of social class in health studies: old measures and new formulations. IARC Sci Publ 1997(138):51-64. [Medline: 9353663]
36. Foraker RE, Rose KM, Whitsel EA, Suchindran CM, Wood JL, Rosamond WD. Neighborhood socioeconomic status, Medicaid coverage and medical management of myocardial infarction: atherosclerosis risk in communities (ARIC) community surveillance. BMC Public Health 2010 Oct 21;10:632 [FREE Full text] [doi: 10.1186/1471-2458-10-632] [Medline: 20964853]
37. Coffee NT, Lockwood T, Hugo G, Paquet C, Howard NJ, Daniel M. Relative residential property value as a socio-economic status indicator for health research. Int J Health Geogr 2013 Apr 15;12:22 [FREE Full text] [doi: 10.1186/1476-072X-12-22] [Medline: 23587373]
38. Ghawi H, Crowson CS, Rand-Weaver J, Krusemark E, Gabriel SE, Juhn YJ. A novel measure of socioeconomic status using individual housing data to assess the association of SES with rheumatoid arthritis and its mortality: a population-based case-control study. Br Med J Open 2015 Apr 29;5(4):e006469 [FREE Full text] [doi: 10.1136/bmjopen-2014-006469] [Medline: 25926142]
39. Ministry of Health Singapore. Reorganisation of healthcare system into three integrated clusters to better meet future healthcare needs. URL: https://www.moh.gov.sg/content/moh_web/home/pressRoom/pressRoomItemRelease/2017/reorganisation-of-healthcare-system-into-three-integrated-cluste.html [WebCite Cache ID 6yywI3Yvt]
40. Deng X, Lin WH, Tai ES, Khoo YH, Salloway MK, Tan CS. From descriptive to diagnostic analytics for assessing data quality: An application to temporal data elements in electronic health records. 2016 Presented at: 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); Feb 24-27, 2016; Las Vegas, USA. [doi: 10.1109/BHI.2016.7455878]
41. Salloway MK, Ling ZJ, Yang Q, Lew HY, Deng X, Chia KS, et al. Towards Utilizing Electronic Medical Records for Public Health in Small Steps. 2015 Presented at: 4th Annual Global Healthcare Conference; Jun 29-30, 2015; Singapore. [doi: 10.5176/2251-3833_GHC15.59]
42. Australian Consortium for Classification Development URL: https://www.accd.net.au/Downloads.aspx [accessed 2018-04-27] [WebCite Cache ID 6yywKAGbi]
43. Fung KW, Richesson R, Smerek M, Pereira KC, Green BB, Patkar A, et al. Preparing for the ICD-10-CM Transition: automated methods for translating ICD codes in clinical phenotype definitions. EGEMS (Wash DC) 2016;4(1):1211 [FREE Full text] [doi: 10.13063/2327-9214.1211] [Medline: 27195309]
44. Venepalli NK, Qamruzzaman Y, Li JJ, Lussier YA, Boyd AD. Identifying clinically disruptive International Classification of Diseases 10th Revision Clinical Modification conversions to mitigate financial costs using an online tool. J Oncol Pract 2014 Mar;10(2):97-103. [doi: 10.1200/JOP.2013.001156] [Medline: 24520143]
45. Krive J, Patel M, Gehm L, Mackey M, Kulstad E, Li JJ, et al. The complexity and challenges of the International Classification of Diseases, Ninth Revision, Clinical Modification to International Classification of Diseases, 10th Revision, Clinical Modification transition in EDs. Am J Emerg Med 2015 May;33(5):713-718 [FREE Full text] [doi: 10.1016/j.ajem.2015.03.001] [Medline: 25863652]
46. Farzandipour M, Sheikhtaheri A, Sadoughi F. Effective factors on accuracy of principal diagnosis coding based on International Classification of Diseases, the 10th revision (ICD-10). Int J Inf Manage 2010 Feb;30(1):78-84. [doi: 10.1016/j.ijinfomgt.2009.07.002]
47. Cave AJ, Davey C, Ahmadi E, Drummond N, Fuentes S, Kazemi-Bajestani SM, et al. Development of a validated algorithm for the diagnosis of paediatric asthma in electronic medical records. NPJ Prim Care Respir Med 2016 Dec 24;26:16085 [FREE Full text] [doi: 10.1038/npjpcrm.2016.85] [Medline: 27882997]
48. Cozzolino F, Abraha I, Orso M, Mengoni A, Cerasa MF, Eusebi P, et al. Protocol for validating cardiovascular and cerebrovascular ICD-9-CM codes in healthcare administrative databases: the Umbria Data Value Project. Br Med J Open 2017 Dec 29;7(3):e013785 [FREE Full text] [doi: 10.1136/bmjopen-2016-013785] [Medline: 28360241]
49. Narongroeknawin P, Patkar NM, Shakoory B, Jain A, Curtis JR, Delzell E, et al. Validation of diagnostic codes for subtrochanteric, diaphyseal, and atypical femoral fractures using administrative claims data. J Clin Densitom 2012;15(1):92-102 [FREE Full text] [doi: 10.1016/j.jocd.2011.09.001] [Medline: 22071028]
50. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. Health Serv Res 2005 Oct;40(5 Pt 2):1620-1639 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00444.x] [Medline: 16178999]

XSL•FO
RenderX

51. Agency for Healthcare Research and Quality. Clinical Classifications Software (CCS) for ICD-9-CM 2015. URL: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp [WebCite Cache ID 6yywPdHN8]
52. Thompson DA, Makary MA, Dorman T, Pronovost PJ. Clinical and economic outcomes of hospital acquired pneumonia in intra-abdominal surgery patients. Ann Surg 2006 Apr;243(4):547-552. [doi: 10.1097/01.sla.0000207097.38963.3b] [Medline: 16552208]
53. Brousseau DC, Owens PL, Mosso AL, Panepinto JA, Steiner CA. Acute care utilization and rehospitalizations for sickle cell disease. J Am Med Assoc 2010 Apr 07;303(13):1288-1294. [doi: 10.1001/jama.2010.378] [Medline: 20371788]
54. Anderson FA, Zayaruzny M, Heit JA, Fidan D, Cohen AT. Estimated annual numbers of US acute-care hospital patients at risk for venous thromboembolism. Am J Hematol 2007 Sep;82(9):777-782 [FREE Full text] [doi: 10.1002/ajh.20983] [Medline: 17626254]
55. R Core Team. R-project. Vienna, Austria R: A Language and Environment for Statistical Computing 3.2.0 ed URL: https://www.r-project.org/ [accessed 2018-11-01] [WebCite Cache ID 73cKlsMOZ]
56. Wickham H. Github. Texas, USA multidplyr: Partitioned data frames for 'dplyr' 0.0.0.9000 ed URL: https://github.com/hadley/multidplyr [accessed 2018-11-01] [WebCite Cache ID 73cL7X17d]
57. Monetary Authority of Singapore. Goods & Services Inflation Calculator. URL: https://secure.mas.gov.sg/calculator/goodsandservices.aspx [WebCite Cache ID 6yywRRZ2r]
58. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005 Nov;43(11):1130-1139. [Medline: 16224307]
59. Wasey JO. icd: Comorbidity Calculations and Tools for ICD-9 and ICD-10 Codes. Vienna, Austria URL: https://cran.r-project.org/web/packages/icd/index.html [WebCite Cache ID 73cLLE3Vi]
60. Duerden M, Avery T, Payne R. Kingsfund. London, United Kingdom: The King's Fund; 2013. Polypharmacy and Medicines Optimisation: Making It Safe and Sound URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/polypharmacy-and-medicines-optimisation-kingsfund-nov13.pdf [accessed 2018-11-01] [WebCite Cache ID 73cLdFqHS]
61. Goodney PP, Stukel TA, Lucas FL, Finlayson EV, Birkmeyer JD. Hospital volume, length of stay, and readmission rates in high-risk surgery. Ann Surg 2003 Aug;238(2):161-167. [doi: 10.1097/01.SLA.0000081094.66659.c3] [Medline: 12894006]
62. McClellan CB, Schatz JC, Mark TR, McKelvy A, Puffer E, Roberts CW, et al. Criterion and convergent validity for 4 measures of pain in a pediatric sickle cell disease population. Clin J Pain 2009 Feb;25(2):146-152 [FREE Full text] [doi: 10.1097/AJP.0b013e3181839ac4] [Medline: 19333161]
63. Svedberg P, Nygren JM, Staland-Nyman C, Nyholm M. The validity of socioeconomic status measures among adolescents based on self-reported information about parents occupations, FAS and perceived SES; implication for health related quality of life studies. BMC Med Res Methodol 2016 Dec 29;16:48 [FREE Full text] [doi: 10.1186/s12874-016-0148-9] [Medline: 27130331]
64. Lahey BB, Rathouz PJ, Keenan K, Stepp SD, Loeber R, Hipwell AE. Criterion validity of the general factor of psychopathology in a prospective study of girls. J Child Psychol Psychiatry 2015 Apr;56(4):415-422 [FREE Full text] [doi: 10.1111/jcpp.12300] [Medline: 25052460]
65. OneMap Singapore. Reverse Geocode(WGS84). URL: https://docs.onemap.sg/#reverse-geocode-wgs84 [WebCite Cache ID 6yywTDyuJ]
66. Housing & Development Board (HDB) Singapore. HDB Map Services 2017. URL: https://services2.hdb.gov.sg/web/fi10/emap.html [WebCite Cache ID 73c7S8dxN]
67. Lu JR, Leung GM, Kwon S, Tin KY, Van Doorslaer E, O'Donnell O. Horizontal equity in health care utilization evidence from three high-income Asian economies. Soc Sci Med 2007 Jan;64(1):199-212. [doi: 10.1016/j.socscimed.2006.08.033] [Medline: 17014944]
68. Ministry of Health Singapore. Healthcare we can all afford 2009. URL: https://www.moh.gov.sg/content/dam/moh_web/Publications/Educational%20Resources/2009/MT%20pamphlet%20(English).pdf [WebCite Cache ID 6yywX9K1O]
69. Tsuji N, Kakee N, Ishida Y, Asami K, Tabuchi K, Nakadate H, et al. Validation of the Japanese version of the Pediatric Quality of Life Inventory (PedsQL) Cancer Module. Health Qual Life Outcomes 2011 Apr 10;9:22 [FREE Full text] [doi: 10.1186/1477-7525-9-22] [Medline: 21477361]
70. Swank JM, Mullen PR. Evaluating evidence for conceptually related constructs using bivariate correlations. Meas Eval Couns Dev 2017 Oct 04;50(4):270-274. [doi: 10.1080/07481756.2017.1339562]
71. Elixhauser A, Steiner C. Clinical Classification Software (CCS) 2015. 2016. Agency for Healthcare Research and Quality URL: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/CCSUsersGuide.pdf [WebCite Cache ID 6yywa6ZQW]
72. Szekendi MK, Williams MV, Carrier D, Hensley L, Thomas S, Cerese J. The characteristics of patients frequently admitted to academic medical centers in the United States. J Hosp Med 2015 Sep;10(9):563-568 [FREE Full text] [doi: 10.1002/jhm.2375] [Medline: 26018340]
73. Sockalingam S, Alzahrani A, Meaney C, Styra R, Tan A, Hawa R, et al. Time to consultation-liaison psychiatry service referral as a predictor of length of stay. Psychosomatics 2016;57(3):264-272. [doi: 10.1016/j.psym.2016.01.005] [Medline: 27005725]
74. Brilleman SL, Salisbury C. Comparing measures of multimorbidity to predict outcomes in primary care: a cross sectional study. Fam Pract 2013 Apr;30(2):172-178 [FREE Full text] [doi: 10.1093/fampra/cms060] [Medline: 23045354]

XSL•FO
RenderX

75. Low LL, Wah W, Ng MJ, Tan SY, Liu N, Lee KH. Housing as a social determinant of health in Singapore and its association with readmission risk and increased utilization of hospital services. Front Public Health 2016;4:109 [FREE Full text] [doi: 10.3389/fpubh.2016.00109] [Medline: 27303662]
76. Low LL, Tay WY, Ng MJ, Tan SY, Liu N, Lee KH. Frequent hospital admissions in Singapore: clinical risk factors and impact of socioeconomic status. Singapore Med J 2018 Dec;59(1):39-43 [FREE Full text] [doi: 10.11622/smedj.2016110] [Medline: 27311740]
77. Housing & Development Board (HDB) Singapore. 2014. Annual Reports URL: https://www.hdb.gov.sg/cs/infoweb/about-us/news-and-publications/annual-reports [accessed 2018-10-01] [WebCite Cache ID 73c89co1c]
78. Housing & Development Board (HDB) Singapore. Public Housing - A Singapore Icon 2018 URL: https://www.hdb.gov.sg/cs/infoweb/about-us/our-role/public-housing--a-singapore-icon [accessed 2018-11-01] [WebCite Cache ID 73c9UDB3O]
79. Housing & Development Board (HDB) Singapore. Hbd. 2016/ 2017 Annual Report 2018 URL: https://www.hdb.gov.sg/cs/infoweb/about-us/news-and-publications/annual-reports [accessed 2018-11-02] [WebCite Cache ID 73c9gucMf]
80. Shanmugaratnam T. Straitstimes. The relentless effort that goes into keeping S'pore inclusive URL: https://www.straitstimes.com/opinion/the-relentless-effort-that-goes-into-keeping-spore-inclusive [accessed 2018-10-01]
81. Forster W. 80 Years of Social Housing in Vienna. 80 Years of Social Housing in Vienna: City of Vienna; 2018. URL: https://www.wien.gv.at/english/housing/promotion/pdf/socialhous.pdf [WebCite Cache ID 732RZV4Ek]
82. Yip N. Residential Segregation in an Unequal City: Why are There No Urban Ghettos in Hong Kong? In: Maloutas T, editor. Residential Segregation in Comparative Perspective: Making Sense of Contextual Diversity. New York, USA: Routledge; 2016.
83. Transport and Housing Bureau Hong Kong. Housing Figures 2017. URL: https://www.thb.gov.hk/eng/psp/publications/housing/HIF2017.pdf [WebCite Cache ID 732SbTlX7]
84. Schantl C. Wienerwohnen. 2016. Municipal Housing in Vienna. History, facts & figures URL: https://www.wienerwohnen.at/dms/workspace/SpacesStore/aa75756e-2836-4e77-8cfd-f37cc15e2756/1.0Wiener-Gemeindebau-engl.pdf [accessed 2018-10-09] [WebCite Cache ID 732SSdOqX]
85. Shaw M. Housing and public health. Annu Rev Public Health 2004;25:397-418. [doi: 10.1146/annurev.publhealth.25.101802.123036] [Medline: 15015927]
86. Singstat. Census of population 2010 statistical release 1: Demographic characteristics, education, language and religion URL: https://www.singstat.gov.sg/-/media/files/publications/cop2010/census_2010_release1/cop2010sr1.pdf [accessed 2018-10-09] [WebCite Cache ID 732Sn49tc]
87. Fazeli DS, Hall KS, Dalton VK, Carlos RC. The link between everyday discrimination, healthcare utilization, and health status among a national sample of women. J Womens Health (Larchmt) 2016 Oct;25(10):1044-1051 [FREE Full text] [doi: 10.1089/jwh.2015.5522] [Medline: 27429363]
88. Wang Z, Li X, Chen M, Si L. Social health insurance, healthcare utilization, and costs in middle-aged and elderly community-dwelling adults in China. Int J Equity Health 2018 Feb 02;17(1):17 [FREE Full text] [doi: 10.1186/s12939-018-0733-0] [Medline: 29394933]
89. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff (Millwood) 2014 Jul;33(7):1115-1122. [doi: 10.1377/hlthaff.2014.0147] [Medline: 25006136]
90. Palanisamy V, Thirunavukarasu R. Implications of big data analytics in developing healthcare frameworks–A review. Journal of King Saud University - Computer and Information Sciences 2017 Dec epub ahead of date. [doi: 10.1016/j.jksuci.2017.12.007]
91. Revere D, Turner AM, Madhavan A, Rambo N, Bugni PF, Kimball A, et al. Understanding the information needs of public health practitioners: a literature review to inform design of an interactive digital knowledge management system. J Biomed Inform 2007 Aug;40(4):410-421 [FREE Full text] [doi: 10.1016/j.jbi.2006.12.008] [Medline: 17324632]
92. Lee NS, Whitman N, Vakharia N, Taksler GB, Rothberg MB. High-cost patients: hot-spotters don't explain the half of It. J Gen Intern Med 2017 Dec;32(1):28-34 [FREE Full text] [doi: 10.1007/s11606-016-3790-3] [Medline: 27480529]
93. Fleishman JA, Cohen JW. Using information on clinical conditions to predict high-cost patients. Health Serv Res 2010 Apr;45(2):532-552 [FREE Full text] [doi: 10.1111/j.1475-6773.2009.01080.x] [Medline: 20132341]
94. Salloway MK, Deng X, Ning Y, Kao SL, Chen Y, Schaefer GO, et al. A de-identification tool for users in medical operations and public health. 2016 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); Feb 24-27, 2016; Las Vegas, USA. [doi: 10.1109/BHI.2016.7455951]

Abbreviations

ACCD: Australian Consortium for Classification Development
AM: Australian Modification
AMC: Academic Medical Center
CCI: Charlson Comorbidity Index
CCS: Clinical Classifications Software
CM: Clinical Modification

XSL•FO
RenderX

DRG: Diagnosis-Related Group
ED: emergency department
EMR: electronic medical records
HDB: Housing Development Board
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10-AM: International Classification of Diseases, Tenth Revision, Australian Modification
LOS: length-of-stay
NUH: National University Hospital
OR: odds ratio
PASS: Patient Affordability Simulation System
PD: primary diagnosis
PPS: Polypharmacy Score
RSR: relative subsidy received
SD: secondary diagnosis
SES: socioeconomic status
SOC: specialist outpatient clinic

Edited by G Eysenbach; submitted 02.05.18; peer-reviewed by R Williams, A Boyd, R Carroll; comments to author 03.08.18; revised version received 09.10.18; accepted 10.10.18; published 21.12.18.

Please cite as:
Rahman N, Wang DD, Ng SHX, Ramachandran S, Sridharan S, Khoo A, Tan CS, Goh WP, Tan XQ
Processing of Electronic Medical Records for Health Services Research in an Academic Medical Center: Methods and Validation
JMIR Med Inform 2018;6(4):e10933
URL: http://medinform.jmir.org/2018/4/e10933/
doi:10.2196/10933
PMID:30578188

©Nabilah Rahman, Debby D Wang, Sheryl Hui-Xian Ng, Sravan Ramachandran, Srinath Sridharan, Astrid Khoo, Chuen Seng Tan, Wei-Ping Goh, Xin Quan Tan. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 21.12.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of an eHealth System to Capture and Analyze Patient Sensor and Self-Report Data: Mixed-Methods Assessment of Potential Applications to Improve Cancer Care Delivery

Alexander R Lucas[1], PhD; Michael B Bass[2], MS; Nan E Rothrock[2], PhD; Mary L O'Connor[2], MS; Mia R Sorkin[2], MPH; Jason Nawyn[3], SM; Fahd Albinali[4], PhD; Lynne I Wagner[1,5], PhD

[1]Department of Social Sciences and Health Policy, Wake Forest School of Medicine, Winston-Salem, NC, United States

[2]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Massachusetts Institute of Technology Media Lab, School of Architecture + Planning, Massachusetts Institute of Technology, Cambridge, MA, United States

[4]QMedic, Cambridge, MA, United States

[5]Comprehensive Cancer Center, Wake Forest Baptist Health, Winston-Salem, NC, United States

Corresponding Author:
Lynne I Wagner, PhD
Comprehensive Cancer Center
Wake Forest Baptist Health
1 Medical Center Boulevard
Winston-Salem, NC, 27157
United States
Phone: 1 336 713 1478
Fax: 1 336 716 7554
Email: lywagner@wakehealth.edu

Abstract

Background: Capturing and Analyzing Sensor and Self-Report Data for Clinicians and Researchers (COMPASS) is an electronic health (eHealth) platform designed to improve cancer care delivery through passive monitoring of patients' health status and delivering customizable reports to clinicians. Based on data from sensors and context-driven administration of patient-reported outcome (PRO) measures, key indices of patients' functional status can be collected between regular clinic visits, supporting clinicians in the delivery of patient care.

Objective: The first phase of this project aimed to systematically collect input from oncology providers and patients on potential clinical applications for COMPASS to refine the system.

Methods: Ten clinicians representing various oncology specialties and disciplines completed semi-structured interviews designed to solicit clinician input on how COMPASS can best support clinical care delivery. Three cancer patients tested a prototype of COMPASS for 7 days and provided feedback. Interview data were tabulated using thematic content analysis to identify the most clinically relevant objective and PRO domains.

Results: Thematic content analysis revealed that clinicians were most interested in monitoring vital statistics, symptoms, and functional status, including the physical activity level (n=9), weight (n=5), fatigue (n=9), sleep quality (n=8), and anxiety (n=7). Patients (2 in active treatment and 1 in remission) reported that they would use such a device, were enthusiastic about their clinicians monitoring their health status, especially the tracking of symptoms, and felt knowing their clinicians were monitoring and reviewing their health status provided valuable reassurance. Patients would, however, like to provide some context to their data.

Conclusions: Clinicians and patients both articulated potential benefits of the COMPASS system in improving cancer care. From a clinician standpoint, data need to be easily interpretable and actionable. The fact that patients and clinicians both see potential value in eHealth systems suggests wider adoption and utilization could prove to be a useful tool for improving care delivery.

(JMIR Med Inform 2018;6(4):e46) doi:10.2196/medinform.9525

XSL•FO
RenderX

KEYWORDS

cancer; care delivery; decision support; eHealth; mobile phone; survivorship; symptom monitoring

Introduction

In 2014, there were an estimated 14.5 million cancer survivors in the United States, and this number is expected to reach 19 million in 2024 [1]. The aging population, increased rates of screening [2,3], and improved availability and quality of treatments have resulted in cancer survivors living longer [4]; over the next decade, the proportion of 5-year cancer survivors is expected to increase by approximately 37% [5]. The provision of high-quality medical care for this growing segment of the population has been identified by the American Society of Clinical Oncology as a priority [6]. Specific priorities include (1) monitoring of patient-reported outcomes (PROs) such as symptoms, health status, and quality of life; (2) adherence to treatment regimens; and (3) monitoring of lifestyle and health-protective behaviors [7-10].

Current evidence shows that collecting PROs including symptoms (pain, fatigue, and nausea), psychosocial well-being, and quality of life yields better clinical outcomes [11,12], including potential survival benefit [13]. Worsening symptoms might signal disease recurrence or progression or the need for medication or dosage adjustments. However, there is a great deal of variability in how PROs are actually being collected and implemented in oncology care [14]. Typically, this type of information is only gathered at routine clinic visits, if at all, and, therefore, may not accurately reflect day-to-day functioning. Furthermore, as patients do not want to burden their care team, they tend not to report symptoms unless specifically asked or they will wait for their next scheduled clinic visit to report concerns [15]. The systematic integration of PROs into clinical care through utilizing an electronic platform that offers the ability to communicate with patients in real time could lead to shorter response times, better symptom management, and ultimately better outcomes [16]. Furthermore, efforts to integrate precise and robust symptom measures such as Patient-Reported Outcomes Measurement Information System (PROMIS) have emerged over the last few years [17].

Adherence to medication has emerged as a particular concern because of the rise in the use of oral anticancer drugs [18]. In terms of lifestyle and health-protective behaviors, physical activity, time spent in sedentary pursuits (television watching), and diet or nutrition markedly impact the physical, functional, and psychological health status of patients and survivors [19-21]. Physical activity is an especially important indicator of physical functioning, health-related quality of life, risk for a decline in health status [22,23], and mortality outcomes [24], meaning its maintenance is an important clinical goal. Failure to manage these aspects of care can lead to increased risk of developing comorbidities and, therefore, to excess economic burden associated with medical care, time off work, lost productivity at home, and additional medical visits [25,26].

However, there are challenges to monitoring the health status of patients in-between clinic visits, soliciting clinically relevant data from patients and family members in an efficient manner, corroborating patient self-report data (eg, physical activity), and integrating multiple sources of clinically relevant data in the context of busy oncology practice. Addressing these challenges can be a daunting task, chiefly because clinicians and their staff are already plagued by numerous competing demands. Therefore, developing innovative ways for clinicians to monitor their patients' behaviors and when needed providing guidance to help them with adherence to medicines, adoption and maintenance of healthy lifestyles, and cope with stress could potentially enhance the overall quality of supportive care and reduce the burden on both patients and their care teams. Electronic health (eHealth) models of care that leverage electronic health records, as well as digital and wearable technologies, are now emerging as an innovative strategy to reduce unmet care needs and support regular monitoring and interaction with patients between scheduled clinic visits [27].

Capturing and Analyzing Sensor and Self-Report Data for Clinicians and Researchers (COMPASS) is a device-agnostic eHealth technology platform that can passively and remotely monitor multiple domains of function and PROs. The COMPASS system includes (1) a device worn by patients to passively monitor physiological function; (2) an interface to sync with patients' smartphone; and (3) a Web-based clinician interface to deliver customizable reports. The purpose of this study was to explore the user requirements for such a system to ensure that it can adequately support the breadth and range of functionality typically requested by practitioners in the field. A user needs assessment was conducted to establish design and use metrics of a prototype COMPASS system before conducting more comprehensive testing and evaluation in a larger-scale phase II study.

Methods

Participants and Procedures

This study was conducted at the Robert H Lurie Comprehensive Cancer Center (RHLCCC) of Northwestern University and was approved by its Institutional Review Board. Potential participants, including clinicians and patients, provided written informed consent prior to participation. Clinicians were provided with a description of the COMPASS system 7 days prior to the interview with instructions to think about potential benefits of the system. Patients were provided with a prototype of the COMPASS app and wearable sensors (Mio Alpha Sports Watch) for 7 days. Both clinicians and patients then completed a semistructured interview with a trained interviewer. All interviews were audiorecorded and transcribed.

Clinician Interviews

Clinicians were eligible if they were current oncology providers at RHLCCC. A purposive sampling strategy was used to recruit a diverse sample of 10 oncology providers with regard to specialty and clinical practice foci. Clinicians agreeing to participate provided written, informed consent. We provided clinician participants with a description of COMPASS

XSL•FO
RenderX

approximately 1 week prior to the interview and asked them to think about how they would utilize the system to inform clinical care delivery (Multimedia Appendix 1, "Thought Exercise"). Clinicians then completed in-person semistructured interviews with a trained interviewer. Software developers who designed COMPASS (FA and JN) participated in the interviews via conference call. Each interview lasted approximately 1 hour. Interview content included information about their clinical practice (eg, specialty and types of patients typically seen), metrics most pertinent to treatment decision making and goals of care, how COMPASS could help to inform clinical visits, and preferences for how patient data collected through COMPASS are summarized and presented. Multimedia Appendix 2 provides examples of clinician interviews.

Textbox 1 presents examples of questions posed to clinicians. The semistructured interview solicited preferences for the types of possible wearable sensors, PROs most relevant to clinical practice, preferences for the format of data visualization options, and communication preferences including the sharing of patients' data and the frequency of contact with patients as a way of improving the delivery of care. The interview template evolved as interviews were conducted; therefore, only a subset of providers was asked to discuss the sharing of data with patients. Multimedia Appendix 3 provides a list of potential sensors that could be incorporated into the COMPASS system. Clinicians were provided with this list during the interviews to inform them of potential metrics that could be captured.

Patient Interviews and Testing

To be eligible for participation, patients had to be aged ≥18 years, diagnosed with cancer (any type, all stages), and had to own a smartphone. Patients could be at any stage of treatment, including posttreatment. Patients provided written informed consent prior to participating in any research activities. Patients were identified through participating clinicians and were approached in a clinic regarding participation. In addition, patients were recruited through a study brochure and flyers placed in RHLCCC clinical practice areas and through outreach on social media sites such as Twitter and Facebook. Multimedia Appendix 2 provides examples of patient interviews.

Each patient participated in two study visits. The first was conducted in-person at RHLCCC. Demographics and disease-specific information were gathered. A commercially available smartphone, with an armband (should they prefer to wear the smartphone rather than carry it), and Mio Alpha heart rate monitor wristwatch were provided to participants for the duration of study participation (7 days). The correct use of the technology was demonstrated, and participants were instructed to wear and interact with the wristwatch and smartphone app for 1 week. We asked participants to respond to brief PRO measures, with the explanation that data collected from smartphone-based surveys would not be assessed for content but rather for evaluating the feasibility and usability of this feature. Following the 7-day testing period, the second study visit was conducted at which time the devices were returned; the second study visit included a follow-up semistructured interview to collect data on wearability or usability of the device and patient experiences communicating with health care providers regarding data collected using the COMPASS system. The interview was approximately 1-hour long and was audiorecorded and transcribed. At the conclusion of the interview, a US $50 gift card was provided. Examples of patient interview questions are provided in Textbox 1. Figure 1 provides screenshots of the smartphone app and wristwatch device.

Analyses

All interviews were subjected to thematic content analysis (TCA) [28,29]. For clinician interviews, TCA was used to organize data according to the frequency and, therefore, the relative importance of the responses. The TCA was completed by two independent coders to tabulate the most common symptoms, clinical concerns, potential applications of the system and themes. Any coding discrepancies were discussed and resolved by the senior author (LIW). In addition, a list was generated and a frequency for each topic recorded. For patient interviews, a conventional qualitative content analysis [30] was used to analyze responses. Transcripts were read several times by the analyst, who determined a coding scheme inductively. Transcripts were coded in ATLAS.ti 8.0 (ATLAS.ti Scientific Software Development GmbH, Berlin, Germany). After coding, segments of text were abstracted by code, reviewed for themes, and summarized.

Textbox 1. Examples of clinician and patient questions.

Clinician Questions

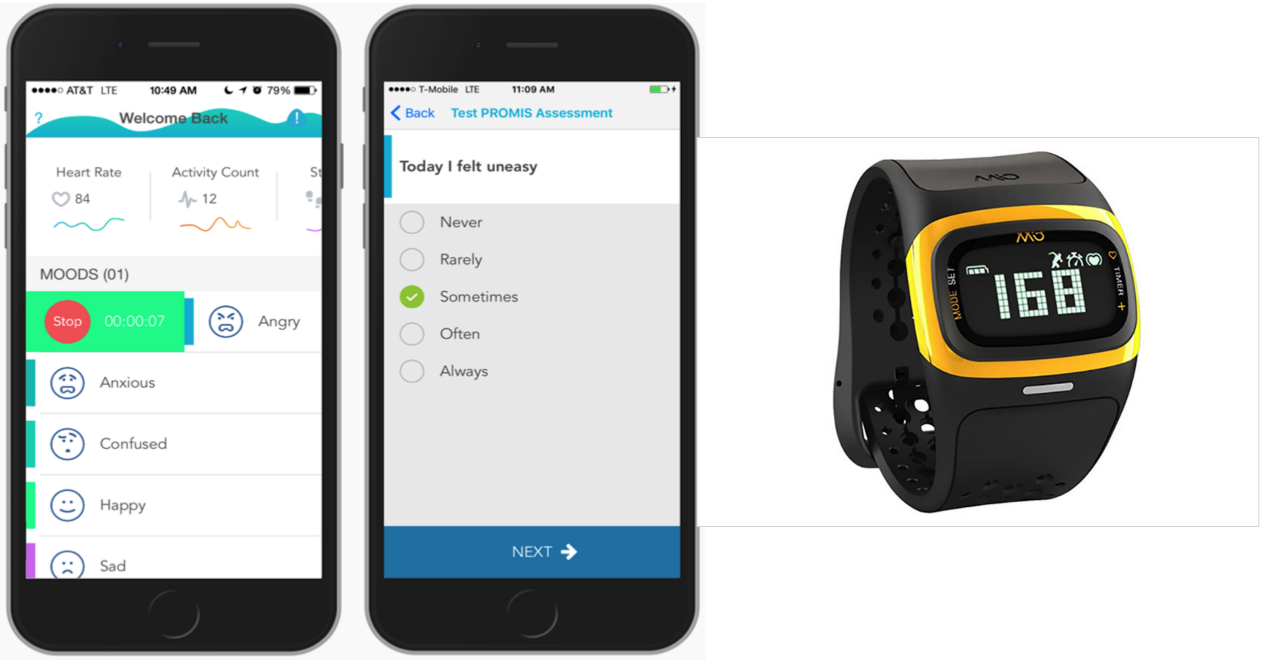
- What would you want this system to measure and how?
- Which patient populations might benefit the most from Capturing and Analyzing Sensor and Self-Report Data for Clinicians and Researchers (COMPASS)?
- What do you perceive as the benefits of COMPASS?
- How would you like collected data to be presented or reported?

Patient Questions

- What was your experience of wearing the device?
- What is the frequency of conversations with your care team and what do you discuss?
- Some people find it difficult to keep track of certain things about their health in order to talk about them at a doctor's visit. Do you ever find that's true for you? What would make that easier?
- Is there anything you would not be comfortable with your medical team monitoring?

XSL•FO
RenderX

Figure 1. Screenshots of smartphone app (QMedic Health, Inc) and Mio ALPHA 2 (Mio Global) wristwatch device.



Results

Clinician Interviews

We enrolled 10 of 17 clinicians who were invited to participate (59% participation rate), including 5 physicians, 3 nurse practitioners, 1 clinical psychologist, and 1 physical therapist. Table 1 presents the details on clinicians' characteristics and responses by specialty area and COMPASS features.

Priority Areas for COMPASS to Assess

Figures 2-4 present frequency distributions for content to assess outside of clinic visits. The objective measures that were deemed most valuable and relevant for informing clinical care were general vital statistics (n=7), heart rate (n=6), weight or body mass index (n=5), caloric expenditure (n=5), and glucose or electrolyte monitoring (n=4). When clinicians were asked about the use of a global positioning system or a monitoring system to track movements inside and outside the home, clinicians reported concern that patients would experience this as an invasion of privacy, and global positioning system data would not necessarily yield actionable results. However, data on general physical activity obtained from accelerometers or pedometers were considered potentially more useful. Domains that could be measured using PROs that clinicians were most interested in routinely assessing included psychological well-being (n=8), anxiety (n=7), mood, depression, and stress (n=6), pain and neuropathy (n=5), medication tolerance (n=4), and nausea (n=2). Figure 3 reflects an interest in types of data that could be captured by a combination of both wearable sensors and PROs. Functional status (including physical activity) and fatigue were mentioned by most clinicians (n=9), with sleep (n=8), gait (n=7), and adherence to treatment in general (n=7) the next most commonly reported. Adherence to medication (n=6) and lymphedema or swelling (n=5) was identified by roughly half of the clinician respondents.

Notably, an overall reflection from most clinicians was the need for data collected to be actionable. The following five main themes emerged representing data that were useful and actionable: (1) *Functional status (physical and cognitive)*, including information about daily physical activity, mobility constraints (ability to walk, balance, and engage in activities of daily living), weakness, cognitive abilities (forgetfulness), sleep (sleep-wake cycles and naps) and capacity for independent self-care; (2) *Indicators of disease progression (ie, health deterioration)*, including falls, seizures, and declining level of physical activity; (3) *Symptoms from disease (particularly disease progression) or treatment*, including side effects such as headaches, nausea, and increased pain from rehabilitation; (4) *Psychological well-being*, including anxiety, depression, and fear of recurrence; and (5) *Adherence to treatment and health behaviors*, including medication adherence, alcohol reduction, smoking cessation, and engaging in physical activity. In addition, clinicians discussed *test and imaging results* as important for evaluating response to treatment and determining ongoing treatment plans.

Patients Most Appropriate for COMPASS

Clinicians' perspectives on patients who could benefit the most from COMPASS differed by specialty. For example, all 3 rehabilitation clinicians felt that patients with functional limitations and who were at highest risk for events such as falls would benefit most from a system like COMPASS. For example, patients being treated for brain and spinal cord tumors were considered high risk for falls because of significant functional limitations secondary to disease. Other distinct groups that rehabilitation specialists felt could benefit from COMPASS were survivors, no longer in active treatment and who no longer were actively engaged in health care services, patients with pain (focusing on these patients during treatment could help to offset problems down the line), and older, overweight, and sedentary patients who typically have more comorbidities (eg, diabetes),

XSL•FO
RenderX

and who were, therefore, also at greater risk for falls and frailty-related declines in function. Conversely, other clinicians indicated that younger patients who were more impacted by their diagnosis and who would also spend longer in survivorship were more likely to benefit from a system like COMPASS.

In addition, clinicians identified a separate patient group at risk and who could benefit from COMPASS—elderly men with a substance abuse history and poor social support. Other psychosocial concerns were for patient groups like breast cancer survivors with large treatment burden that could lead to sleep issues and depression. Three clinicians specifically talked about patients (and caregivers) who experienced high anxiety being able to communicate regularly via COMPASS. Finally, patients who had difficulty with treatment adherence could benefit from a system that could incorporate reminders, for example, those who had complicated treatment regimens and who could, therefore, be helped with structured reminders and regular check-ins.

Summarizing Patient Data Collected Through COMPASS

Table 1 reports how clinicians of different specialties felt data should be presented and which platforms would make this most accessible. Overall, 5 clinicians wanted to see charts and graphs supported by qualitative written information to aid in interpretation, while other respondents wanted only charts and graphs or only written information. Owing to concerns about data overload, it was suggested that a summary of data with an option to expand to a more detailed view would offer the best usability. Data should only be provided when something abnormal was indicated.

Table 1. Clinicians' characteristics and responses by specialty area and Capturing and Analyzing Sensor and Self-Report Data for Clinicians and Researchers (COMPASS) features.

Specialty area	Main concerns for patients	Uses for COMPASS	Format of data presented	Platform for viewing data	Sharing patient data
Neuro-oncology					
Physician 1	Current rehab activity or need for referrals	Triggering alerts	Graphs or charts and qualitative data	Desktop personal computer (PC)	No
Physician 2	Monitoring adverse effects of medications or interventions	Triggering alerts and triggering electronic interventions (eInterventions; eg, reminders)	Graphs or charts and qualitative data	Desktop PC	Yes
Nurse practitioner	Current rehab activity or need for referrals and monitoring adverse effects of medications or interventions	Triggering alerts and eInterventions	Graphs or charts and qualitative data	Desktop PC	Yes
Rehabilitation					
Physiatrist 1	Monitoring adverse effects of medications, tracking physical activity, and identifying comorbidity	Triggering eInterventions	Graphs or charts and qualitative data	Desktop PC and laptop	Yes
Physiatrist 2	Nothing mentioned	Triggering alerts and eInterventions	Graphs or charts only	Laptop and tablet	Yes
Physical therapist	Patient vitals for the safety of exercise	Triggering alerts and eInterventions and data summary at the point of care	Graphs or charts and qualitative data	No preference	Yes
Cancer survivorship					
Physician (gastrointestinal cancers)	Medication adherence	Data summary at the point of care	Graphs or charts	Device agnostic (all)	Yes
Nurse practitioner (breast cancer)	Monitoring adverse effects of medications, identifying late effects, and addressing nutrition concerns	Triggering alerts	Graphs or charts and qualitative data	Desktop PC	Yes
Surgical oncology nurse practitioner (gastrointestinal cancers)	Determining the need for referrals	Triggering alerts	Graphs or charts	Desktop PC	Yes
Supportive oncology clinical psychologist (general oncology, head and neck cancer)	Tracking lifestyle behaviors (smoking, alcohol, and physical activity)	Triggering alerts and eInterventions and patient networking	Graphs or charts	No preference	Yes

XSL•FO
RenderX

Figure 2. Objective measures. GPS: global positioning system; BMI: body mass index.

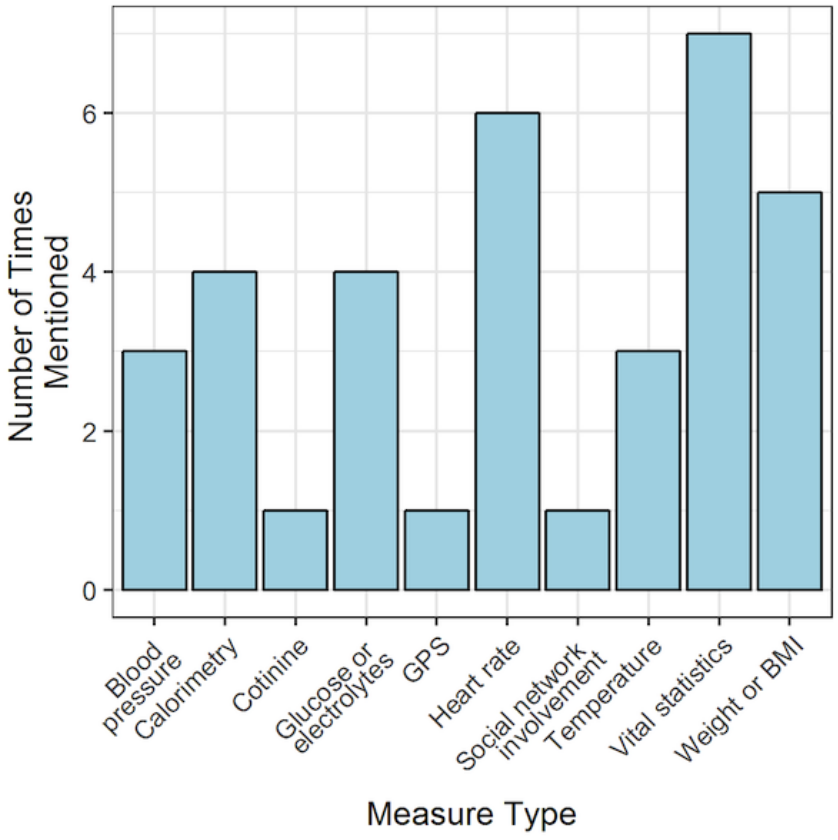
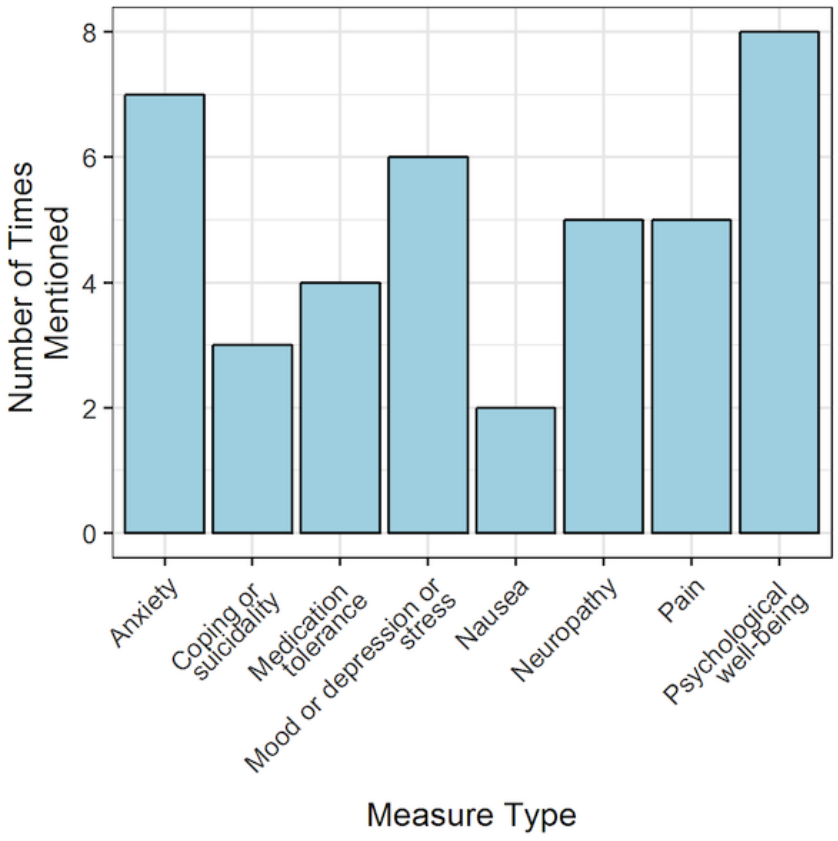
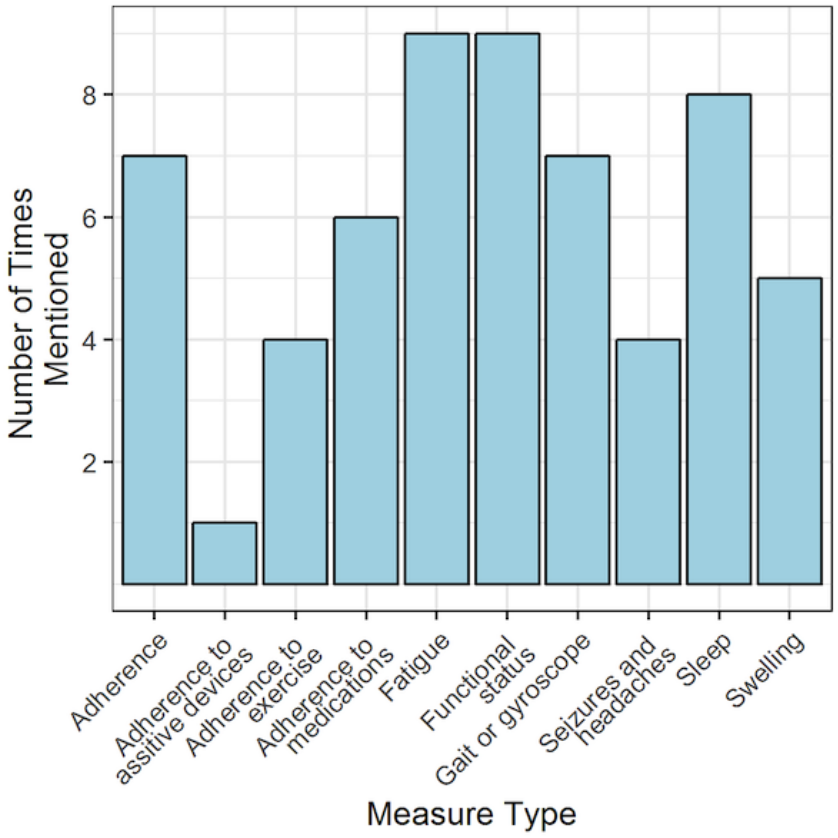


Figure 3. Patient-reported outcomes.



XSL•FO
RenderX

Figure 4. Content measured by both objective sensors and patient-reported outcomes.



Five clinicians preferred to view data on a desktop computer as they reportedly felt it was more secure and had a large area for viewing data; 4 of 10 reported they were comfortable on all platforms. However, some clinicians (n=3) did not like to look at important data on the phone, and 2 clinicians expressed concerns that having data delivered to their personal phones might remove an important barrier between themselves and their patients. All but 1 of the clinicians were interested in the data being sharable between those who were providing care, reflecting that this may be the most useful part of the entire system, especially because cancer care is so multidisciplinary. Conversely, another clinician expressed that time is a valuable and limited resource and having a care plan where one person brings the data together to decide a course of action may be better than the raw data being viewed by multiple people, all of whom are reviewing and, perhaps, deciding a duplicative plan.

Sharing of Data With Patients

A subset of 6 clinicians was asked whether they thought patients should be provided with their own data. Most clinicians (n=5) were in favor of seeing a data summary being provided. Reasons given were that it would help to engage patients with the technology and, perhaps, lead to better long-term use and could show patients their patterns. It was suggested by one clinician, who was in favor of data sharing, that, perhaps, some data should be presented in the clinic rather than via COMPASS, as this could allow the clinician to frame the information appropriately. For example, data related to imaging or other clinical interpretation should not be available outside the clinic where it could not be explained properly, and patients may become more anxious if they do not understand the data. In this study, 9 clinicians were asked about the utility of patients being able to annotate their data to provide context, and all agreed this would be valuable; however, they would still call or contact patients for clarification anyway. Furthermore, 3 of the clinicians felt that while valuable, this should only be optional and not required as it may increase the patients' burden.

Patient Interviews

We enrolled 3 patients, 2 of whom were diagnosed with brain cancer (one receiving active treatment and one 2 year posttreatment) and 1 had a previous diagnosis of lung cancer and was also 2-year posttreatment (Table 2). Owing to the variety of recruitment methods utilized, including social media, flyers in clinics, and on RHLCCC notice boards, we could not track a participation rate for patients.

Wearability

Patients wore the Mio wristwatch devices for an average of 7 days prior to their interviews. Patients found the watch-like style of the device acceptable as it was generally comfortable, as long as it was not worn too tightly. However, 1 of 3 patients seemed somewhat uncomfortable. When asked if potentially wearing a tracking device like the Mio on a chain around the neck so it was close to the chest rather than on the wrist or whether they would prefer to wear the smartphone on an arm strap, 2 of 3 patients commented that these options were not favored. Patient 3 specifically responded that he or she would not want to wear a cell phone while running saying "I think it would be difficult, so no." While not considered a particular theme, 1 patient also made comments relating to specific

XSL•FO
RenderX

concerns about privacy, professional life, and intimacy. As part of wearability, patients also specifically noted the battery life of the wrist device. While Patient 1 was able to charge the device and iPhone nightly, Patients 2 and 3 had some trouble. Both patients reported a short battery life that was inconvenient.

Data Capture

All patients reported times when they were unsure their data were being captured; this was either because the device was not positioned correctly or because the device and cell phone were not communicating. In addition, patients were asked to enter basic data on the dietary intake, mood, and activity. A small number of options were available in each category and patients reported that this task was easy.

Tracking of Symptoms and Other Items

The tracking of symptoms during treatment and recovery was a particular area of interest for patients. Patients agreed that tracking a symptom was helpful; however, the symptoms experienced and preferences for tracking these symptoms varied by patient. Table 3 shows the kinds of symptoms that patients reported experiencing during their treatment, and Table 4 shows that in addition to tracking symptoms, patients also felt it would be useful to track other items.

All patients felt it would be useful to track their moods or feelings on an app during treatment to discuss with providers. One patient, in particular, felt this was important with regards to medications that may need adjustment. In addition, all patients reflected that having the ability to add freeform text to responses they had provided was important for linking activities or events with mood, changes in heart rate or stress, providing context to their data.

Suggestions for Further Development of the System

Finally, patients reflected about preferences and suggestions for improvements for COMPASS; these included suggestions for improving the PRO descriptions and adding functionality to the surveys depending on responses given. For example, the device could provide guided relaxation exercises and imagery if they reported high levels of anxiety. The contact frequency should also be limited. One patient stated, “15 notifications in a row is unnecessary.” All patients wanted to be able to review their personal data, as they felt this would be “motivating.” In addition, the system could provide better support for reminders, including for taking medicines, filling prescriptions, and upcoming appointments. There were mixed responses to questions about the capacity of the device to contact significant others if there were concerns about their health.

Table 2. Patient characteristics and device preferences.

Characteristic	Patient 1	Patient 2	Patient 3
Cancer site	Brain	Brain	Lung
Treatment status	Current chemotherapy	2-year posttreatment	2-year posttreatment
Platform or device currently using	iPhone, laptop	Android	Personal computer, Mac laptop, Android
Platform or device preferred	iPad	Desktop and mobile	iPhone

Table 3. Patients’ symptoms.

Symptoms	Patient 1	Patient 2	Patient 3
Anxiety	✓		✓
Confusion	✓		
Constipation	✓		
Feeling sick		✓	
Fevers			✓
Headaches		✓	
Memory loss	✓		
Neuropathy	✓		
Pain	✓		✓
Respiratory infection			✓
Seizures or auras		✓	
Vision loss (peripheral)	✓		
Weakness	✓		
Weight loss	✓		

XSL•FO
RenderX

Table 4. Items to track.

Item to Track	Patient 1	Patient 2	Patient 3
Appointments		✓	
Confusion	✓		
Dietary intake or nutrition	✓	✓	
Exercise or activity, including heart rate and steps	✓	✓	✓
Fatigue or exhaustion	✓		
Fevers			✓
Headaches	✓	✓	
Medication use or prescriptions		✓	✓
Mood, including anxiety	✓	✓	✓
Pain	✓		✓
Progress	✓		
Seizures or auras		✓	
Weight		✓	

Design Impacts of User Needs Assessment

The design of the prototype system represents an attempt to accommodate the most frequently requested features from the clinician assessment while minimizing the user burden on the patient population. Because a goal of this phase I study was to evaluate the feasibility of a system that integrates across multiple devices and data sources, efforts were made to realistically reflect the type of activities required to operate and maintain such a system. Specifically, the type and position of the wearable device were necessitated by technical constraints related to physiological sensing and capturing physical activity. Similar considerations were taken to provide a user interface on the mobile phone that accurately reflects that number, type, and duration of interactions required to collect the types of PROs most frequently requested by clinicians. Subsequent design iterations over both the clinician and patient interfaces should incorporate feedback on utility and usability from participants until an optimal balance between these objectives is achieved.

Discussion

Principal Findings

This study yields important insights regarding the initial feasibility and priority domains to inform the development of COMPASS, an eHealth platform designed to facilitate the patient-provider communication and improve supportive care outcomes in the cancer care setting. Clinicians were most interested in measuring and monitoring general vital statistics (heart rate, body weight, caloric expenditure, and glucose levels), functional status, symptoms (mood, depression, anxiety, and pain), and medication adherence. Importantly, measures needed to be actionable and integrate both objective metrics and PROs together to provide the richest and most clinically relevant understanding of the patients' status. This was echoed by patients, who also wanted to be able to provide context to their data and responses. Patients were most interested in monitoring their symptoms, including pain, headaches, mood, and anxiety. In addition, they wanted to be able to track physical activity, diet or nutrition, medications, and appointments. Overall, patients who may benefit the most from a system like COMPASS depended on the clinicians' specialty. Particular groups mentioned were the elderly, those with comorbidities such as diabetes, and those with complicated treatment regimens (oral chemotherapy).

The results of this study indicate that both clinicians and patients felt that a system like COMPASS had potential benefits for the delivery of cancer care. Given the shared interest and importance of symptom monitoring by both clinicians and patients, finding ways to improve this aspect of patient care seems critical. Recent studies have suggested that oncologists are often not sufficiently aware of their patients' symptoms [31,32] but that when prompted, patients willingly provide this information [33]. However, integrating such self-report systems into routine clinical care with minimal disruption is of key importance [11], which is where eHealth-based platforms may offer significant potential. A recent qualitative needs assessment among 30 head and neck and breast cancer survivors supported these findings, revealing that survivors often felt their symptoms remained unknown to care providers [34]. In addition, they reported that the advantage of an eHealth app would be that monitoring could provide insight into the course of symptoms, providing information for follow-up visits and receiving personalized advice and tailored supportive care.

Both clinicians and patients were interested in tracking mood, anxiety, and lifestyle behaviors, such as physical activity and diet or nutrition, as areas of shared importance. This may be related to the growing popularity of fitness trackers and health apps in general; nonetheless, the monitoring of physical activity and lifestyle behaviors is of clinical importance, as it strongly reflects the functional status. Furthermore, the fact that patients are interested in monitoring these aspects of their function points to a higher likelihood of adherence when asked to monitor their behaviors. A previously conducted randomized clinical trial tested a system similar to COMPASS in patients with type II

XSL•FO
RenderX

diabetes; in that study, the intervention group supported with 24-hour access to mobile health coaching, monitoring, and communication obtained better disease control in half the time that it took the comparison group [35]. This finding highlights the potential benefits of more regular contact that can be facilitated with a system such as COMPASS. The growing interest in eHealth-based approaches in the clinical care setting seemingly reflects the benefits that such systems can provide for both patients and providers.

Other themes representing actionable data included the monitoring of the functional status and the ability of clinicians to communicate the results of tests and imaging with their patients more readily—something that really tapped into the role that innovative systems such as COMPASS could play in ongoing patient care. However, as was mentioned several times, too much data of no relevance to the clinical care of patients would potentially create more of a burden than a benefit. Kuijpers et al reached a similar conclusion when evaluating a Web-based intervention focused on patient care for lung and breast cancer survivors in the Netherlands [36]. While health care professionals supported access to the electronic medical record for providing reports and results, they also expected it would lead to the increased workload because patients were unlikely to understand the information provided, prompting greater burden on clinicians to follow-up. This has important implications for the development and potential of similar eHealth-based approaches as the goal is to provide improved patient care, but not lead to a greater workload, which is already a burden on the system [27].

Limitations

Because this study was designed to assess user needs relating to an eHealth-based system for monitoring and facilitating communication between patients and clinicians, we could not test the prototype with a large number of patients. The study included only 3 patients with 2 types of diagnoses compared with what would typically be seen in the clinical setting. Therefore, a primary limitation of this study was the small number of patient interviews conducted in phase I of the study, which may somewhat limit any generalizability of our findings to other patient subsets. To address this limitation and facilitate a continuing update of system features and further refine the technology, patients who are involved in phase II will provide regular feedback on usability and features, which will be incorporated with those provided in phase I. Evidence suggests that between 6 and 12 participants can provide adequate data for determining meta-themes [37]. Given the fast pace of obsolescence in technology settings, the priority of phase I of the study was to develop a prototype of the device-agnostic platform that incorporated evidence-based features identified by clinicians and that could be more rigorously tested in phase II. Another limitation is that while the device was given to participants who had previous experience with a smartphone for logistical purposes, this may not reflect well on the ease of adoption for a smartphone-naïve individual.

Strengths

It is critical that people engaging with the technology—in this case, clinicians, patients, and researchers—have input as to the design of the tools and content they will be using. A significant strength of this study is that we used an iterative process that involved input from researchers, engineers, clinicians, and patients in identifying the aspects of a mobile and eHealth-based platform that would be most effective for monitoring, capturing, and reporting of relevant data related to patient care in the time between standard clinic visits. Other significant strengths of the study were as follows: (1) the use of qualitative methods that allowed us to gather and synthesize provider and patient perspectives “in their own words” and (2) the broad cross-section of clinicians that provided a detailed description of the types of metrics deemed most important and suitable for monitoring and the types of patients who could benefit most from a system like COMPASS.

Conclusions

Technology is being increasingly integrated into the care of cancer patients and survivors, who as a group require significant resources in terms of time and personal contact. One of the primary goals of technology-driven approaches is to help improve the efficiency of the current care delivery system while providing high-quality care. Technology needs to support the decision-making process of providers in an evidence-based manner and in a way that makes its use easy for patients to comply. It is critical to include the perspective of clinicians and their patients, as well as researchers and engineers, when designing such systems. Furthermore, to provide a continual refinement of such technologies, it is critical that ongoing feedback be sought from patients and clinicians using the system. Many practical and logistical issues may only arise after ongoing use and in response to changing environmental conditions. Therefore, a degree of flexibility in design iteration is preferable. However, the other side of this argument is that asking patients to utilize systems that are burdensome will lead to low compliance while providing clinicians with vast amounts of unusable data will lead to similar results. An ideal system would streamline patient-provider interactions while also highlighting clinically relevant domains that are important to clinicians and ensure patients most salient concerns are addressed. Future research needs to determine whether such systems can be integrated into current practice settings and whether there are, in fact, improvements in care and outcomes for a variety of different cancer patients and survivors.

Acknowledgments

This research was supported by a Small Business and Innovation Research grant from the National Cancer Institute (261201300055C) and was partly supported by a National Cancer Institute training grant (R25 CA122061).

XSL•FO
RenderX

Conflicts of Interest

FA is cofounder and Chief Technology Officer of QMedic Health, Inc, which develops and manufactures older adult activity monitoring and safety wearable devices. QMedic Health may or may not directly or indirectly profit from this work by employing techniques and methods to motivate and encourage older adults to wear the monitors. LIW received research funding and consulting fees from EveryFit, Inc.

Multimedia Appendix 1

Thought exercise.

[PDF File (Adobe PDF File), 17KB - medinform_v6i4e46_app1.pdf]

Multimedia Appendix 2

Clinician and patient interview quotes.

[PDF File (Adobe PDF File), 34KB - medinform_v6i4e46_app2.pdf]

Multimedia Appendix 3

Potential sensor list.

[PDF File (Adobe PDF File), 13KB - medinform_v6i4e46_app3.pdf]

References

1. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, et al. Cancer treatment and survivorship statistics, 2014. CA Cancer J Clin 2014;64(4):252-271 [FREE Full text] [doi: 10.3322/caac.21235] [Medline: 24890451]
2. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med 2012 Nov 22;367(21):1998-2005. [doi: 10.1056/NEJMoa1206809] [Medline: 23171096]
3. Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. JAMA 2009 Oct 21;302(15):1685-1692. [doi: 10.1001/jama.2009.1498] [Medline: 19843904]
4. Zeng C, Wen W, Morgans AK, Pao W, Shu X, Zheng W. Disparities by Race, Age, and Sex in the Improvement of Survival for Major Cancers: Results From the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) Program in the United States, 1990 to 2010. JAMA Oncol 2015 Apr;1(1):88-96 [FREE Full text] [doi: 10.1001/jamaoncol.2014.161] [Medline: 26182310]
5. de Moor JS, Mariotto AB, Parry C, Alfano CM, Padgett L, Kent EE, et al. Cancer survivors in the United States: prevalence across the survivorship trajectory and implications for care. Cancer Epidemiol Biomarkers Prev 2013 Apr;22(4):561-570 [FREE Full text] [doi: 10.1158/1055-9965.EPI-12-1356] [Medline: 23535024]
6. McCabe MS, Bhatia S, Oeffinger KC, Reaman GH, Tyne C, Wollins DS, et al. American Society of Clinical Oncology statement: achieving high-quality cancer survivorship care. J Clin Oncol 2013 Feb 10;31(5):631-640 [FREE Full text] [doi: 10.1200/JCO.2012.46.6854] [Medline: 23295805]
7. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Survivorship (NCCN Guidelines). 2017. URL: https://www.nccn.org/professionals/physician_gls/pdf/survivorship.pdf [accessed 2018-09-20] [WebCite Cache ID 72Zl4AlYR]
8. Leach CR, Weaver KE, Aziz NM, Alfano CM, Bellizzi KM, Kent EE, et al. The complex health profile of long-term cancer survivors: prevalence and predictors of comorbid conditions. J Cancer Surviv 2015 Jun;9(2):239-251. [doi: 10.1007/s11764-014-0403-1] [Medline: 25319681]
9. Stein KD, Syrjala KL, Andrykowski MA. Physical and psychological long-term and late effects of cancer. Cancer 2008 Jun 1;112(11 Suppl):2577-2592 [FREE Full text] [doi: 10.1002/cncr.23448] [Medline: 18428205]
10. Henry DH, Viswanathan HN, Elkin EP, Traina S, Wade S, Cella D. Symptoms and treatment burden associated with cancer treatment: results from a cross-sectional national survey in the U.S. Support Care Cancer 2008 Jul;16(7):791-801. [doi: 10.1007/s00520-007-0380-2] [Medline: 18204940]
11. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
12. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014 May 10;32(14):1480-1501. [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]

XSL•FO
RenderX

13. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017 Jun 04. [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
14. Jensen RE, Snyder CF, Abernethy AP, Basch E, Potosky AL, Roberts AC, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. J Oncol Pract 2014 Jul;10(4):e215-e222 [FREE Full text] [doi: 10.1200/JOP.2013.001067] [Medline: 24301843]
15. Basch E. Missing Patients' Symptoms in Cancer Care Delivery--The Importance of Patient-Reported Outcomes. JAMA Oncol 2016 Apr;2(4):433-434. [doi: 10.1001/jamaoncol.2015.4719] [Medline: 26720842]
16. Basch E, Abernethy AP. Supporting clinical practice decisions with real-time patient-reported outcomes. J Clin Oncol 2011 Mar 10;29(8):954-956. [doi: 10.1200/JCO.2010.33.2668] [Medline: 21282536]
17. Wagner LI, Schink J, Bass M, Patel S, Diaz MV, Rothrock N, et al. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. Cancer 2015 Mar 15;121(6):927-934 [FREE Full text] [doi: 10.1002/cncr.29104] [Medline: 25376427]
18. Foulon V, Schöffski P, Wolter P. Patient adherence to oral anticancer drugs: an emerging issue in modern oncology. Acta Clin Belg 2011;66(2):85-96. [doi: 10.2143/ACB.66.2.2062525] [Medline: 21630604]
19. Pachman DR, Price KA, Carey EC. Nonpharmacologic approach to fatigue in patients with cancer. Cancer J 2014;20(5):313-318. [doi: 10.1097/PPO.0000000000000064] [Medline: 25299140]
20. Mishra SI, Scherer RW, Geigle PM, Berlanstein DR, Topaloglu O, Gotay CC, et al. Exercise interventions on health-related quality of life for cancer survivors. Cochrane Database Syst Rev 2012;8:CD007566. [doi: 10.1002/14651858.CD007566.pub2] [Medline: 22895961]
21. Betof AS, Dewhirst MW, Jones LW. Effects and potential mechanisms of exercise training on cancer progression: a translational perspective. Brain Behav Immun 2013 Mar;30 Suppl:S75-S87 [FREE Full text] [doi: 10.1016/j.bbi.2012.05.001] [Medline: 22610066]
22. Morey MC, Snyder DC, Sloane R, Cohen HJ, Peterson B, Hartman TJ, et al. Effects of home-based diet and exercise on functional outcomes among older, overweight long-term cancer survivors: RENEW: a randomized controlled trial. JAMA 2009 May 13;301(18):1883-1891 [FREE Full text] [doi: 10.1001/jama.2009.643] [Medline: 19436015]
23. Fong DYT, Ho JWC, Hui BPH, Lee AM, Macfarlane DJ, Leung SSK, et al. Physical activity for cancer survivors: meta-analysis of randomised controlled trials. BMJ 2012;344:e70 [FREE Full text] [Medline: 22294757]
24. Ballard-Barbash R, Friedenreich CM, Courneya KS, Siddiqi SM, McTiernan A, Alfano CM. Physical activity, biomarkers, and disease outcomes in cancer survivors: a systematic review. J Natl Cancer Inst 2012 Jun 06;104(11):815-840 [FREE Full text] [doi: 10.1093/jnci/djs207] [Medline: 22570317]
25. Banegas MP, Guy GP, de Moor JS, Ekwueme DU, Virgo KS, Kent EE, et al. For Working-Age Cancer Survivors, Medical Debt And Bankruptcy Create Financial Hardships. Health Aff (Millwood) 2016 Jan;35(1):54-61. [doi: 10.1377/hlthaff.2015.0830] [Medline: 26733701]
26. Guy GP, Ekwueme DU, Yabroff KR, Dowling EC, Li C, Rodriguez JL, et al. Economic burden of cancer survivorship among adults in the United States. J Clin Oncol 2013 Oct 20;31(30):3749-3757 [FREE Full text] [doi: 10.1200/JCO.2013.49.1241] [Medline: 24043731]
27. American Society of Clinical Oncology. The State of Cancer Care in America, 2016: A Report by the American Society of Clinical Oncology. J Oncol Pract 2016 Apr;12(4):339-383 [FREE Full text] [doi: 10.1200/JOP.2015.010462] [Medline: 26979926]
28. Neuendorf KA. The content analysis guidebook. Thousand Oaks, CA: Sage Publications; 2002.
29. Marshall C, Rossman GB. Designing qualitative research. Thousand Oakes, CA: Sage Publications; 1999.
30. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
31. Detmar SB, Aaronson NK, Wever LD, Muller M, Schornagel JH. How are you feeling? Who wants to know? Patients' and oncologists' preferences for discussing health-related quality-of-life issues. J Clin Oncol 2000 Sep 15;18(18):3295-3301. [doi: 10.1200/JCO.2000.18.18.3295] [Medline: 10986063]
32. Fromme EK, Eilers KM, Mori M, Hsieh Y, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol 2004 Sep 1;22(17):3485-3490. [doi: 10.1200/JCO.2004.03.025] [Medline: 15337796]
33. Montemurro F, Mittica G, Cagnazzo C, Longo V, Berchialla P, Solinas G, et al. Self-evaluation of Adjuvant Chemotherapy-Related Adverse Effects by Patients With Breast Cancer. JAMA Oncol 2016 Apr;2(4):445-452. [doi: 10.1001/jamaoncol.2015.4720] [Medline: 26720497]
34. Lubberding S, van Uden-Kraan CF, Te Velde EA, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. J Clin Nurs 2015 May;24(9-10):1367-1379. [doi: 10.1111/jocn.12753] [Medline: 25677218]
35. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health Coaching Reduces HbA1c in Type 2 Diabetic Patients From a Lower-Socioeconomic Status Community: A Randomized Controlled Trial. J Med Internet Res 2015;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]

XSL•FO
RenderX

36. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Development of MijnAVL, an Interactive Portal to Empower Breast and Lung Cancer Survivors: An Iterative, Multi-Stakeholder Approach. JMIR Res Protoc 2015;4(1):e14 [FREE Full text] [doi: 10.2196/resprot.3796] [Medline: 25614924]
37. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]

Abbreviations

COMPASS: Capturing and Analyzing Sensor and Self-Report Data for Clinicians and Researchers
eHealth: electronic health
PC: personal computer
PRO: patient-reported outcome
RHLCCC: Robert H Lurie Comprehensive Cancer Center
TCA: thematic content analysis

Edited by G Eysenbach; submitted 29.11.17; peer-reviewed by W Groen, J Guattery, R Pankomera; comments to author 15.03.18; revised version received 03.05.18; accepted 21.06.18; published 22.10.18.

Please cite as:
Lucas AR, Bass MB, Rothrock NE, O'Connor ML, Sorkin MR, Nawyn J, Albinali F, Wagner LI
Development of an eHealth System to Capture and Analyze Patient Sensor and Self-Report Data: Mixed-Methods Assessment of Potential Applications to Improve Cancer Care Delivery
JMIR Med Inform 2018;6(4):e46
URL: http://medinform.jmir.org/2018/4/e46/
doi:10.2196/medinform.9525
PMID:30348634

©Alexander R Lucas, Michael B Bass, Nan E Rothrock, Mary L O'Connor, Mia R Sorkin, Jason Nawyn, Fahd Albinali, Lynne I Wagner. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 22.10.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Identifying Principles for the Construction of an Ontology-Based Knowledge Base: A Case Study Approach

Xia Jing[1], MD, PhD; Nicholas R Hardiker[2], RN, PhD, FACMI; Stephen Kay[3], PhD, FACMI; Yongsheng Gao[4], MD, PhD

[1]Department of Social and Public Health, College of Health Sciences and Professions, Ohio University, Athens, OH, United States

[2]School of Human and Health Sciences, University of Huddersfield, Huddersfield, United Kingdom

[3]THISC Ltd, Manchester, United Kingdom

[4]International Health Terminology Standards Development Organization, London, United Kingdom

Corresponding Author:
Xia Jing, MD, PhD
Department of Social and Public Health
College of Health Sciences and Professions
Ohio University
Grover Center W357
Athens, OH, 45701
United States
Phone: 1 7405930750
Fax: 1 7405930555
Email: xia.xjing@gmail.com

Abstract

Background: Ontologies are key enabling technologies for the Semantic Web. The Web Ontology Language (OWL) is a semantic markup language for publishing and sharing ontologies.

Objective: The supply of customizable, computable, and formally represented molecular genetics information and health information, via electronic health record (EHR) interfaces, can play a critical role in achieving precision medicine. In this study, we used cystic fibrosis as an example to build an Ontology-based Knowledge Base prototype on Cystic Fibrobis (OntoKBCF) to supply such information via an EHR prototype. In addition, we elaborate on the construction and representation principles, approaches, applications, and representation challenges that we faced in the construction of OntoKBCF. The principles and approaches can be referenced and applied in constructing other ontology-based domain knowledge bases.

Methods: First, we defined the scope of OntoKBCF according to possible clinical information needs about cystic fibrosis on both a molecular level and a clinical phenotype level. We then selected the knowledge sources to be represented in OntoKBCF. We utilized top-to-bottom content analysis and bottom-up construction to build OntoKBCF. Protégé-OWL was used to construct OntoKBCF. The construction principles included (1) to use existing basic terms as much as possible; (2) to use intersection and combination in representations; (3) to represent as many different types of facts as possible; and (4) to provide 2-5 examples for each type. HermiT 1.3.8.413 within Protégé-5.1.0 was used to check the consistency of OntoKBCF.

Results: OntoKBCF was constructed successfully, with the inclusion of 408 classes, 35 properties, and 113 equivalent classes. OntoKBCF includes both atomic concepts (such as amino acid) and complex concepts (such as “adolescent female cystic fibrosis patient”) and their descriptions. We demonstrated that OntoKBCF could make customizable molecular and health information available automatically and usable via an EHR prototype. The main challenges include the provision of a more comprehensive account of different patient groups as well as the representation of uncertain knowledge, ambiguous concepts, and negative statements and more complicated and detailed molecular mechanisms or pathway information about cystic fibrosis.

Conclusions: Although cystic fibrosis is just one example, based on the current structure of OntoKBCF, it should be relatively straightforward to extend the prototype to cover different topics. Moreover, the principles underpinning its development could be reused for building alternative human monogenetic diseases knowledge bases.

(JMIR Med Inform 2018;6(4):e52) doi:10.2196/medinform.9979

XSL•FO
RenderX

KEYWORDS

cystic fibrosis; knowledge base; knowledge representation; molecular genetics information; ontology; OntoKBCF; phenotypes

Introduction

"The Semantic Web is a vision for the future of the Web, in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the Web" [1]. The major objectives of the Semantic Web are to improve data reuse, data sharing, and data integration in the data-centric Web era [2]. The World Wide Web, in contrast, is document-centric, whereby a document, not the data, is the basic processing unit. Therefore, the Semantic Web, owing to a finer granularity, has the potential to achieve a more precise result in its application.

Ontologies are key enabling technologies for the Semantic Web [3]. Ontologies play a key role in information use via knowledge sharing and reuse [4]. According to the Oxford Living Dictionary, an ontology is defined as "a set of concepts and categories in a subject area or domain that shows their properties and the relations between them." Another definition by Tom Gruber is "an ontology is a specification of conceptualization" [5]. In this paper, we use ontology to refer a collection of representations of the domain knowledge units (ie, entities and properties) and their relationships that can be utilized by humans, databases, and applications. The Web Ontology Language (OWL) is a semantic markup language for publishing and sharing data by ontologies on the Web [1,6]. OWL is intended to be used in information processing by machines, as well as understood by human beings [1,6]. Description logic can be subjected to automated reasoning to check the consistency of ontologies and determine subsumption relationships [6].

The specific application of ontologies includes standard conceptual vocabularies, services for queries, and reusable knowledge bases, all of which can facilitate interoperability across different systems [7]. Biomedicine is one of the active application domains of ontologies. The *motivation* of our work is to demonstrate the feasibility of providing formal and computable clinical information and molecular genetic information, derived from an ontology-based knowledge base, via an electronic health record (EHR) interface. Molecular genetic information plays a key role in medicine; this information, however, is not the kind that is traditionally included in an EHR system. Therefore, we aim to build an Ontology-based Knowledge Base prototype on Cystic Fibrosis (OntoKBCF) [8,9] to supply formal and machine-processable, clinically relevant molecular genetics information dynamically via an EHR prototype.

Cystic fibrosis is used as an example disease in building OntoKBCF for a feasibility demonstration. The reasons for this choice include the following: cystic fibrosis has a relatively stable and comprehensive molecular genetic profile; it is one of the most common, fatal genetic diseases in the United States [10], especially in white people of Northern European ancestry [11,12]; the mechanism of the disease has been well studied and well understood; and cystic fibrosis is a single gene disease, which is less complicated than multiple-gene diseases. The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene causes cystic fibrosis.

An ontology-based knowledge base is built on top of domain entities, properties, and relationships (ie, the ontology); however, it also includes sharable instances, rules, and inference capabilities [13], which extend the ontology. Therefore, an ontology-based knowledge base can be utilized via an EHR system, for example, by incorporating the characteristics of patients. In this paper, we use a *knowledge base model* to refer to the logic specifications of a knowledge base at the abstract level and a *knowledge base prototype* (ie, OntoKBCF) to refer to the physical artifact that we eventually built at the concrete level, based on a knowledge base model. We use a *knowledge base prototype* instead of a *knowledge base* because we include different levels of facts in OntoKBCF, for example, nucleotide changes, amino acid changes, and clinical phenotypes. The facts are not exhaustive at any horizontal level within OntoKBCF; however, the facts are sufficient to organize the relationships at cross-horizontal levels [8,9] and to demonstrate how to generate a knowledge resource including molecular genetics information and health information via an EHR prototype [14,15].

An overview of the project, showing how OntoKBCF supplies computable and formally represented molecular and health information to an EHR in a dynamic manner has been published previously [14], as has the technical integration of OntoKBCF and the EHR prototype [15]. A further related source [8], published at an early stage of the project, mainly describes the *domain coverage* of OntoKBCF. In contrast, this paper provides *a holistic view of the construction principles* of OntoKBCF, including the *representations of the integration* between OntoKBCF and an EHR. As such, this paper may provide a reference for content developers operating in the Semantic Web space, especially in building knowledge bases and knowledge base modeling on monogenetic human diseases.

Methods

Representation Language and Terms Used in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

OntoKBCF was built using Protégé-OWL [16]. We chose OWL description logic (DL) [1] as the language for OntoKBCF. We used basic, combined, and complex concepts, as well as final facts, to represent selected content within OntoKBCF. Table 1 provides definitions for the major terms that we have used in the remainder of the paper, with examples. Figure 1 shows the relationships between the terms and their relative volumes of each category within OntoKBCF.

XSL•FO
RenderX

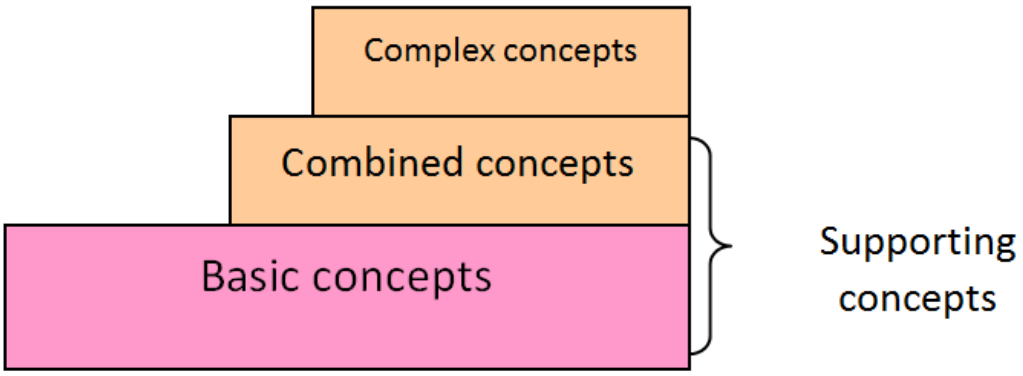
Table 1. Major construction terms used in the paper.

Term name	Annotation and example
Basic concept	The atomic concept from Unified Medical Language System [17], Gene Ontology [18], or domain knowledge, such as Gly, which is the amino acid glycine.
Combined concept	Combination of two or three basic concepts, such as Gly542, which is Gly's location, 542 in the cystic fibrosis transmembrane conductance regulator amino acid chain.
Complex concept	Combinations of more than three basic concepts and usually they are subjects. A subset of the combined concepts is explained through the EHR[a] interface, such as Patient_CF_with_Gly542X, which is a group of cystic fibrosis patients with the Gly542X variant.
Supporting concept	Basic concepts and combined concepts, used in explaining the complex concept.
Final fact	A domain statement represented in OntoKBCF[b] through a combination of basic concepts or combined concepts with properties and logic relationships, such as property and description of Patient_CF_with_Gly542X.
Fact	Any concept above is a subset of facts, which includes (1) hierarchy of concepts, both basic and combined concepts, and (2) a property description of the concept.

[a]EHR: electronic health record.

[b]OntoKBCF: Ontology-based Knowledge Base prototype on Cystic Fibrosis.

Figure 1. Relationships between construction terms used in this paper and their relative sizes.



Organization and Main Construction Procedures of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

The main axes used in OntoKBCF are time and problem in line with the content of the included domain. The main axes are utilized to organize the atomic, combined, and complex concepts in OntoKBCF. Figure 2 shows the main construction processes for OntoKBCF.

Scope of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis and Sources

Figure 3 presents the main content of OntoKBCF. One of the purposes of OntoKBCF is for clinical use; therefore, the following factors were considered when determining domain content: (1) treatment or therapy, which has been reported as the main category of information needs for clinicians, including gene therapy and other treatments for cystic fibrosis [19-25]; (2) availability of relevant molecular genetic information, such as the most common *CFTR* mutations, particularly those related to health information or phenotypes (eg, symptoms, ethnic groups) [26-28]; (3) cystic fibrosis has a time-related development cycle, and usually, a patient's age is available in an EHR; thus, time-oriented descriptions of cystic fibrosis are included in OntoKBCF [21,23-25].

Top-to-Bottom Content Analysis in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

The selected content required analysis to set up the construction strategies. Based on the scope and the knowledge resources available, analyzing the facts by utilizing a series of dissections was necessary to make the task realistic and practically accomplishable. The starting point was sentences of facts that were dissected into atomic units (ie, basic concepts in Figure 1). The atomic units were represented with their necessary hierarchies, they were combined, and their properties included to express the complex facts in a logical, machine-processable format. Usually, complex facts can be represented by one or multiple concepts, described in Figure 1, and their properties.

The general criteria of the dissection were determined by the scope and granularity of OntoKBCF; otherwise, dissection would be an endless process. Only necessary superclass concepts, rather than detailed and complete subsets of Unified Medical Language System (UMLS) [17] and Gene Ontology (GO) [18], have been included in OntoKBCF. Only the properties and constraints within the scope of OntoKBCF, rather

XSL•FO
RenderX

than a complete and comprehensive description for each class, are represented. The main reason for doing so is to ensure that the tasks are accomplishable with limited resources and within a restricted timeline. The dissection process was followed by reconstruction. Figure 4 shows the analysis and reconstruction order that we utilized in constructing OntoKBCF.

Figure 2. General construction procedures of the ontology-based knowledge base prototype on cystic fibrosis. KB: knowledge base; UMLS: Unified Medical Language System; GO: gene ontology.

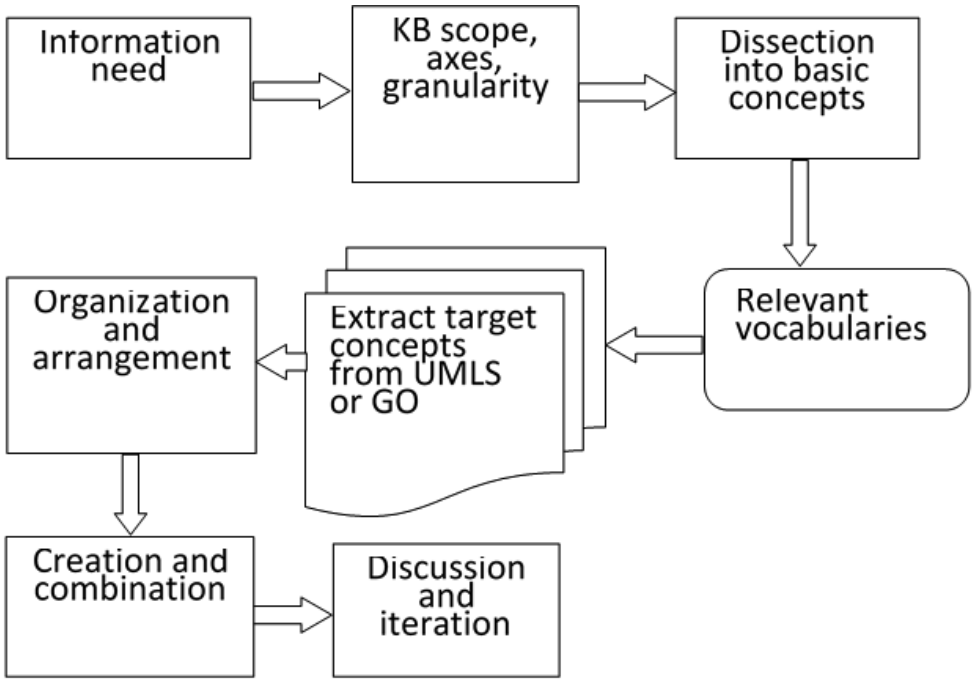
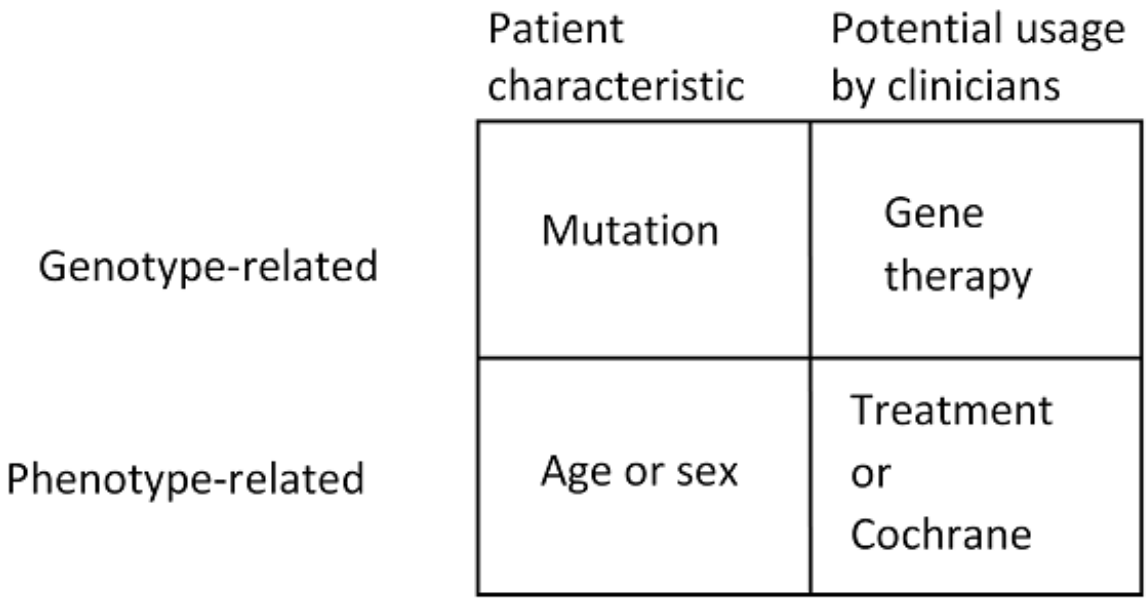
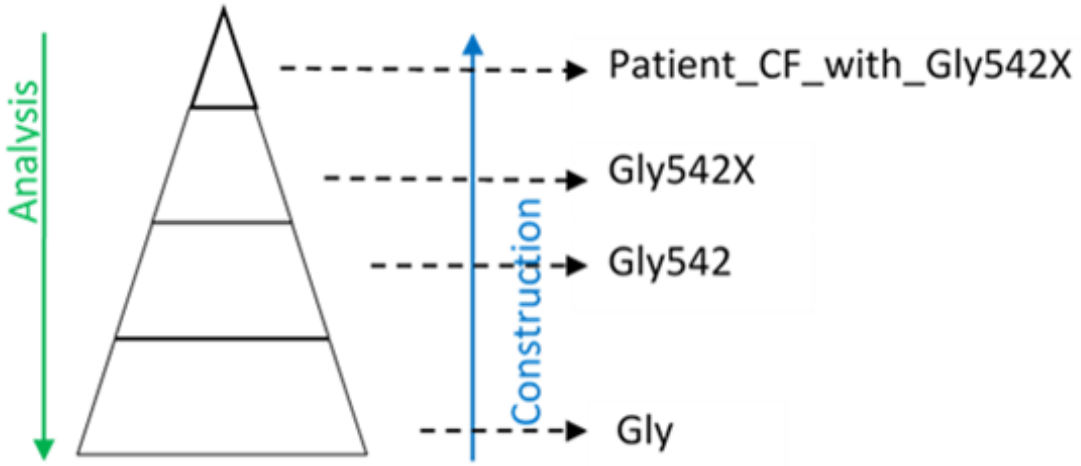


Figure 3. Main content categories represented in the ontology-based knowledge base prototype on cystic fibrosis and their potential usage.



XSL•FO
RenderX

Figure 4. General analysis (top-to-bottom) and construction (bottom-up) for ontology-based knowledge base prototype on cystic fibrosis.



Bottom-Up Construction Examples in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

OntoKBCF was constructed bottom-up [29] as follows. After dissection analysis, reconstruction work started from basic atomic concepts. Then, the basic concepts were modified step-by-step until the basic concept was turned into a meaningful composite (a combined or complex concept). Every later concept used the former ones in its representation. Then, final facts were represented by the combination of a subject (ie, its definition) and its properties. A combination of basic concepts was utilized to build combined and complex concepts. The complex concept could be utilized as the subject to describe characteristics for a subgroup of patients. Figure 4 shows the construction process of OntoKBCF.

By way of example, if a group of patients with the Gly542X variant needs to be represented, the following steps would be used (see Figure 4):

- "Gly" stands for glycine, a type of amino acid; this is a basic concept.
- "X" is the nonsense codon, which terminates the translation; this is a basic concept.
- "Gly542" is defined as an amino acid location in the human *CFTR* amino acid chain; this is a combined concept.
- "Gly542X" is defined as an amino acid substitution in human *CFTR* amino acid chain; this is a combined concept.
- "Patient_CF_with_Gly542X" is defined as a group of cystic fibrosis patients with the amino acid change; this is a complex concept.
- Final facts about the group of cystic fibrosis patients with the Gly542X variant are represented by the combination of both the logical definition of "Patient_CF_with_Gly542X" and its properties.

Most of the hierarchies of basic concepts in the knowledge base prototype follow UMLS and GO. In addition, relationships or classes were added or adjusted if there were no obvious choices. For example, in OntoKBCF, "sex_group" was used to connect "population_group" and "female," and the concept "*CFTR* gene" in UMLS was adapted to "Human *CFTR* gene"; considering the clinical usage potential, we distinguished the human origins of the *CFTR* gene.

Intersection and Combination Representation Principles

We used the intersection among basic concepts to represent combined and complex concepts in OntoKBCF. For example, "female adolescent cystic fibrosis patient" is represented as a complex concept in OntoKBCF. Figure 5 shows the intersectional representation principle in OntoKBCF for female (ie, the purple circle), adolescent (ie, the rose circle), and cystic fibrosis patients (ie, the black circle) as a group, which represents *all* possible properties of this group. Figure 6 shows the exact representation (ie, final facts are represented as the combination of a definition of a complex concept and all its properties on the right lower corner of the figure) that we created within Protégé-OWL. The description in Figure 6 is a subset of the intersection in Figure 5. The entire set of properties for the subject (intersection of 3 bigger ellipses), compared with the filled ellipse (in blue) represented in OntoKBCF, refers to all the possible properties related to an adolescent (in rose) female (in purple) CF patient (in black). The reasons that we did not represent all possible properties are listed in the section Top-to-Bottom Content Analysis in OntoKBCF.

XSL•FO
RenderX

Figure 5. Representation of adolescent female cystic fibrosis (CF) patient in ontology-based knowledge base prototype on cystic fibrosis.

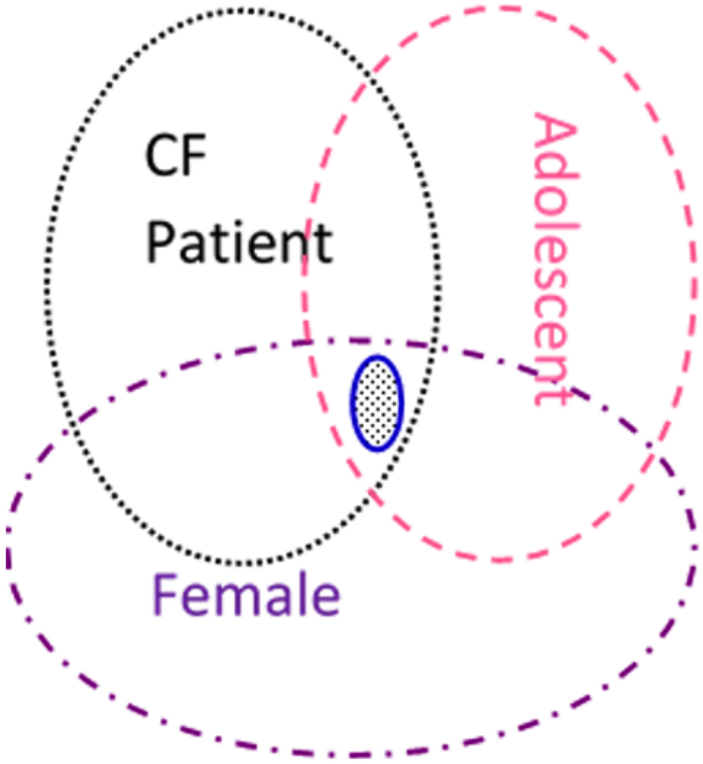
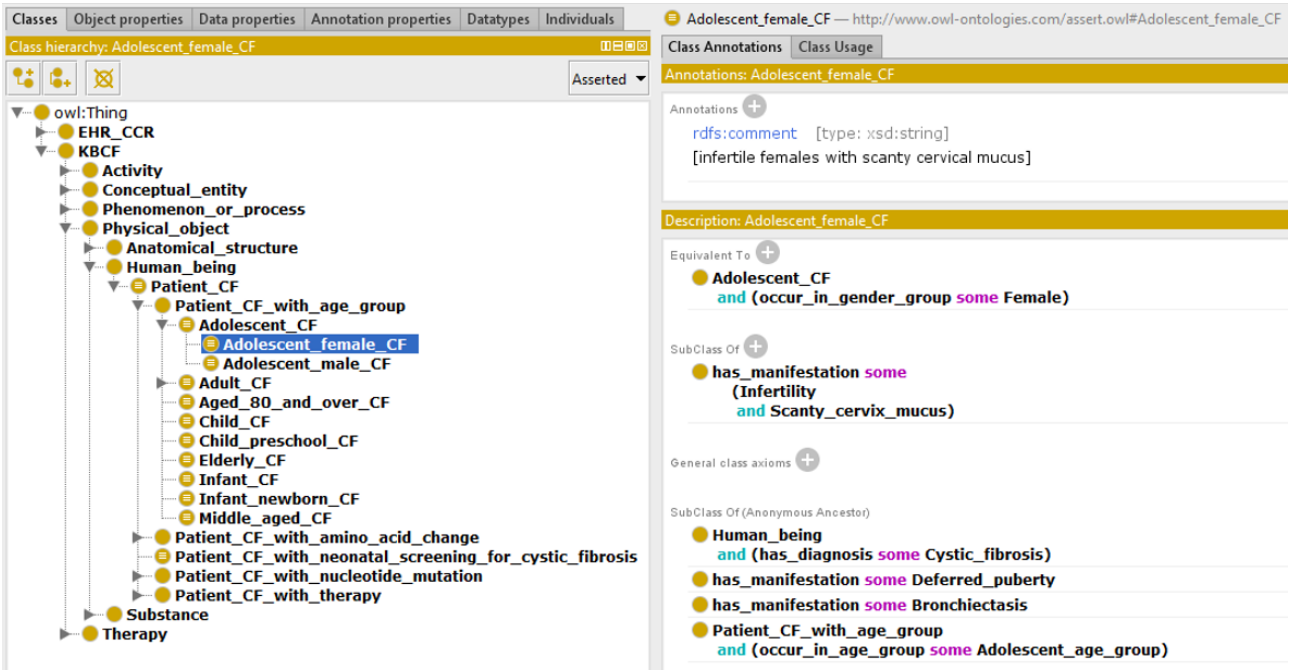


Figure 6. The representation of an adolescent female cystic fibrosis patient group in ontology-based knowledge base prototype on cystic fibrosis via Protégé-Web Ontology Language. (Protégé Source: Standford University).



Ontology-Based Knowledge Base Prototype on Cystic Fibrosis Construction Principles

Our *philosophy* is to use the existing concepts as much as possible and to create new ones only if it is absolutely necessary. If a concept is from UMLS, then the unique concept identifier from UMLS is kept in the Annotations section for future reference (eg, reuse, mapping, and communication). Other construction *principles* include the following: (1) to represent more types (eg, nucleotide deletion, insertion, transition, and transversion) of facts within the same level; (2) to represent 2-5 examples within the same type, for example, for the single-nucleotide transition of *CFTR*, we represented 5 examples in OntoKBCF instead of representing all existing facts

XSL•FO
RenderX

exhaustively; (3) to have the same representation listed in all possible hierarchies in OntoKBCF; for example, *AA2183_minus_G* is listed under *Del_A*, *Nucleotide_deletion_in_human_CFTR_gene*, *A_transition_G*, and *Nucleotide_transition_in_human_CFTR_gene*; and (4) to represent only the facts within the scope of OntoKBCF. The purpose of our work is to demonstrate the feasibility of the construction of a knowledge base by utilizing Semantic Web technology and demonstrating that the facts from the knowledge base can be utilized via an EHR interface dynamically. Therefore, representing the exhaustive list of variations of the same gene is not a priority. In addition, as most of the mutation types have been covered in the current OntoKBCF, we believe it to be extendable. To include an exhaustive list of the variations of the *CFTR* gene should be straightforward when following our construction principles and approaches.

Naming Conventions in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

To create the potential for broader usage, some of the classes were specifically named for human beings in OntoKBCF, such as "Human_CFTR_gene" and "Human_CFTR_gene_exon"; this differentiates them from any additional classes that may concern other organisms.

The Nomenclature for the Description of Sequence Variations [30] is followed in OntoKBCF. In addition, the following basic name rules are used in OntoKBCF: (1) for amino acid changes (ie, protein level), a 3-letter abbreviation name was used; (2) for coding DNA (cDNA)-level nucleotide changes, a 1-letter name (such as "A," "T," "C," and "G") was used. These design decisions are present in OntoKBCF as follows: (1) a mutation name, which strictly follows the nomenclature recommendation [30,31], has been kept in the "Annotation" section and (2) there were only cDNA- and protein-level descriptions of the *CFTR* mutation name, which did not include the RNA-level description. The reason is that cDNA is used to code protein, and RNA is more of a middle layer with regards to protein synthesis. Amino acids can be named in 1 or 3 letters. There are both amino acids and nucleotides in OntoKBCF, and 1-letter abbreviation names of amino acids can be confused with nucleotides. For example, A can be Alanine (an amino acid; the 3-letter abbreviation is Ala) or Adenine. Thus, we use 3-letter abbreviation names for amino acids and 1-letter names for nucleotides.

As there are 22 types of amino acids, there are many more possibilities for amino acid changes compared with nucleotide mutations; only the amino acid changes that fell within the scope of OntoKBCF were included. In contrast, because there are 8 types of nucleotides, which consist of DNA and RNA, with a smaller number of possible mutations, the complete set of possible nucleotide mutations in cDNA for deletion, insertion, transversion, and transition were included in OntoKBCF. The symbols "+" or "−" indicate the nucleotide position at the beginning of the intron (+) or the end of the intron (−); this would be represented as "plus" and "minus" in OntoKBCF for nucleotide substitutions. Amino acid substitutions are handled similarly.

In OntoKBCF, the final expression of nucleotide mutations can be distinguished from amino acid changes as follows: (1) their labels contained "minus" or "plus," or "Ins" or "Del" (such as "AA2183_minus" or "G621_plus_1", and "Del394" or "Ins3905"); or (2) a 3-letter abbreviation name was used for amino acids to avoid confusion. For example, Gly was utilized as the abbreviation for glycine, not G, which can be confused with the nucleotide "G." "Minus" and "plus" were used to represent the nucleotide position. "Ins" is used for nucleotide insertion, and "Del" is used for nucleotide deletion. In theory, not all nucleotide mutations have to be labeled with one of these 4 values; however, no such example has been found in the construction of OntoKBCF. This strategy avoids the potential limitations of the special characters that one can use in a name.

Other Considerations in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

For properties in the OWL representation, only existential restriction "someValuesFrom" was used as the property restriction (appearing as "some" in the text); it would be inappropriate to use allValuesFrom (appearing as "only" in the text,) because it cannot be guaranteed that their value classes are fully specified and mutually exclusive and exhaustive. It is sufficient to use existential restriction in most representations to leave space for future efforts to complete the description. For example, for a cystic fibrosis patient with Gly542X, the subject had the mutation property "some Gly542X" because it is possible that Gly542X is not the only mutation for this type of patient.

According to Horridge et al [32], the "Domain" and "Range" conditions in the "Property" tab of the Protégé-OWL interface are not used as a constraint. These properties are outside of the scope of this project, and consequently, the "Domain" and "Range" widgets were left blank for all properties.

Results

Ontology-Based Knowledge Base Prototype on Cystic Fibrosis Content Summary

Table 2 summarizes the metrics of OntoKBCF. Description logic expressivity was ALCHIF (attribute language; complex concept negation; role hierarchy; inverse properties; functional properties).

In OntoKBCF, the most fine-grain level of biological information starts at nucleobase, which is the most important component for the elementary units (nucleotides) of RNA and DNA. In contrast, for health information in OntoKBCF, the most fine-grain level starts from relatively *atomic concepts*, such as diarrhea, nausea, or coughing.

Currently, OntoKBCF includes answers to only questions about "what" (eg, what is Gly542X? What is the hierarchy of cystic fibrosis?), not about "how" (eg, disease mechanisms and disease pathways) or "why" (eg, reasons and explanations) [8]. This is a pragmatic approach to make this study manageable and feasible. We focus on "what" questions initially. We do recognize the importance of "how" and "why" questions in clinical usage, especially in explanation of pathophysiological

XSL•FO
RenderX

mechanisms and disease pathways. However, we leave these questions for future development. Textbox 1 lists the main categories of content in OntoKBCF. In the Multimedia Appendix 1, we also include a screenshot of all the object properties in OntoKBCF.

OntoKBCF has been shared via BioPortal [9]. We used a reasoner, HermiT 1.3.8.413 within Protégé-5.1.0, to check the consistency of OntoKBCF.

Ontology-Based Knowledge Base Prototype on Cystic Fibrosis Application Functions in an Electronic Health Record

OntoKBCF has been demonstrated to be able to supply personalized, nonduplicate molecular genetic information and health information according to a patient's characteristics successfully, dynamically, and automatically via an EHR prototype [14,15]. The connection between OntoKBCF and the EHR prototype can be made through a combination of automatic and manual means. OntoKBCF includes a section for the EHR main structure, which is utilized to map the domain knowledge in OntoKBCF to the EHR structure and to make the connection automatic.

OntoKBCF can provide reference facts (eg, hierarchies of basic concepts, combined concepts, and final facts) and reminders to EHR users. When a user applies the final facts via EHR interfaces, all the basic concepts and combined concepts utilized for building a complex concept can be accessed. Figure 7 shows a conceptual architecture about how OntoKBCF can be utilized within an EHR user interface.

Table 2. Metrics of ontology-based knowledge base prototype on cystic fibrosis content summary [9].

Name	Count (n)
Axiom	1375
Logical axiom	665
Declaration axiom	445
Class	408
Individual	0
Object property	35
SubClassOf	498
EquivalentClasses	113
DisjointClasses	37
Maximum depth	9
Maximum number of children	20
Average number of children	3
Classes with a single child	69
AnnotationAssertion	265

Textbox 1. Main categories of content in ontology-based knowledge base prototype on cystic fibrosis.

Biological information (genotype-related)

- Nucleotide mutation
- Amino acid change
- Gene mutation location
- Gene or protein mutation dissection

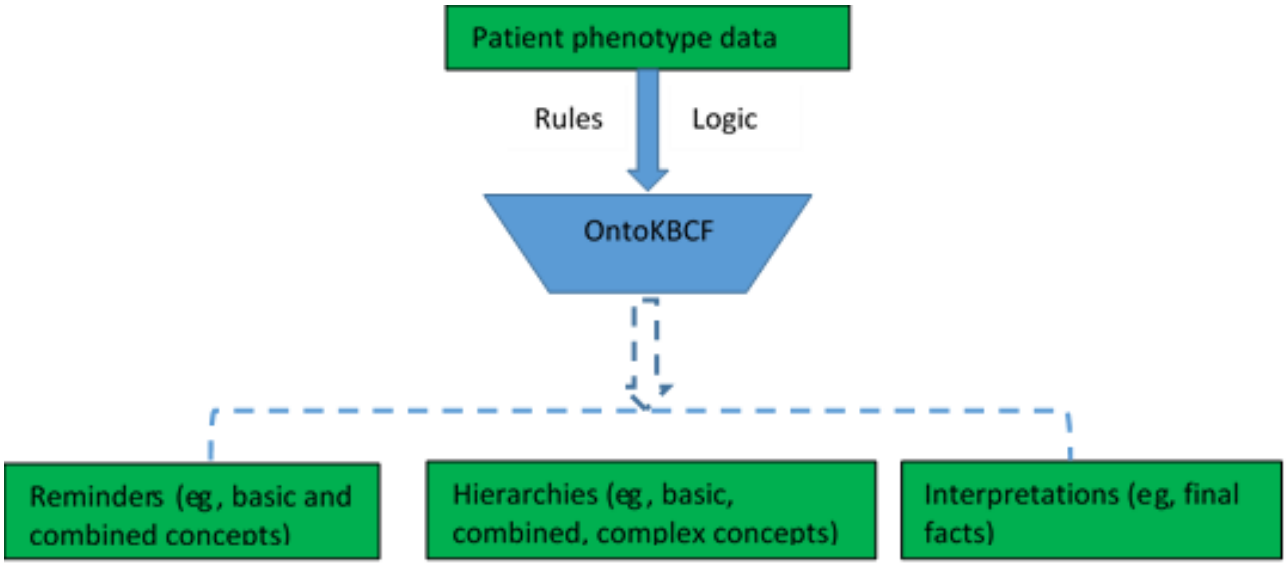
Health information (phenotype-related)

- Demographic data (eg, age, sex, and ethnicity)
- Symptoms
- Diagnostic tests or diagnosis
- Treatment facts

Electronic health record data representation structure

XSL•FO
RenderX

Figure 7. The conceptual architecture between ontology-based knowledge base prototype on cystic fibrosis (OntoKBCF) and an electronic health record interface.



Discussion

Significance of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

OntoKBCF meets our design objectives and has been integrated successfully with an EHR prototype [15]. This paper illustrates the construction principles, representation approaches, and challenges in constructing OntoKBCF in Protégé-OWL. We believe that the methods we reported to construct OntoKBCF will provide a useful reference for Semantic Web content developers, particularly in constructing knowledge bases on human monogenetic diseases for clinical usage. The basic and combined concepts and similar facts in both biological and health fields represented in OntoKBCF can be found in other diseases. OWL statements make OntoKBCF sharable and computer processable. The possibility of automated reasoning assures the consistency of the underlying ontology. The organization, structure, and use of existing, controlled vocabularies and terminologies as much as possible within OntoKBCF improve its potential for compatibility and reusability.

Patient Group Representation Challenge Example

In OntoKBCF, different groups of cystic fibrosis patients are represented separately. For example, there are cystic fibrosis patients in different age groups or with different genetic variations. Figure 8 shows the main categories of cystic fibrosis patient groups included in OntoKBCF.

Currently, for each cystic fibrosis patient group, we represent only the available facts from the selected resources and within the scope of OntoKBCF. We defined patient groups logically, rather than representing them comprehensively. Some cystic fibrosis patient groups, such as an adolescent cystic fibrosis patient group (Figure 9), are fairly straightforward, while some are fairly complicated because of available facts. Figure 6 demonstrates a partial representation of the adolescent female cystic fibrosis patient group.

For the complicated groups in OntoKBCF, the representation is far from comprehensive or exhaustive. We feel that if we were to set comprehensive representation as a goal, the task would be endless, as there are many properties related to the patient groups. The properties can include, among other things, clinical manifestations, treatment plans, nursing care plans, patient self-care plans and patient education. The detailed categories, as well as the actual representation for each property, would result in an exhaustive list. Our pragmatic solution to this scalability problem is to focus on practical applications, stay within the defined scope, and make it possible to expand in the future.

Ambiguous Concepts and Uncertain and Negative Statement Representation Challenge

There are some common expressions in clinical medicine, for example, “sometimes,” “especially,” “very common,” and “very rarely” that are challenging to represent in a computable and formal manner as their interrelationships are vague. These expressions are very important for human experts (such as physicians) to absorb or to apply the information. For example, for older children with cystic fibrosis, one important symptom is “nasal polys, especially if recurrent” [21]. However, we cannot represent this type of fact in a computable form. Likewise, Tao et al [33] recognized that representing uncertain expressions is a challenge in creating a temporal-related ontology.

In addition to uncertain concepts, negative descriptions cannot be represented precisely, either. Considering the limited reasoning support, less than ideal reasoning efficiency, and the complexity in modeling concepts, currently, for demonstration purposes, we included only the definite and positive results in OntoKBCF. We do realize that although this is necessary, it is a limitation of OntoKBCF as it does not reflect the real world comprehensively.

XSL•FO
RenderX

Figure 8. Main cystic fibrosis patient groups in ontology-based knowledge base prototype on cystic fibrosis via Protégé-Web Ontology Language. (Protégé Source: Standford University).

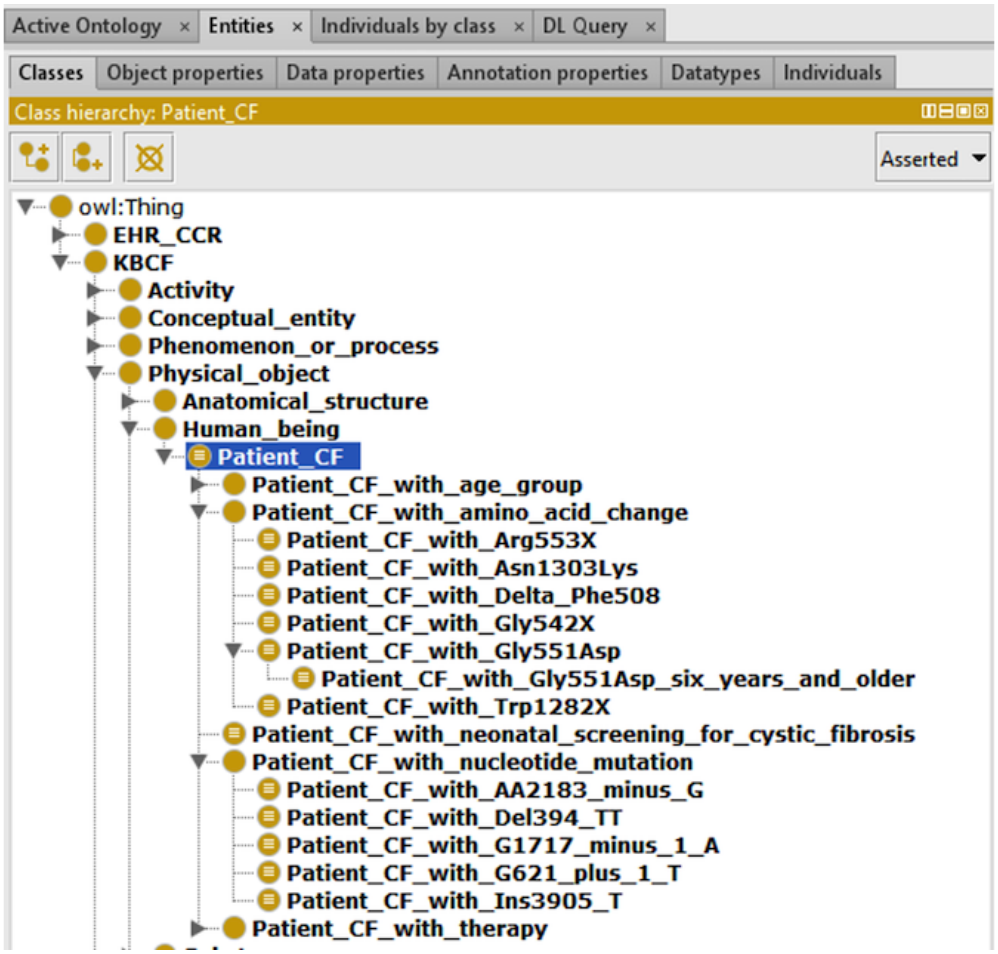
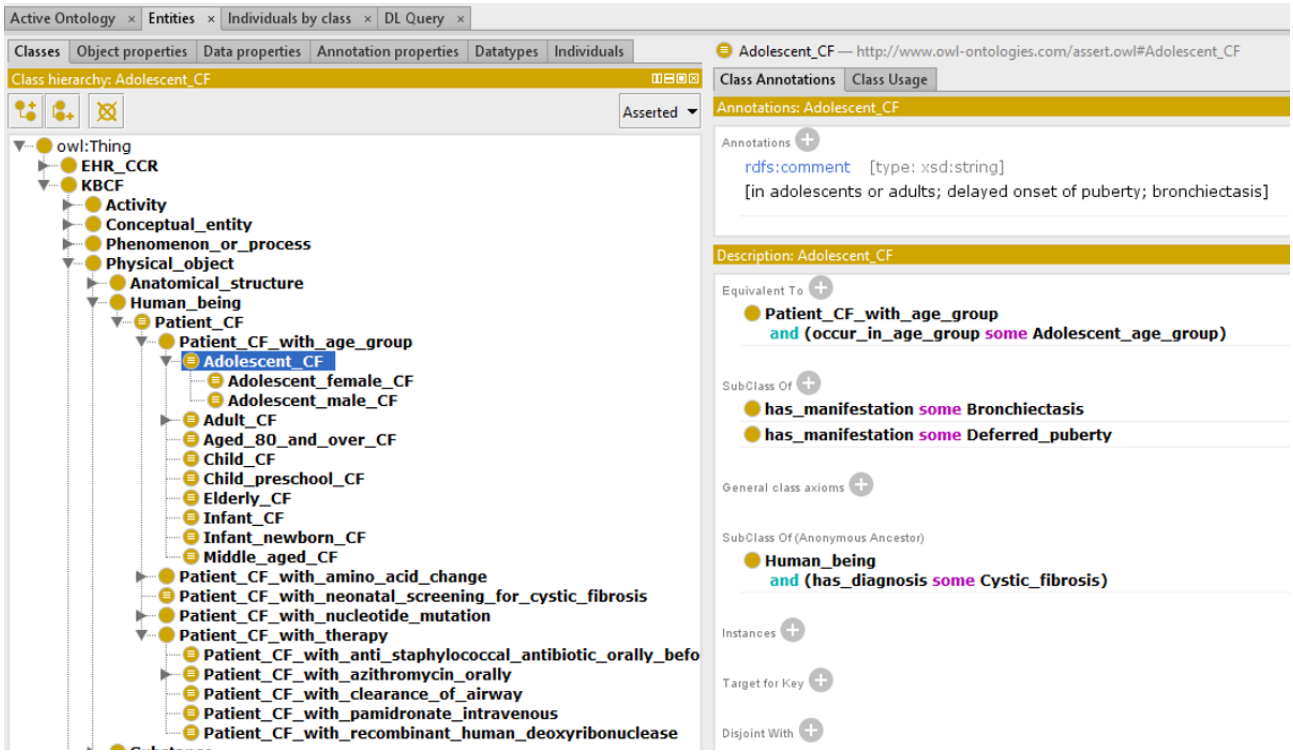
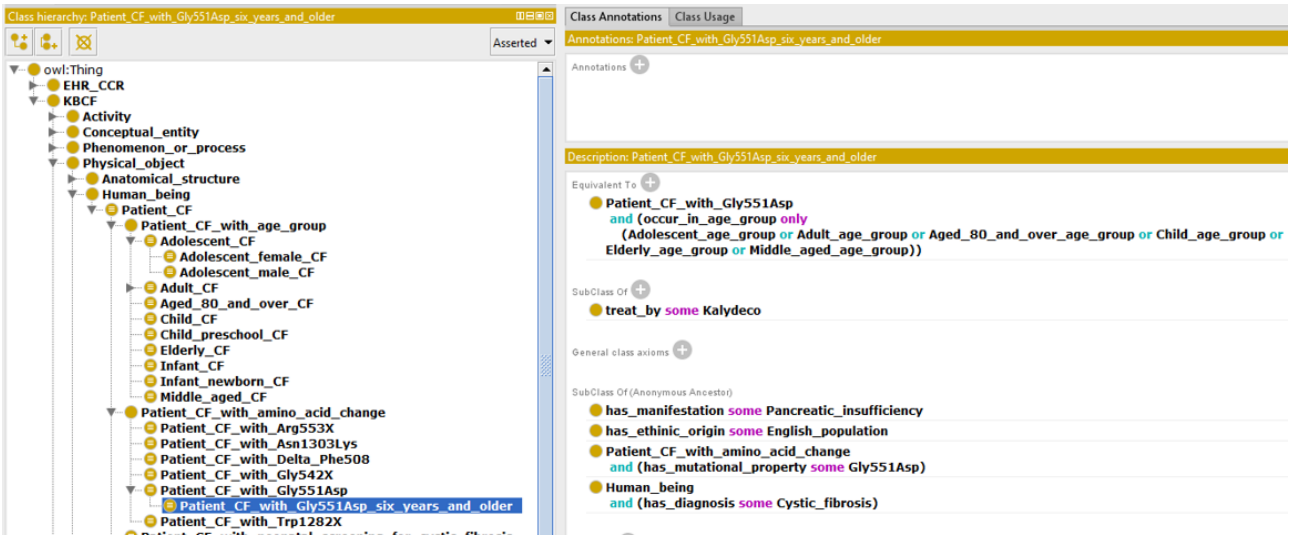


Figure 9. The representation of an adolescent cystic fibrosis patient group in ontology-based knowledge base prototype on cystic fibrosis via Protégé-Web Ontology Language. (Protégé Source: Standford University).



XSL•FO
RenderX

Figure 10. A cystic fibrosis patient group with the Gly551Asp variation and aged at least 6 years represented in ontology-based knowledge base prototype on cystic fibrosis via Protégé-Web Ontology Language. (Protégé Source: Standford University).



Complex Facts Representation Challenges

One of the complex facts represented in OntoKBCF is shown in Figure 10. This is a representation of a cystic fibrosis patient group whose patients have a mutation from Gly to Asp at 552 positions and are at least 6 years old. The current properties include diagnosis, treatment (ie, medication) information, age groups, ethnic origin, clinical manifestation, and exact amino acid change. This example has positive and definite statements, which are represented in the current OntoKBCF. However, how to represent more complicated facts, such as detailed molecular mechanisms or pathway information, with higher than first-order logic and how to process those descriptions also would be challenging.

Broader Challenges in the Construction of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

The challenges of introducing information technologies and computer science applications to medicine have been recognized for a long time [34]. Knowledge representation in medicine is one of the many areas that face issues. It is difficult to say that the challenge is in knowledge representation, *per se*, or that the challenges are in the broader area of medicine, logic, or even mathematics.

Regarding OntoKBCF, we sought to determine how many properties are sufficient or ideal to describe a cystic fibrosis patient group. From a pragmatic point of view, setting up the scope of the project is necessary to make the project manageable. Otherwise, the representation would become an endless task, as there are many possible properties related to a group of cystic fibrosis patients, as Davis et al [35] described. If we assume that we constructed OntoKBCF and others constructed similar knowledge bases about cystic fibrosis, then other management-related challenges would emerge, for example, (1) how to validate classes from different sources; (2) how to match the same classes from different sources; (3) how to integrate the properties for the same class that originates from different sources; (4) how to integrate different ontologies about the same topic; and (5) how to authorize and achieve consensus for the combined classes or ontologies conveniently. Currently, BioPortal [36] is an effective way to collect submitted biomedical ontologies; it can be utilized to check existing ontologies before building any new ontologies. Ontobee [37] provides a finer search on ontology terms, which goes further than index-based searches on ontologies, by providing an accessible unique resource identifier (URI) for ontology terms. These examples provide great potential in reusing and mapping ontology terms; however, many of the abovementioned challenges are still not solved. In addition to using URI, including the coded concepts, such as unique identifiers in UMLS or the clinical terminology systematized nomenclature of medicine-clinical terms, may improve the reusability of the knowledge base in general.

One critical challenge in building OntoKBCF is a lack of sufficient clinical actionable knowledge that is related to molecular genetic information. This is one of the main reasons that we constructed OntoKBCF as a prototype rather than as a product knowledge base.

Other challenges that we faced during the construction of OntoKBCF included the following: (1) how to communicate the construction principles and design considerations in a consistent, explicit, and sharable manner; (2) whether we can make the construction process automatic; and (3) how to update the domain content automatically. Solving these challenges may help us build larger-scale ontologies and help in their integration automatically.

Other Related Work

Our work was originally conceived in 2004 and mostly completed by 2009. At the time, there were no available resources that could meet *all* of the following criteria: consistent molecular genetic information and a clinical actionable, machine-processable format that was shareable, reusable, and customizable. We, thus, had to create our own resource to meet all the requirements. Biobanks might be a good source to provide such information. Nevertheless, the UK Biobank [38] started to recruit participants only in 2006, and the USA Biobank, as

XSL•FO
RenderX

a part of the Precision Medicine Initiative, was founded only in 2016 [39].

We used an ontology-based knowledge base to provide customizable molecular genetics and health information to EHR settings successfully. The ontology-based knowledge base provides potential in reusing and sharing, as well as the consistency of information. The creation of knowledge resources with fine granularity (not only disease names) has been recognized repeatedly [40,41] as the main challenge in bringing new information (eg, molecular genetics information) to an EHR. Although there are many existing databases for both genotypes and phenotypes of human beings, not all of them are organized in a machine-processable format. Therefore, the detailed construction principles and approaches for OntoKBCF reported in this paper can provide a reference to peers in the field.

One recent example of knowledge base was reported by Samwald et al [42,43] who built a resource description framework or OWL knowledge base to provide support for clinical pharmacogenetics. Their work shares some similarities with OntoKBCF. For example, it provides consistent genomics information in clinical settings in a machine-processable format. Both projects use reasoners to maintain consistency. The two projects, however, have different focused application domain areas; Samwald et al's work focuses on pharmacogenomics information, and OntoKBCF uses cystic fibrosis to demonstrate possible broader applications. In addition, Samwald et al's work includes a query and answer part that supersedes OntoKBCF. Meanwhile, OntoKBCF demonstrates its usage via an EHR prototype; such demonstration was not included in Samwald et al's [42,43] publications. There is no detailed description of the knowledge base construction in Samwald et al's project. Thus, it is difficult to compare the representation and construction principles and approaches of the two projects in depth.

In recent years, the broad importance of Semantic Web technologies has been recognized. There are many more studies that have used ontology-based knowledge bases to assist clinical tasks within an EHR. For example, Robles-Bykbaev et al [40] reported the use of a formal knowledge base model to assist in the generation of decision making and recommendations for communication disorders. They stated that their knowledge base is used to support data analysis and inference processes; however, the paper does not include the details of the organization and structure of the knowledge base. The Clinical Narrative Temporal Relation Ontology (CNTRO) [33,44] is a temporal-related ontology in the OWL. CNTRO is used for inference purposes in processing clinical narratives. The use cases [45] show promising results in processing short and simple adverse event narratives. The OWL query application [46] for CNTRO and the harmonization of CNTRO with other existing time ontologies may improve CNTRO [44] further and provide broader applications in analyzing clinical narratives. Another example is the research of Wang et al [41] and Hu et al [47]. They utilized an ontology-based clinical pathways knowledge base to generate personalized clinical pathways for clinicians by incorporating patients' data. The clinical pathways knowledge base is independent of an EHR, and it can be shared by other systems. All these examples used ontology-based knowledge bases and integrated them with an EHR. There are other ontology examples that are utilized in settings other than in EHR, such as in mobile devices [48] or to facilitate social media data mining [49] for health purposes. While these recent studies demonstrate the general application of semantic technologies in health care, they do not consider the important clinical usage of relevant molecular genetic information.

Limitations of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

OntoKBCF is a knowledge base prototype rather than a knowledge base, which may include substantial instances for each class. OntoKBCF, however, includes content from nucleotides to clinical phenotypes across different granularity levels. If we can image a pyramid that moves from gene to protein to cell to tissue to organ up to human beings [50], we can see that the current OntoKBCF is not thorough at any horizontal level; however, it includes all vertical levels of knowledge. The expansion of each level should not be too challenging, as each horizontal level has some classes represented in OntoKBCF.

There has been no formal evaluation of OntoKBCF. However, we do have two limited levels of validation of the work: (1) automated reasoning assures the consistency of underlying ontology and (2) successful demonstration that OntoKBCF can deliver customizable molecular information and health information dynamically via an EHR prototype. As a knowledge base model, these levels of validation provide evidence that the original design objectives of OntoKBCF were met successfully.

Conclusion

We have presented the robust construction principles, approaches, design decisions, and challenges in constructing OntoKBCF. OntoKBCF is constructed in OWL and can be utilized as a knowledge resource to provide information on molecular genetics and other health topics dynamically to an EHR. The construction principles and approaches can be referenced for other topic areas, and based on its current structure, OntoKBCF could be expanded in a straightforward manner to a full knowledge base.

Acknowledgments

The authors would like to thank Mr Tom Marley for his contribution to the project and Dr Miao Sun for her patience, support, and expert knowledge in molecular genetics. Sincere thanks go to Prof James J Cimino for highly constructive discussions and valuable suggestions offered for the project. The work was funded by the University of Salford and partially through Ohio University, College of Health Sciences and Professions' new faculty startup funds. This work was conducted using the Protégé

XSL•FO
RenderX

resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional screenshot of properties of OntoKBCF.

[PNG File, 310KB - medinform_v6i4e52_app1.png]

References

1. McGuinness DL, van Harmelen F. World Wide Web Consortium (W3C). 2004 Feb 10. OWL Web Ontology Language Overview URL: https://www.w3.org/TR/owl-features/ [accessed 2006-03-30] [WebCite Cache ID 5EmVp1112]
2. Updegrove C. ConsortiumInfo.org. 2005 Jun. The Semantic Web: An Interview With Tim Berners-Lee URL: http://www.consortiuminfo.org/bulletins/semanticweb.php [accessed 2018-01-29] [WebCite Cache ID 6wq92rzvh]
3. Fensel D, Davies J, Harmelen F. Towards the Semantic Web: Ontology-driven Knowledge Management. Hoboken, NJ: Wiley, John & Sons, Incorporated; 2002.
4. World Wide Web Consortium (W3C). World Wide Web Consortium Issues RDF and OWL Recommendations: Semantic Web emerges as commercial-grade infrastructure for sharing data on the Web URL: https://www.w3.org/2004/01/sws-pressrelease [accessed 2018-01-29] [WebCite Cache ID 6wq9PlUMm]
5. Gruber T. Stanford Knowledge Systems, AI Laboratory. 1999. What is an Ontology? URL: http://www-ksl.stanford.edu/kst/what-is-an-ontology.html [accessed 2009-03-08] [WebCite Cache ID 5f87rn7k1]
6. W3C OWL Working Group. World Wide Web Consortium (W3C). 2012 Dec 11. OWL 2 Web Ontology Language Document Overview (Second Edition) URL: https://www.w3.org/TR/owl2-overview/ [accessed 2012-03-29] [WebCite Cache ID 66VqVQDlO]
7. Ozsu MT, Liu L. Encyclopedia Of Database Systems. New York City, NY: Springer-Verlag New York Inc; 2018.
8. Jing X, Kay S, Hardiker N, Marley T. Ontology-based knowledge base model construction-OntoKBCF. Stud Health Technol Inform 2007;129(Pt 1):785-790. [Medline: 17911824]
9. Jing X, Kay S, Hardiker N, Marley T, Cimino JJ. NCBO BioPortal. 2013 Jan 16. Ontological Knowledge Base Model for Cystic Fibrosis --BioPortal Internet URL: http://bioportal.bioontology.org/ontologies/ONTOKBCF [accessed 2018-06-29] [WebCite Cache ID 70XneQXIb]
10. National Human Genome Research Institute. Learning About Cystic Fibrosis URL: https://www.genome.gov/10001213/learning-about-cystic-fibrosis/ [accessed 2018-01-29] [WebCite Cache ID 6wq9nLDB7]
11. Mayo Clinic. Cystic fibrosis URL: https://www.mayoclinic.org/diseases-conditions/cystic-fibrosis/symptoms-causes/syc-20353700 [accessed 2018-01-29] [WebCite Cache ID 6wqAnQW0e]
12. Genetics Home Reference: NIH. Cystic fibrosis URL: https://ghr.nlm.nih.gov/condition/cystic-fibrosis [accessed 2018-01-29] [WebCite Cache ID 6wqBiqmuj]
13. Bassiliades N, Vlahavas I. Active Knowledge-Based Systems: Techniques and Applications. In: Leondes CT, editor. Knowledge Engineering: Systems Techniques and Applications. London: Academic Press; 2000.
14. Jing X, Kay S, Marley T, Hardiker NR, Cimino JJ. Incorporating personalized gene sequence variants, molecular genetics knowledge, and health knowledge into an EHR prototype based on the Continuity of Care Record standard. J Biomed Inform 2012 Feb;45(1):82-92 [FREE Full text] [doi: 10.1016/j.jbi.2011.09.001] [Medline: 21946299]
15. Jing X, Kay S, Marley T, Hardiker NR. Integration of an OWL-DL knowledge base with an EHR prototype and providing customized information. J Med Syst 2014 Sep;38(9):75. [doi: 10.1007/s10916-014-0075-4] [Medline: 24997857]
16. Musen MA, Protégé T. The Protégé Project: A Look Back and a Look Forward. AI Matters 2015 Jun;1(4):4-12 [FREE Full text] [doi: 10.1145/2757001.2757003] [Medline: 27239556]
17. Lindberg D, Humphreys B, McCray A. The Unified Medical Language System. Methods Inf Med 1993;32(4):281-291 [FREE Full text]
18. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000 May;25(1):25-29 [FREE Full text] [doi: 10.1038/75556] [Medline: 10802651]
19. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007;24(2):78-94.
20. Arroll B, Pandit S, Kerins D, Tracey J, Kerse N. Use of information sources among New Zealand family physicians with high access to computers. J Fam Pr 2002;51(8):706.
21. Harris A, Super M. Cystic fibrosis: the facts. 3rd ed. Oxford: Oxford University Press; 1995.
22. Orenstein D, Rosenstein B, Stern R. Cystic fibrosis: medical care. Philadelphia: Lippincott Williams & Wilkins; 2000.

XSL•FO
RenderX

23. Cystic Fibrosis Foundation. Adding Tomorrows and Living Today URL: https://www.cff.org/ [accessed 2006-04-07] [WebCite Cache ID 5Ey9oAgZ9]
24. Cystic Fibrosis Trust. What is CF Internet URL: https://www.cysticfibrosis.org.uk/what-is-cystic-fibrosis [accessed 2018-01-29] [WebCite Cache ID 6wqC9e8bP]
25. Rosenstein BJ. The Merck Manuals. Cystic Fibrosis URL: http://www.merckmanuals.com/professional/pediatrics/cystic-fibrosis-cf/cystic-fibrosis [accessed 2018-01-29] [WebCite Cache ID 6wqCTk96o]
26. Zielenski J. Genotype and phenotype in cystic fibrosis. Respiration 2000;67(2):117-133. [doi: 10.1159/000029497] [Medline: 10773783]
27. Zielenski J, Tsui L. Cystic fibrosis: genotypic and phenotypic variations. Annu Rev Genet 1995;29:777-807. [doi: 10.1146/annurev.ge.29.120195.004021] [Medline: 8825494]
28. Cystic Fibrosis: Mutation Database. URL: http://www.genet.sickkids.on.ca/cftr/ [accessed 2005-12-07] [WebCite Cache ID 5C0czfzNe]
29. Rector A. School of Computer Science: The University of Manchester. 2005. Foundations of the Semantic Web: Ontology Engineering Internet URL: http://www.cs.man.ac.uk/~rector/modules/CS646-2005/ [accessed 2018-01-29] [WebCite Cache ID 6wqCrJkUB]
30. den Dunnen JT, Antonarakis S. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 2000;15(1):7-12. [doi: 10.1002/(SICI)1098-1004(200001)15:1<7::AID-HUMU4>3.0.CO;2-N] [Medline: 10612815]
31. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Hum Mutat 2016 Dec;37(6):564-569. [doi: 10.1002/humu.22981] [Medline: 26931183]
32. Horridge M, Knublauch H, Rector A, Stevens R, Wroe C. CO-ODE Project. A Practical Guide To Building OWL Ontologies Using The Prot´eg´e-OWL Plugin and CO-ODE Tools Edition 1 URL: http://mowl-power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf [accessed 2009-03-08] [WebCite Cache ID 5f88oV7Wk]
33. Tao C, Wei W, Solbrig H, Savova G, Chute C. CNTRO: a semantic web onotlogy for temporal relation inferencing in clinical narratives. In: AMIA Annu Symp Proc. 2010 Presented at: AMIA 2010; Nov 13-17; Washington DC p. 787.
34. Shortliffe E, Cimino J, editors. Biomedical informatics: computer applications in health care and biomedicine. 4th ed. London: Springer; 2014.
35. Davis R, Shrobe H, Szolovits P. What is a knowledge representation? AI Magazine 1993;14(1):33. [doi: 10.1609/aimag.v14i1.1029]
36. Whetzel P, Noy N, Shah N, Alexander P, Nyulas C, Tudorache T, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res 2011 Jul;39(Web Server issue):W541-W545 [FREE Full text] [doi: 10.1093/nar/gkr469] [Medline: 21672956]
37. Ong E, Xiang Z, Zhao B, Liu Y, Lin Y, Zheng J, et al. Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. Nucleic Acids Res 2017 Jan 04;45(D1):D347-D352 [FREE Full text] [doi: 10.1093/nar/gkw918] [Medline: 27733503]
38. UK Biobank. URL: http://www.ukbiobank.ac.uk/ [accessed 2018-06-29] [WebCite Cache ID 70XnSxIey]
39. NIH Office of the Director. NIH funds biobank to support Precision Medicine Initiative Cohort Program Internet URL: https://www.nih.gov/news-events/news-releases/nih-funds-biobank-support-precision-medicine-initiative-cohort-program [accessed 2018-01-29] [WebCite Cache ID 6wqDkX0dV]
40. Robles-Bykbaev V, López-Nores M, Pazos-Arias J, Quisi-Peralta D, García-Duque J. An Ecosystem of Intelligent ICT Tools for Speech-Language Therapy Based on a Formal Knowledge Model. Stud Health Technol Inform 2015;216:50-54. [Medline: 26262008]
41. Wang H, Li J, Zhang Y, Suzuki M, Araki K. Creating personalised clinical pathways by semantic interoperability with electronic health records. Artif Intell Med 2013 Jun;58(2):81-89. [doi: 10.1016/j.artmed.2013.02.005] [Medline: 23466439]
42. Samwald M, Miñarro GJA, Boyce RD, Freimuth RR, Adlassnig K, Dumontier M. Pharmacogenomic knowledge representation, reasoning and genome-based clinical decision support based on OWL 2 DL ontologies. BMC Med Inform Decis Mak 2015 Feb 22;15:12 [FREE Full text] [doi: 10.1186/s12911-015-0130-1] [Medline: 25880555]
43. Samwald M, Freimuth R, Luciano J, Lin S, Powers R, Marshall M, et al. An RDF/OWL knowledge base for query answering and decision support in clinical pharmacogenetics. Stud Health Technol Inform 2013;192:539-542 [FREE Full text] [Medline: 23920613]
44. Tao C, Solbrig H, Chute C. CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives. AMIA Jt Summits Transl Sci Proc 2011;2011:64-68 [FREE Full text] [Medline: 22211182]
45. Clark K, Sharma D, Qin R, Chute CG, Tao C. A use case study on late stent thrombosis for ontology-based temporal reasoning and analysis. J Biomed Semantics 2014;5(1):49 [FREE Full text] [doi: 10.1186/2041-1480-5-49] [Medline: 25540680]
46. Zhu Q, Tao C, Ding Y, Chute C. Time-related patient data retrieval for the case studies from the pharmacogenomics research network. J Med Syst 2012 Nov;36 Suppl 1:S37-S42 [FREE Full text] [doi: 10.1007/s10916-012-9888-1] [Medline: 23076712]

XSL•FO
RenderX

47. Hu Z, Li J, Zhou T, Yu H, Suzuki M, Araki K. Ontology-based clinical pathways with semantic rules. J Med Syst 2012 Aug;36(4):2203-2212. [doi: 10.1007/s10916-011-9687-0] [Medline: 21445676]
48. Kim H, Park H, Min YH, Jeon E. Development of an obesity management ontology based on the nursing process for the mobile-device domain. J Med Internet Res 2013;15(6):e130 [FREE Full text] [doi: 10.2196/jmir.2512] [Medline: 23811542]
49. Jung H, Park H, Song T. Ontology-Based Approach to Social Data Sentiment Analysis: Detection of Adolescent Depression Signals. J Med Internet Res 2017 Jul 24;19(7):e259 [FREE Full text] [doi: 10.2196/jmir.7452] [Medline: 28739560]
50. Blois M. Information and medicine: the nature of medical descriptions Internet. Berkeley: University of California Press; 1984.

Abbreviations

***CFTR*:** cystic fibrosis transmembrane conductance regulator
CNTRO: Clinical Narrative Temporal Relation Ontology
EHR: electronic health record
GO: Gene Ontology
OntoKBCF: Ontology-based Knowledge Base prototype on Cystic Fibrosis
OWL: Web Ontology Language
UMLS: Unified Medical Language System
URI: unique resource identifier

Edited by G Eysenbach; submitted 29.01.18; peer-reviewed by W Goossen, C Tao; comments to author 21.03.18; revised version received 25.04.18; accepted 19.07.18; published 21.12.18.

Please cite as:
Jing X, Hardiker NR, Kay S, Gao Y
Identifying Principles for the Construction of an Ontology-Based Knowledge Base: A Case Study Approach
JMIR Med Inform 2018;6(4):e52
URL: http://medinform.jmir.org/2018/4/e52/
doi:10.2196/medinform.9979
PMID:30578220

©Xia Jing, Nicholas R Hardiker, Stephen Kay, Yongsheng Gao. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 21.12.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Identifying Patients Who Are Likely to Receive Most of Their Care From a Specific Health Care System: Demonstration via Secondary Analysis

Gang Luo[1], PhD; Peter Tarczy-Hornoch[1,2,3], MD; Adam B Wilcox[1], PhD; E Sally Lee[4], PhD

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

[2]Division of Neonatology, Department of Pediatrics, University of Washington, Seattle, WA, United States

[3]Department of Computer Science and Engineering, University of Washington, Seattle, WA, United States

[4]Population Health Analytics, University of Washington Medicine Finance, University of Washington, Seattle, WA, United States

Corresponding Author:
Gang Luo, PhD
Department of Biomedical Informatics and Medical Education
University of Washington
University of Washington Medicine South Lake Union
850 Republican Street, Building C
Box 358047
Seattle, WA, 98195
United States
Phone: 1 206 221 4596
Fax: 1 206 221 2671
Email: gangluo@cs.wisc.edu

Abstract

Background: In the United States, health care is fragmented in numerous distinct health care systems including private, public, and federal organizations like private physician groups and academic medical centers. Many patients have their complete medical data scattered across these several health care systems, with no particular system having complete data on any of them. Several major data analysis tasks such as predictive modeling using historical data are considered impractical on incomplete data.

Objective: Our objective was to find a way to enable these analysis tasks for a health care system with incomplete data on many of its patients.

Methods: This study presents, to the best of our knowledge, the first method to use a geographic constraint to identify a reasonably large subset of patients who tend to receive most of their care from a given health care system. A data analysis task needing relatively complete data can be conducted on this subset of patients. We demonstrated our method using data from the University of Washington Medicine (UWM) and PreManage data covering the use of all hospitals in Washington State. We compared 10 candidate constraints to optimize the solution.

Results: For UWM, the best constraint is that the patient has a UWM primary care physician and lives within 5 miles of at least one UWM hospital. About 16.01% (55,707/348,054) of UWM patients satisfied this constraint. Around 69.38% (10,501/15,135) of their inpatient stays and emergency department visits occurred within UWM in the following 6 months, more than double the corresponding percentage for all UWM patients.

Conclusions: Our method can identify a reasonably large subset of patients who tend to receive most of their care from UWM. This enables several major analysis tasks on incomplete medical data that were previously deemed infeasible.

(JMIR Med Inform 2018;6(4):e12241) doi:10.2196/12241

KEYWORDS

data analysis; inpatients; emergency departments; health care system

XSL•FO
RenderX

Introduction

In the United States, health care is fragmented in numerous distinct health care systems including private, public, and federal organizations like private physician groups and academic medical centers. Frequently, a given health care system has incomplete medical data on many of its patients, as these patients' complete data are recorded across multiple health care systems [1,2]. Finnell et al [2] showed that during a 3-year period in Indiana, 40.7% of emergency department visits came from patients who also had emergency department visits at other health care systems. Bourgeois et al [1] showed that during a 5-year period in Massachusetts, 56.5% of adult hospital encounters (inpatient stays and emergency department visits) came from patients who also had encounters at other hospitals. Incomplete data are particularly problematic in academic health care systems such as the University of Washington Medicine (UWM), where many patients are referred from other health care systems. As shown in the Results section, less than one-third of the hospital encounters for all UWM patients occur within UWM. Currently, several major data analysis tasks such as predictive modeling using historical data are deemed impractical on incomplete data. This limits the applications based on these analysis tasks. For example, predictive modeling is widely used for identifying future high-cost patients [3] for care management [4] to prevent high costs and health status degradation [5-7]. Typical models for projecting a patient's cost assume complete historical data [8-10] and are not used by a health care system with incomplete data on its patients. As a result, many future high-cost patients are not identified and enrolled in care management, contributing to undesirable outcomes.

This study presents, to the best of our knowledge, the first method to use a geographic constraint to identify a reasonably large subset of patients who tend to receive most of their care from a specific health care system. This is to enable these data analysis tasks on incomplete medical data. Although the health care system has incomplete data on many of its patients, it has more complete data on this subset of patients. For a data analysis task requiring relatively complete medical data, we can conduct the task for this subset of patients, with the understanding that the analysis results apply to only this subset of patients rather than all patients of the health care system. This could be an improvement compared with the current practice of not conducting the task at all, in cases when conducting the task on all patients is impractical. Our previous work [11] sketched the method's main goal but did not complete the method, do a computer coding implementation, or evaluate the method's performance. This study aims to fill these gaps and demonstrate our method using data at UWM.

Methods

Patient Population

The patient cohort included all adult patients (aged ≥18 years) who had encounters at UWM facilities (hospitals and clinics) with information stored in UWM's enterprise data warehouse during the 1-year period of April 1, 2016, to March 31, 2017. In this paper, an encounter can be of any type, unless it is explicitly specified as a hospital encounter or an outpatient visit. UWM is the largest academic health care system in Washington State and has both hospitals and clinics for adults.

Dataset

We used administrative data in UWM's enterprise data warehouse during the 2-year period of April 1, 2015, to March 31, 2017. The dataset included encounter and primary care physician (PCP) information of our patient cohort. We also used PreManage data that UWM has on all of its patients during the 6-month period of April 1, 2017, to September 30, 2017. PreManage is Collective Medical Technologies Inc's commercial product offering encounter and diagnosis data on hospital encounters (inpatient stays and emergency department visits) at many US hospitals [12]. PreManage data cover all hospitals in Washington State. Starting April 1, 2017, UWM has been receiving relatively complete PreManage data on its patients. In this paper, we chose April 1, 2017, as the index date separating the prior and subsequent periods for the analysis task.

Our Constraint-Based Patient Identification Method

Our goal is to use a constraint to identify a reasonably large subset of patients who tend to receive most of their care from UWM. We considered 3 UWM hospitals whose administrative and clinical data are stored in the UWM's enterprise data warehouse: Harborview Medical Center, University of Washington Medical Center, and Northwest Hospital. All the 3 hospitals are in Seattle, Washington. We considered the following candidate constraints that all include the component of living within r miles of at least one of the 3 UWM hospitals, with r being a parameter whose optimal value was to be determined in the study:

1. Distance only: The patient lives within r miles of at least one of the 3 UWM hospitals. Intuitively, with everything else being equal, the closer a patient lives to UWM hospitals, the larger portion of the patient's care tends to be received from UWM. In addition, the smaller the r, the smaller the number of UWM patients satisfying the constraint.
2. PCP: The patient has a UWM PCP and lives within r miles of at least one of the 3 UWM hospitals. UWM PCPs tend to make referrals within UWM. Hence, intuitively, compared with others, patients with a UWM PCP may receive a larger portion of their care from UWM.
3. ≥2 encounters in the past year: The patient had ≥2 encounters at UWM facilities in the past year (April 1, 2016, to March 31, 2017) and lives within r miles of at least one of the 3 UWM hospitals. Intuitively, patients with more previous encounters at UWM facilities may receive a larger portion of their care from UWM in the future.
4. ≥2 encounters in the past 2 years: The patient had ≥2 encounters at UWM facilities in the past 2 years (April 1, 2015, to March 31, 2017) and lives within r miles of at least one of the 3 UWM hospitals.
5. ≥1 hospital encounter in the past year: The patient had ≥1 hospital encounter in 1 of the 3 UWM hospitals in the past year and lives within r miles of at least one of the 3 UWM hospitals.

XSL•FO
RenderX

6. ≥2 hospital encounters in the past year: The patient had ≥2 hospital encounters across the 3 UWM hospitals in the past year and lives within *r* miles of at least 1 of the 3 UWM hospitals.
7. ≥2 hospital encounters in the past 2 years: The patient had ≥2 hospital encounters across the 3 UWM hospitals in the past 2 years and lives within *r* miles of at least one of the 3 UWM hospitals.
8. ≥1 outpatient visit in the past year: The patient had ≥1 outpatient visit to UWM in the past year and lives within *r* miles of at least one of the 3 UWM hospitals.
9. ≥2 outpatient visits in the past year: The patient had ≥2 outpatient visits to UWM in the past year and lives within *r* miles of at least one of the 3 UWM hospitals.
10. ≥2 outpatient visits in the past 2 years: The patient had ≥2 outpatient visits to UWM in the past 2 years and lives within *r* miles of at least one of the 3 UWM hospitals.

In each candidate constraint, distance is no longer a factor when $r=+\infty$.

Data Analysis

Using the distVincentyEllipsoid function in R's geosphere package version 1.5-5 [13], we computed the ellipsoid great circle distance between a patient's home and a UWM hospital based on the longitude and latitude coordinates of the patient's 5-digit home address zip code and the hospital's address. This distance serves as a rough proxy of the travel distance between the patient's home and hospital, is easy to compute, and is sufficient for our patient identification purpose, as shown in the Results section. For other researchers wanting to adopt our constraint-based patient identification method for their studies, using zip codes instead of exact patient home addresses can facilitate data acquisition because a limited dataset is easier to obtain than an identified one.

We compared performance across the 10 candidate constraints for identifying patients likely to receive most of their care from UWM. We used administrative data in UWM's enterprise data warehouse to check whether a patient satisfied a specific constraint. For each candidate constraint, we computed the percentage of UWM patients satisfying it. For all patients satisfying the constraint, we used PreManage data to compute the percentage of their hospital encounters that occurred within UWM in the following 6 months (April 1, 2017, to September 30, 2017). As hospital encounters are usually much more expensive than other encounters, this percentage reflects the portion of these patients' care received from UWM. In computing this percentage, every patient satisfying the constraint was included, regardless of whether the patient had ≥1 hospital encounter in the following 6 months. When selecting the final constraint to be used, we struck a balance between the following 2 criteria:

> Criterion 1: The percentage of UWM patients satisfying the constraint should be as large as possible. As multiple data analysis tasks will be conducted on these patients, this will maximize the usefulness of the applications based on these tasks.
>
> Criterion 2: For the patients satisfying the constraint, the percentage of their hospital encounters that occurred within UWM should be as large as possible. This is to maximize the degree of completeness of the medical data that UWM has on these patients. For the data analysis tasks that will be conducted on these patients, this degree impacts the biases in the analysis results.

As mentioned in the Discussion section, the selected constraint has a special property, increasing our confidence that the patients identified by the constraint also tend to incur most of their outpatient visits within UWM.

To show how the constraint-based method works for individual UWM hospitals, for all patients satisfying the selected constraint and each of the 3 UWM hospitals, we used PreManage data to compute the percentage of these patients' hospital encounters that occurred at the UWM hospital in the following 6 months.

Ethics Approval

The institutional review board of UWM reviewed and approved this study and waived the need for informed consent for all patients.

Results

Table 1 shows the demographic characteristics of our patient cohort.

Figures 1 and 2 show the percentage of UWM patients satisfying each of the 10 candidate constraints. The percentage increases with *r*, initially quickly when *r* is small and then more slowly as *r* becomes larger. Recall that *r* is the maximum allowed distance in miles between the patient's home and the closest UWM hospital. As UWM mainly serves the Seattle metropolitan area, 88.92% (309,483/348,054) of UWM patients live within 60 miles of at least one of the 3 UWM hospitals. About 44.76% (138,530/309,483) of these patients live within 5 miles.

For each of the 10 candidate constraints and all the patients satisfying it, Figures 3 and 4 show the percentage of their hospital encounters that occurred within UWM in the following 6 months. With a few exceptions when *r* is small, as *r* increases, the percentage decreases, initially quickly when *r* is small and then more slowly as *r* becomes larger. This is consistent with our intuition that with everything else being equal, patients living further from UWM hospitals are less likely to use them. Regardless of how small *r* is, this percentage never approaches 100%, partly because UWM patients could also use several non-UWM hospitals that are within 1 mile of certain UWM hospitals. As we want this percentage to be as large as possible, we should choose *r* to be ≤5 or ≤10, depending on the constraint.

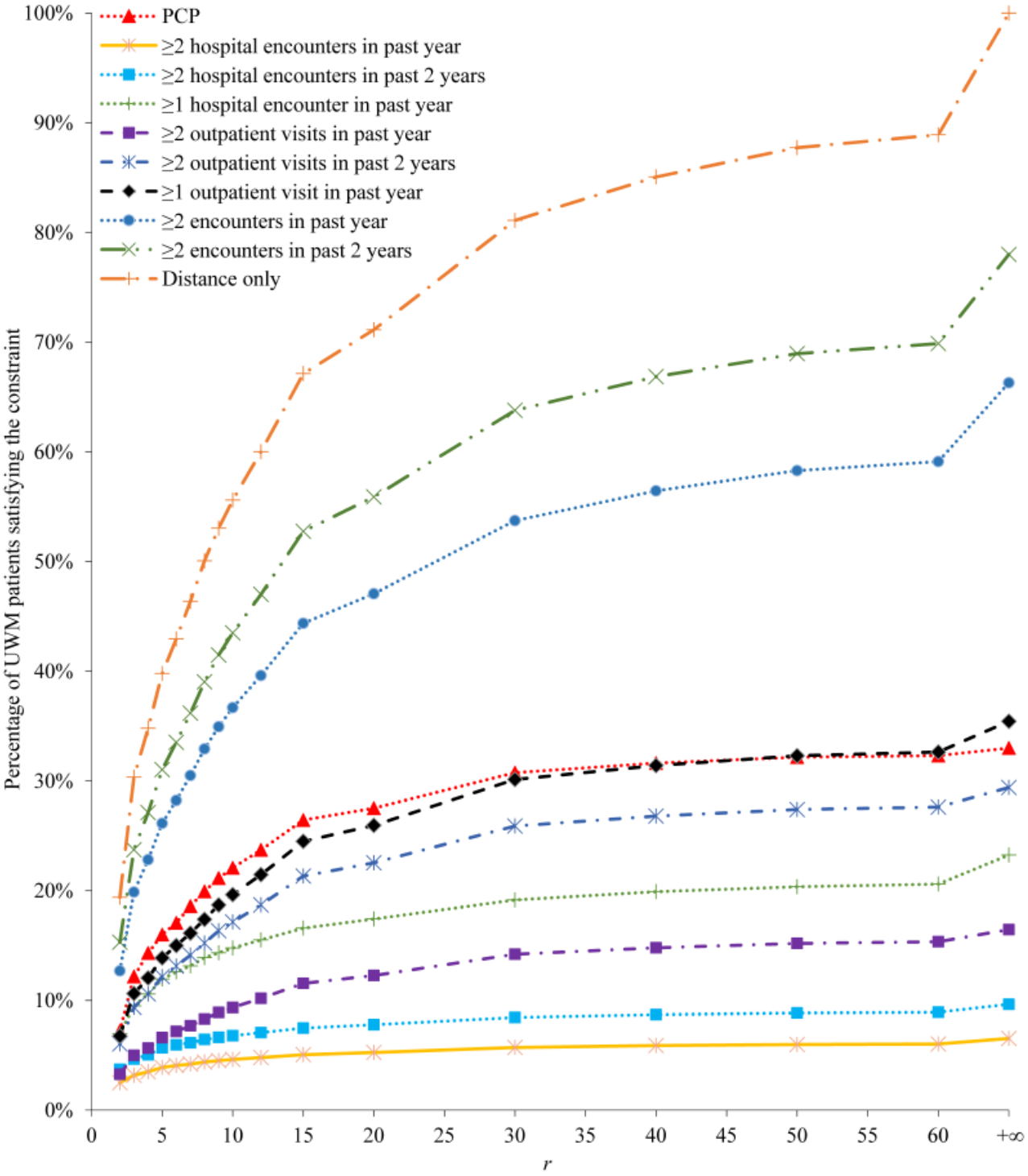
XSL•FO
RenderX

Table 1. Demographic characteristics of adult patients who had encounters at the University of Washington Medicine (UWM) facilities with information stored in UWM's enterprise data warehouse during April 1, 2016, to March 31, 2017 (N=348,054).

Demographic characteristics	n (%)
Age in years	
18 to <30	64,311 (18.48)
30 to <65	209,033 (60.06)
≥65	74,710 (21.47)
Gender	
Male	154,511 (44.39)
Female	193,506 (55.60)
Unknown or not reported	37 (0.01)
Race	
American Indian or Alaska native	5107 (1.47)
Asian	33,595 (9.65)
Black or African American	25,224 (7.25)
Multiple races	2772 (0.80)
Native Hawaiian or other Pacific islander	2500 (0.72)
White	232,354 (66.76)
Unknown or not reported	46,502 (13.36)
Ethnicity	
Hispanic	21,436 (6.16)
Non-Hispanic	270,043 (77.59)
Unknown or not reported	56,575 (16.25)
Insurance	
Private	167,640 (48.16)
Public	154,008 (44.25)
Self-paid or charity	26,406 (7.59)

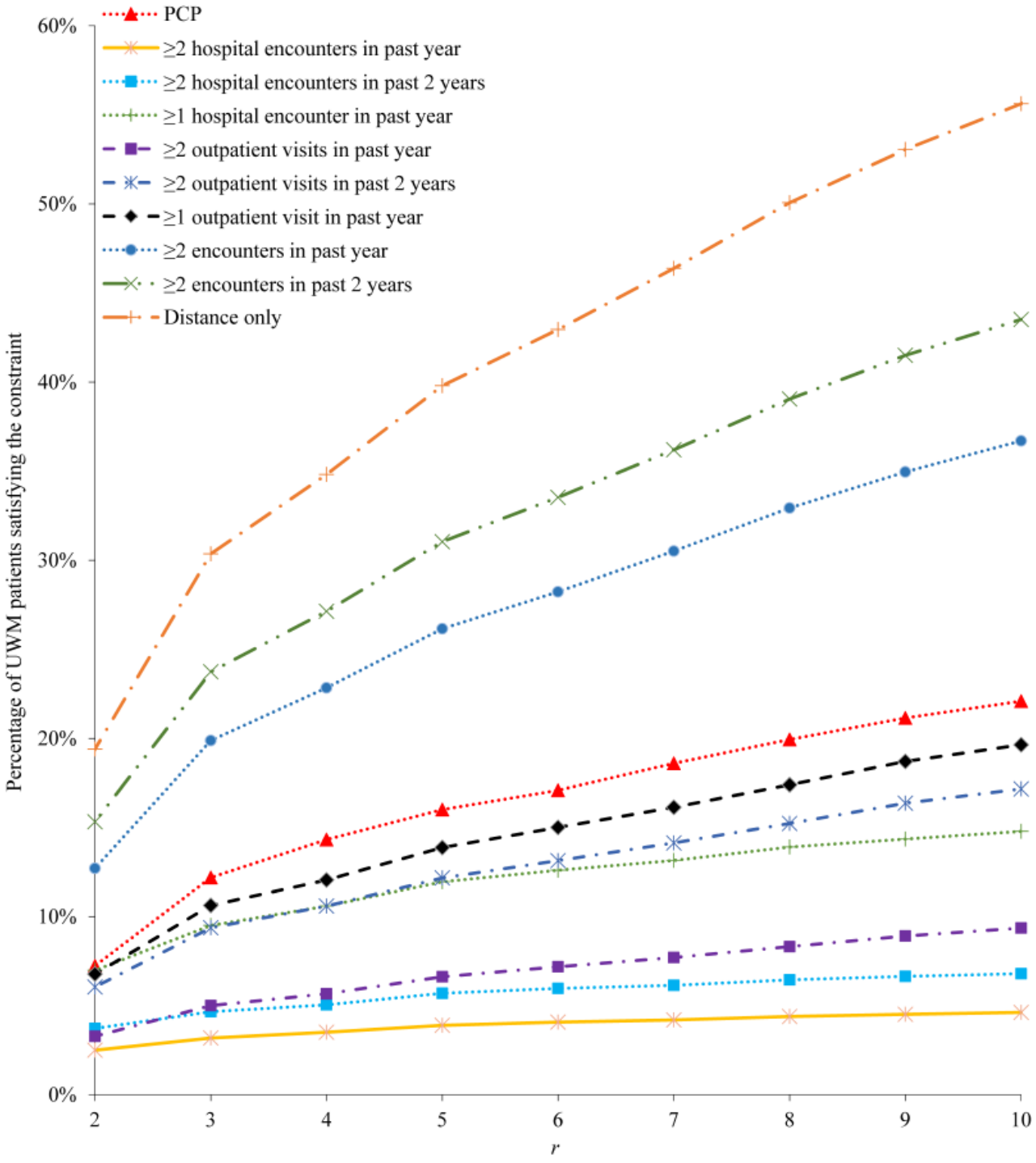
XSL•FO
RenderX

Figure 1. The percentage of University of Washington Medicine (UWM) patients satisfying each of the 10 candidate constraints. PCP: primary care physician.



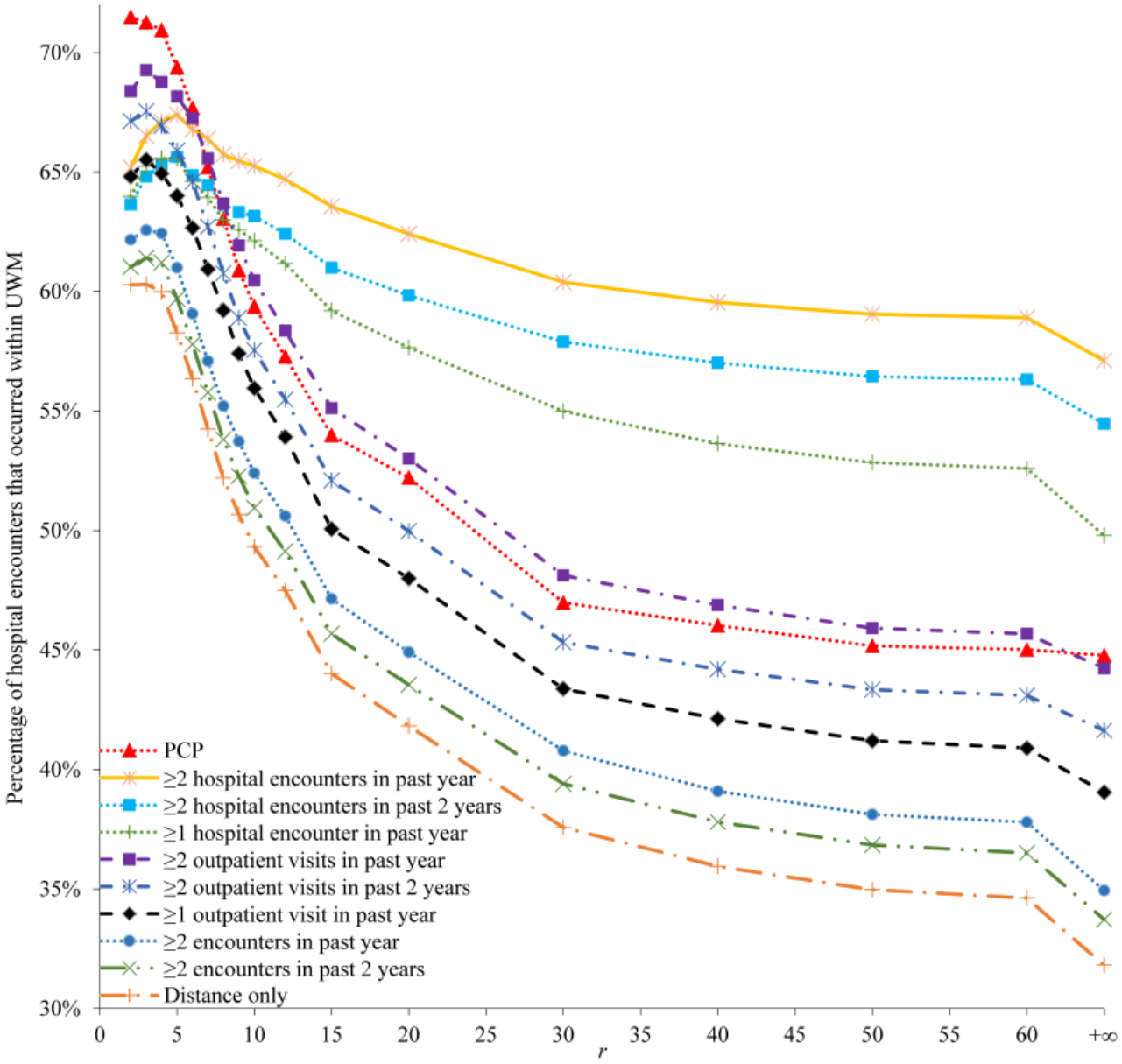
XSL•FO
RenderX

Figure 2. The percentage of University of Washington Medicine (UWM) patients satisfying each of the 10 candidate constraints, when r is ≤10. PCP: primary care physician.



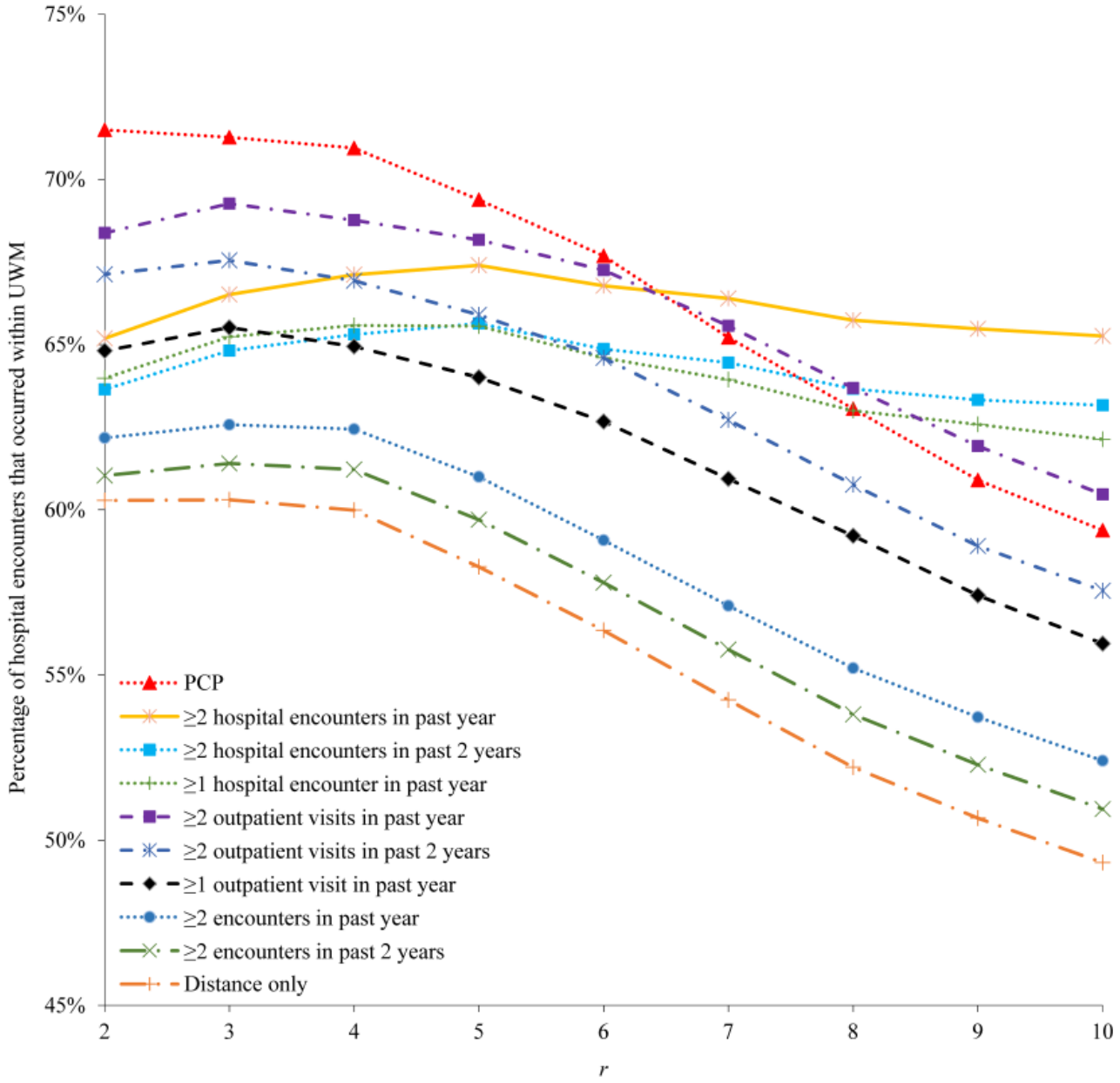
XSL•FO
RenderX

Figure 3. The percentage of hospital encounters that occurred within the University of Washington Medicine (UWM) in the following 6 months for each of the 10 candidate constraints and all patients satisfying it. PCP: primary care physician.



XSL•FO
RenderX

Figure 4. The percentage of hospital encounters that occurred within the University of Washington Medicine (UWM) in the following 6 months for each of the 10 candidate constraints and all patients satisfying it, when r is ≤10. PCP: primary care physician.

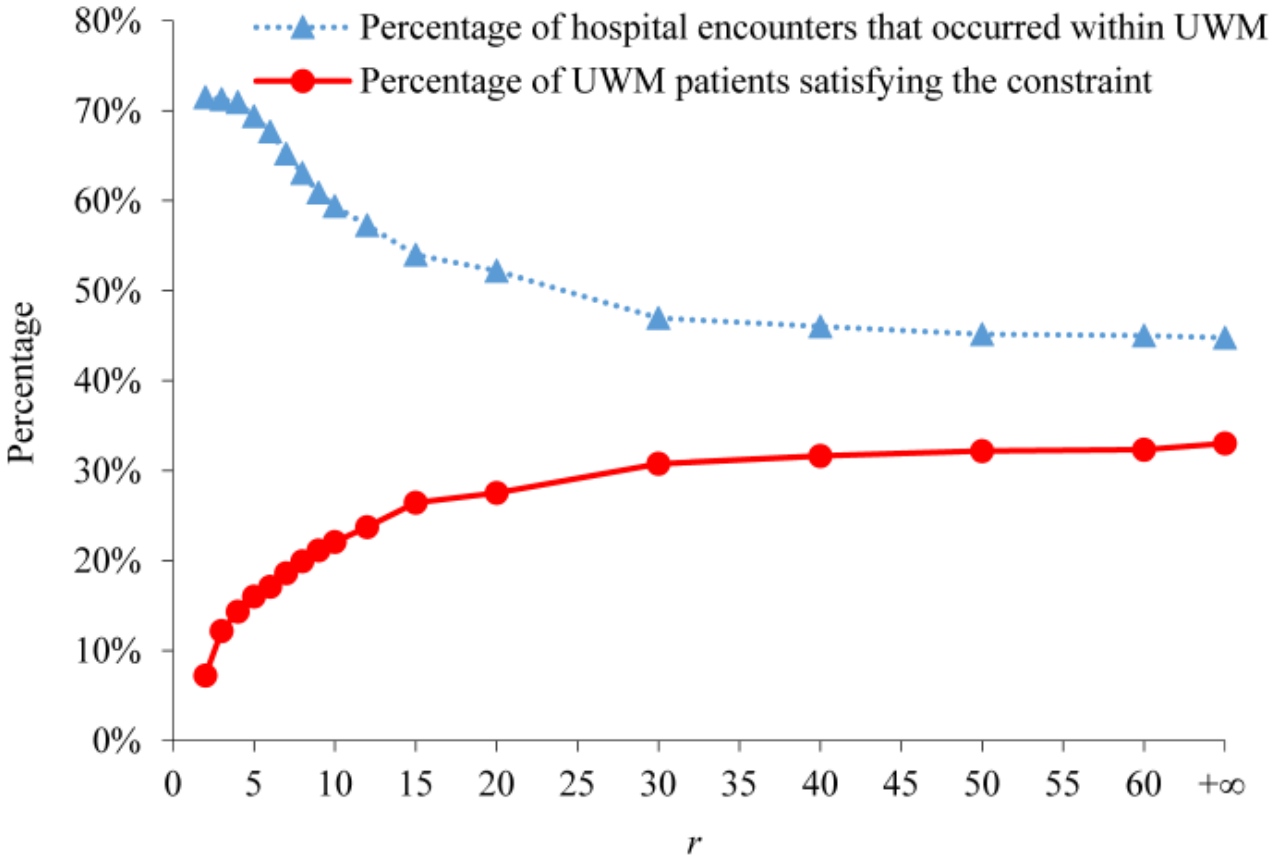


In selecting the final constraint to be used, we struck a balance between Criteria 1 and 2 listed at the end of the Methods section. The *PCP* constraint significantly outperforms 6 other constraints on Criterion 1. When r is ≤6, the *PCP* constraint outperforms all the other constraints under Criterion 2. Also, when $r=+\infty$ and distance is no longer a factor, no constraint outperforms the *PCP* constraint with r ≤6 under Criterion 2. Figure 5 shows the percentage of UWM patients satisfying the *PCP* constraint as well as the percentage of these patients' hospital encounters that occurred within UWM in the following 6 months. When the *PCP* constraint was used with r=5, 16.01% (55,707/348,054) of UWM patients satisfied the constraint. For these patients, 69.38% (10,501/15,135) of their hospital encounters occurred within UWM in the following 6 months. In comparison, for all UWM patients, 31.80% (39,171/123,162) of their hospital encounters occurred within UWM in the following 6 months.

For each of the 3 UWM hospitals and all patients satisfying the *PCP* constraint, Figure 6 shows the percentage of their hospital encounters that occurred at the UWM hospital in the following 6 months. The percentage varies across the 3 UWM hospitals. As r increases, the percentage decreases at similar rates across the 3 UWM hospitals.

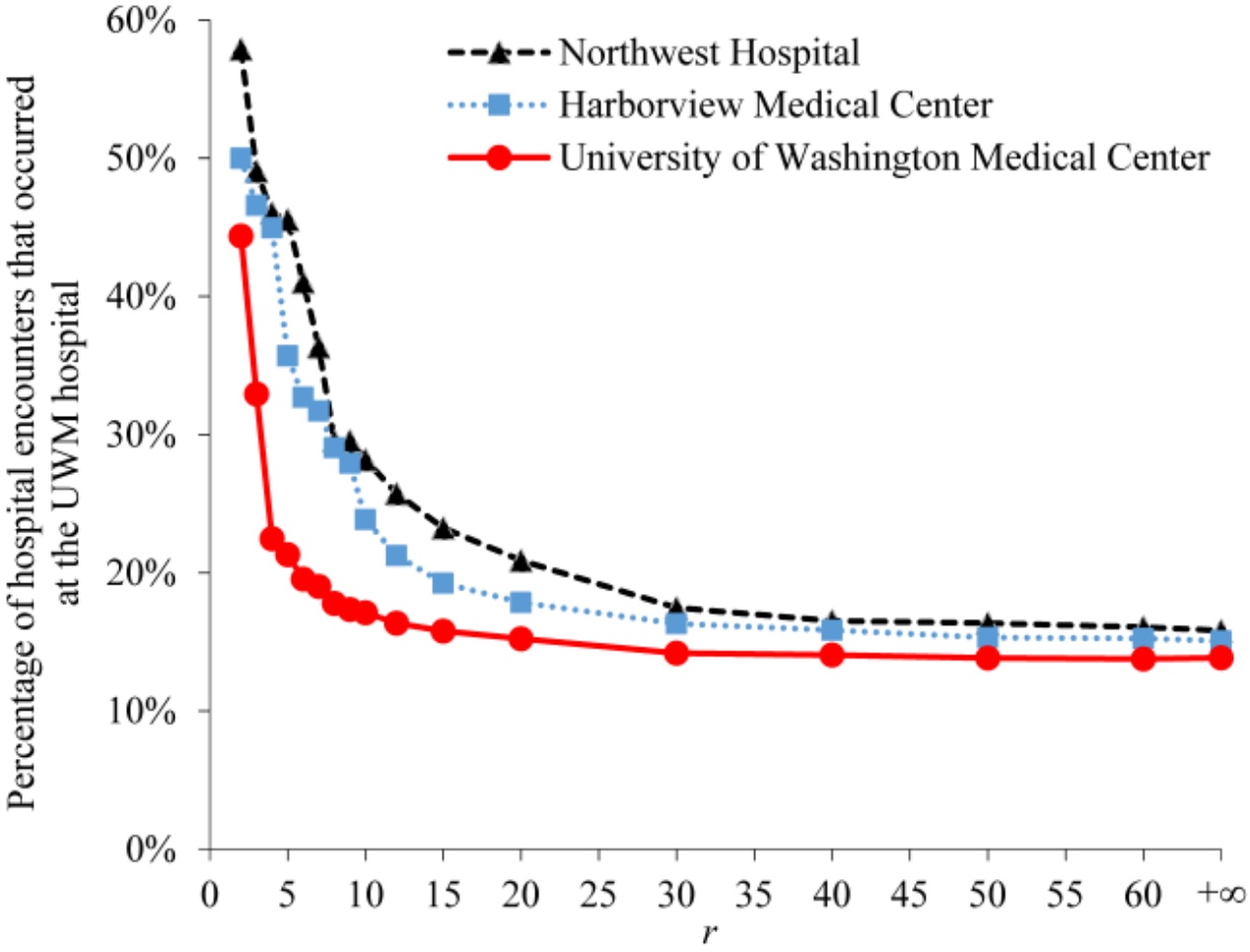
XSL•FO
RenderX

Figure 5. The percentage of University of Washington Medicine (UWM) patients satisfying the PCP constraint and the percentage of these patients' hospital encounters that occurred within UWM in the following 6 months. PCP: primary care physician.



XSL•FO
RenderX

Figure 6. For each of the 3 University of Washington Medicine (UWM) hospitals and all patients satisfying the PCP constraint, the percentage of their hospital encounters that occurred at the UWM hospital in the following 6 months. PCP: primary care physician.



Discussion

Principal Findings

By striking a balance between Criteria 1 and 2, we chose the *PCP* constraint with r=5 as the final one to be used. Using our constraint-based method to identify the right subset of patients, we more than doubled the percentage of patient hospital encounters that occurred within UWM in the following 6 months: from 31.80% (39,171/123,162) to 69.38% (10,501/15,135). Moreover, as each identified patient has a UWM PCP, we are confident that the identified patients incurred most of their outpatient visits within UWM in the following 6 months, even if we do not have data to verify this.

Potential Use of Our Results

Our results show that for patients living within 5 miles of at least one of the 3 UWM hospitals, UWM provides most of their care and has reasonably complete medical data on them. For a data analysis task requiring relatively complete data, such as predictive modeling using historical data, we can conduct the task on this subset of patients and obtain useful results, even if conducting the task on all UWM patients is impractical. For example, we can build a predictive model to identify future high-cost patients among this subset [3]. Enrolling such patients in care management can help prevent high costs and improve outcomes [5-7].

Our results show that patients living further from the 3 UWM hospitals tend to receive a smaller portion of their care from UWM. This suggests UWM to consider using different preventive interventions for patients living at differing distances from the UWM hospitals, for example, for care management to achieve better results. For patients who will receive only a small portion of their care from UWM, it is difficult for UWM to use expensive preventive interventions in a cost-effective manner.

This study used PreManage data covering adult patients in all age groups. It cannot be done using Medicare claims data that mainly cover patients aged ≥65 years and patients with certain disabilities and diseases. Similar to many other health care systems, UWM does not have complete claims data covering all its patients' health care use both within and outside of UWM. We could use claims data to do a similar study for another health care system if it has complete claims data covering all its patients' health care use both within and outside of that system.

This study used PreManage data to validate the *PCP* constraint's effectiveness for UWM. However, the *PCP* constraint does not depend on PreManage data's availability and can be used by another health care system even if it cannot access PreManage data. In this case, one way to estimate the *PCP* constraint's

XSL•FO
RenderX

effectiveness is to survey some of its patients about their health care use both within and outside of the system.

This paper focuses on identifying patients likely to receive most of their care from a specific health care system. If multiple health care systems exchange data, we could use a similar method to identify patients likely to receive most of their care from these health care systems combined. This could enable several data analysis tasks across these health care systems.

Limitations

This study has several limitations that can serve as interesting areas for future work:

- So far, UWM has accumulated PreManage data only over a limited period. After UWM accumulates more PreManage data, we should redo our analysis, check the percentage of patient hospital encounters that occur within UWM in the next 2 to 3 years, and see whether any of our conclusions will change.
- This study demonstrates our constraint-based patient identification method at a single health care system, UWM, which provides both inpatient and outpatient care mainly for an urban area. To understand how our method generalizes, we should repeat our analysis on several other health care systems, some mainly serving urban areas and others offering many services in rural areas, and see whether the optimal constraint will change. For a health care system offering many services in rural areas, we would expect the optimal value of r to be >5, as patients are more scattered in rural areas than in urban areas.
- For a health care system with incomplete medical data on many of its patients, we can use our method to identify a subset of patients on whom the health care system has more complete data and estimate the data's incompleteness level on this subset of patients. For a data analysis task, using incomplete data to do the analysis on this subset of patients would produce biased results, which could still be better than no result if the degree of bias is acceptable. Yet, the exact relationship between data incompleteness level and degree of bias in the analysis results is unknown. In particular, we have no idea of the threshold for the data's incompleteness level, beyond which the analysis' conclusion could become invalid. To address this issue, we can take a reasonably complete dataset from another health care system such as Kaiser Permanente, remove different portions of the dataset, and check the resulting impact on the analysis results. This will help us understand whether our method is good enough for enabling the data analysis task in the current health care system.

Conclusions

To the best of our knowledge, for a health care system with incomplete medical data on many of its patients, we provided the first method to use a geographic constraint to identify a reasonably large subset of patients who tend to receive most of their care from the system. Our results show that our method performs reasonably well at UWM. Our method opens the door for conducting several major analysis tasks on incomplete medical data, which were previously deemed impractical to undertake.

Acknowledgments

The authors thank Xinran Liu, Richard Crouch, Lisa Swan, Kristal Mauritz-Miller, Christie Benevich, Philip J Brewster, Ross Lordon, Lucy Wang, and Mika N Sinanan for helpful discussions. GL was partially supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number R01HL142503. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

GL was mainly responsible for the paper. He conceptualized and designed the study, performed literature review and data analysis, and wrote the paper. ESL, PTH, and ABW offered feedback on various medical issues, contributed to conceptualizing the presentation, and revised the paper. ESL also contributed to conceptualizing the study. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Bourgeois FC, Olson KL, Mandl KD. Patients treated at multiple acute health care facilities: quantifying information fragmentation. Arch Intern Med 2010 Dec 13;170(22):1989-1995. [doi: 10.1001/archinternmed.2010.439] [Medline: 21149756]
2. Finnell JT, Overhage JM, Grannis S. All health care is not local: an evaluation of the distribution of Emergency Department care delivered in Indiana. AMIA Annu Symp Proc 2011;2011:409-416 [FREE Full text] [Medline: 22195094]
3. Curry N, Billings J, Darin B, Dixon J, Williams M, Wennberg D. Kingsfund. Predictive Risk Project: Literature Review URL: http://www.kingsfund.org.uk/sites/files/kf/field/field_document/predictive-risk-literature-review-june2005.pdf [accessed 2015-08-10] [WebCite Cache ID 6aga2XBFC]

XSL•FO
RenderX

4. Mays GP, Claxton G, White J. Managed care rebound? Recent changes in health plans' cost containment strategies. Health Aff (Millwood) 2004;Suppl Web Exclusives:W4-427-W4-436 [FREE Full text] [doi: 10.1377/hlthaff.w4.427] [Medline: 15451964]
5. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med 2007 Dec;22(Suppl 3):391-395 [FREE Full text] [doi: 10.1007/s11606-007-0322-1] [Medline: 18026807]
6. Caloyeras JP, Liu H, Exum E, Broderick M, Mattke S. Managing manifest diseases, but not health risks, saved PepsiCo money over seven years. Health Aff (Millwood) 2014 Jan;33(1):124-131. [doi: 10.1377/hlthaff.2013.0625] [Medline: 24395944]
7. Nelson L. Congressional Budget Office. Lessons from Medicare's demonstration projects on disease management and care coordination URL: https://www.cbo.gov/sites/default/files/112th-congress-2011-2012/workingpaper/WP2012-01_Nelson_Medicare_DMCC_Demonstrations_1.pdf [accessed 2015-08-10] [WebCite Cache ID 6agZcFxD1]
8. Duncan IG. Healthcare Risk Adjustment and Predictive Modeling. Winsted, CT: ACTEX Publications Inc; 2011.
9. Ash A, McCall N. Rti. 2005. Risk assessment of military populations to predict health care cost and utilization URL: http://www.rti.org/pubs/tricare_riskassessment_final_report_combined.pdf [accessed 2015-08-10] [WebCite Cache ID 6aga7wYZC]
10. Iezzoni L. Risk Adjustment for Measuring Health Care Outcomes, 4th ed. Chicago, IL: Health Administration Press; 2013.
11. Luo G, Stone BL, Johnson MD, Tarczy-Hornoch P, Wilcox AB, Mooney SD, et al. Automating construction of machine learning models with clinical big data: rationale and methods. JMIR Res Protoc 2017 Aug 29;6(8):e175 [FREE Full text] [doi: 10.2196/resprot.7757] [Medline: 28851678]
12. Collectivemedicaltech. PreManage: Closing communication gaps to close gaps in care URL: http://collectivemedicaltech.com/what-we-do-2/premanage/ [accessed 2017-02-19] [WebCite Cache ID 6oOi8en1e]
13. Cran.r-project. geosphere: Spherical Trigonometry URL: https://cran.r-project.org/web/packages/geosphere/index.html [accessed 2018-09-17] [WebCite Cache ID 72VG1whxE]

Abbreviations

PCP: primary care physician
UWM: University of Washington Medicine

Edited by G Eysenbach; submitted 17.09.18; peer-reviewed by L Sheets; comments to author 08.10.18; revised version received 13.10.18; accepted 16.10.18; published 05.11.18.

Please cite as:
Luo G, Tarczy-Hornoch P, Wilcox AB, Lee ES
Identifying Patients Who Are Likely to Receive Most of Their Care From a Specific Health Care System: Demonstration via Secondary Analysis
JMIR Med Inform 2018;6(4):e12241
URL: http://medinform.jmir.org/2018/4/e12241/
doi:10.2196/12241
PMID:30401670

©Gang Luo, Peter Tarczy-Hornoch, Adam B Wilcox, E Sally Lee. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 05.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Benefits and Costs of Digital Consulting in Clinics Serving Young People With Long-Term Conditions: Mixed-Methods Approach

Sung Wook Kim[1], MSc, PhD; Jason Madan[1], PhD; Melina Dritsaki[2], PhD; Carol Bryce[1], PhD; Vera Forjaz[3], MSc; Joe Fraser[4], MA; Frances Griffiths[1], MBBS, PhD; Kathryn Hamilton[3], PhD; Caroline Huxley[1], PhD; Jackie Sturt[3], PhD

[1]Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry, United Kingdom

[2]Oxford Clinical Trials Research Unit, Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

[3]Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

[4]Joe's Diabetes Ltd, Coventry, United Kingdom

Corresponding Author:
Sung Wook Kim, MSc, PhD
Division of Health Sciences
Warwick Medical School
The University of Warwick
Medical School Building
Gibett Hill Road
Coventry, CV4 7AL
United Kingdom
Phone: 44 02476 151254
Email: s.w.kim@warwick.ac.uk

Abstract

Background: Since the introduction of digital health technologies in National Health Service (NHS), health professionals are starting to use email, text, and other digital methods to consult with their patients in a timely manner. There is lack of evidence regarding the economic impact of digital consulting in the United Kingdom (UK) NHS.

Objective: This study aimed to estimate the direct costs associated with digital consulting as an adjunct to routine care at 18 clinics serving young people aged 16-24 years with long-term conditions.

Methods: This study uses both quantitative and qualitative approaches. Semistructured interviews were conducted with 173 clinical team members on the impacts of digital consulting. A structured questionnaire was developed and used for 115 health professionals across 12 health conditions at 18 sites in the United Kingdom to collect data on time and other resources used for digital consulting. A follow-up semistructured interview was conducted with a single senior clinician at each site to clarify the mechanisms through which digital consulting use might lead to outcomes relevant to economic evaluation. We used the two-part model to see the association between the time spent on digital consulting and the job role of staff, type of clinic, and the average length of the working hours using digital consulting.

Results: When estimated using the two-part model, consultants spent less time on digital consulting compared with nurses (95.48 minutes; $P<.001$), physiotherapists (55.3 minutes; $P<.001$), and psychologists (31.67 minutes; $P<.001$). Part-time staff spent less time using digital consulting than full-time staff despite insignificant result ($P=.15$). Time spent on digital consulting differed across sites, and no clear pattern in using digital consulting was found. Health professionals qualitatively identified the following 4 potential economic impacts for the NHS: decreasing adverse events, improving patient well-being, decreasing wait lists, and staff workload. We did not find evidence to suggest that the clinical condition was associated with digital consulting use.

Conclusions: Nurses and physiotherapists were the greatest users of digital consulting. Teams appear to use an efficient triage system with the most expensive members digitally consulting less than lower-paid team members. Staff report showed concerns regarding time spent digitally consulting, which implies that direct costs increase. There remain considerable gaps in evidence related to cost-effectiveness of digital consulting, but this study has highlighted important cost-related outcomes for assessment in future cost-effectiveness trials of digital consulting.

XSL•FO
RenderX

(JMIR Med Inform 2018;6(4):e48) doi:10.2196/medinform.9577

KEYWORDS

digital consulting; long-term conditions; costing analysis; young people; mental health; diabetes

Introduction

Improving efficiency in the National Health Service (NHS), particularly regarding managing long-term conditions, is a major policy goal in the United Kingdom (UK). NHS aims to secure the greatest improvement in the health of the people living in England under 3 themes: equity, efficiency, and responsiveness [1]. Efficiency is sometimes misinterpreted as involving cost savings and reduced budgets, but it actually refers to maximizing the value generated by whatever resources are available to NHS in terms of the quantity and quality of health care it provides [2]. The use of digital communication for consultation, such as email, text, and mobile, has been proposed as a way of enhancing NHS efficiency [3]. Care home residents in West Yorkshire, for example, were offered a service that connects them with clinicians using a video link [4]. By consulting digitally, it is possible to ensure equitable access to NHS care for people living in geographically isolated areas [5]. Under this current trend, the use of digital consulting is expected to increase in the United Kingdom over the next few years. A particular area of interest for applying digital consulting is the management of individuals with long-term chronic conditions. Such individuals account for 80% of consultations, and 10% of patients with long-term conditions account for 55% of inpatient days [6]. The cost of inpatient days for patients with long-term conditions are expected to be a significant cost burden for NHS. In addition to this, managing long-term conditions is expensive for NHS with treatment costs accounting for approximately 69% of all health care costs in England in 2008 [7]. Given these facts, the potential gains from improving the efficiency of care for people with long-term conditions are enormous [8].

Digital communication has been assumed to be a route for delivering quality care at lower cost. However, this is not necessarily reflective of experience with health investment in information technology to date. The Watcher report warns against such assumptions, particularly over short-term horizons, and argues that financial savings may take 10 years to be realized [9]. This applies to patient-health professional digital consulting because the costs of implementing it are unknown. It is not inevitable that using email and text as examples of digital consulting will reduce the workload of clinics. It was argued that growth in email communication with patients has increased the workload of clinicians because they need to respond to more patients [10].

In terms of the impact of digital clinical communication on both patients and service providers, the evidence is somehow conflicting. One cancer trial study reported that a higher level of use for a Web-based health support system is associated with improved outcomes, such as mood and quality of life [11]. Email consultation can improve continuity of care and thus is likely to improve the quality of care and that of life [12]. Other potential benefits of digital communications include improvement in health care management and an improved patient-doctor relationship [13]. The smartphone is likely to make behavioral health therapy more interactive for patients, improving the delivery of evidence-based medicine [14]. A systematic review found that the clinical outcomes such as the HbA_{1c} level for diabetic patients and forced expiratory volume for asthmatic patients were improved when asynchronous communication such as short message service (SMS) text message was used [15]. Other studies [16,17] suggest that there is no difference between digital and usual care. Bradford et al [16] reported that there was no difference in the quality of life scores between caregivers in control and intervention groups who used a home telehealth service for pediatric palliative care. It was also reported that offering treatment using Web-based consultations for child dermatitis was not effective compared with traditional treatment, such as visiting general practitioners [17]. These examples show that there is no clear agreement in the effectiveness of consulting based on digital communication.

Although digital consulting between clinician and patient is of general interest in NHS, few studies have considered cost. Much of the research has focused only on the "effectiveness" of digital communication [18-20]. This may be because the direct costs of services of using text, mobile, or internet communication appear low [21,22]. Moreover, two systematic reviews on networked communications found that telemedicine cost is cheaper than travel costs [23,24]. Nevertheless, the impact of digital consulting on staff workload is unknown. It is possible that digital consulting will reduce the time involved in each consultation, but an alternative possibility is that it increases the volume of communication and hence increases staff workload without adding significant value to patient care. This information is important as time spent on digital consulting, if substantial, could have a significant impact on NHS.

The purpose of this study was to explore the economic impact of digital clinical consulting that occurs using email, text, mobile phone, and Web portals with young people having long-term conditions as part of the LYNC study [25,26]. The LYNC study explored how health care delivery and receipt are impacted by the use of digital consulting for patients and health professionals. To study the health economic impacts, we used a mixed-methods approach to address the following questions: What are the main drivers of the time spent on digital communication? What are the health benefits of digital communication as perceived by patients? Are there any benefits of digital communication beyond health?

Methods

The LYNC study focused on young people with long-term health conditions because they are more likely to disengage from health services and be associated with additional cost for NHS. The Long-term conditions Young people Networked Communication (LYNC) study attempted to look into whether their engagement can be improved using digital consulting. The

XSL•FO
RenderX

researchers working in the LYNC study both qualitatively and quantitatively interviewed clinical staff to understand how digital consulting is used to communicate with young people. The duration of this study was approximately 3.5 years. The LYNC study was an observational mixed-methods study that aimed to identify the effects, impacts, costs, patient safety, and ethical implications of digital consulting using email, text, social media, and personnel health records between health care professionals and young people living with 1 of 13 different physical or mental long-term health conditions [25].

We considered mobile phone calls differently from other phone calls because they could be initiated by young people when they needed, wherever they were located. The use of mobile phones for calls was important to some young people because it is often difficult to connect with a clinician through standard telephone systems within the UK NHS systems. The rationale for considering mobile phone calls as digital relates to the portability of mobile phones, which changes the relationship between the young person and the communication such that they have control and flexibility over the communication that took place.

Mobile phones were also valued as a communication route by clinical teams because they often found it difficult to reach young people on landlines, given that they were not often there during clinicians' working hours and may not answer the landline or respond to the messages left.

The study involved a total of 173 clinical team members, including clinicians, psychologists, psychiatrists, dieticians, and nurses, 165 young people living with long-term conditions, 13 parents, and 16 information governance specialists from 20 clinics across the UK NHS between November 2014 and March 2016. Clinics were eligible for inclusion if they provided specialist care for young people (age 16-24 years) with long-term conditions and if the clinical team had an interest in using digital consulting as part of the services provided to patients. Ethical approval for the LYNC study (14/WM/0066) was obtained from the National Research Ethics Service Committee West Midlands—The Black Country.

A health economic questionnaire (Multimedia Appendix 1) was designed with the purpose of collecting information from clinical staff members with respect to their use of digital consulting in the clinic. This questionnaire was developed following, and drawing on insights from, the initial 115 semistructured interviews conducted for the main LYNC study. The LYNC participant was asked to complete the questionnaire during the interview wherever possible, although in some cases, it was completed postinterview and returned by email. This questionnaire elicited information about the time spent using digital consulting to communicate with young patients, type of communication being used (eg, email, mobile phone calls, text, or other), staff grade, and the number of hours the staff member worked per week. Clinical staff members were asked to recollect their approximate time spent using digital consulting with patients per week in time intervals (eg, 15-30 minutes and 45-60 minutes). They were also asked about the equipment they used to communicate via digital consulting (eg, laptop, desktop, tablet, or mobile phone).

The time spent digitally consulting, as reported by clinical staff, was costed using the midpoint salary for their grade based on the NHS Agenda for Change 2014/2015 salary scales [27]. In the UK NHS system, "grade" is equivalent to "band" [28]. Grade covers all NHS staff other than doctors, dentists, and senior managers, and the 9 pay bands have multiple pay points [27]. Therefore, we used "grade" to estimate staff salary per hour.

Equipment costs were estimated using price lists provided by the University of Warwick. These costs were annualized assuming a 3.5% discount rate based on methods guidance from National Institute for Health and Care Excellence [29]. For clinical sites with questionnaire completion rates over 50%, estimates of clinic list size were requested to allow estimation of cost per patient. A health service perspective was adopted for the costing analysis in this study. In fact, the LYNC study collected health economic data from various sites with long-term conditions rather than focusing on only one disease site. Hence, the costing analysis in this study covers a wide range of costs. Among them, we could not get sufficient data from 4 sites—Dermatology, Mental health 3, Sickle cell, and Diabetes 2. This is because sites were counted only if their overall completion rate was over 50%. If less than 50% of the questionnaires were completed at a site, we did not attempt to calculate clinic-level costs.

A regression analysis was carried out to estimate the main drivers for the time spent on digital consulting activity. A two-part model [30] was employed to analyze the data collected from the clinical staff to identify any factors associated with the time spent on digital consulting. This model was chosen to reflect the fact that the time spent on using digital consulting is rightly skewed and contains many zeros [30]. The logit model was used for the first part, whereas the generalized linear model with the Poisson family was used for the second part after conducting the modified links test [31]. Log link was used in this regression. All regression analyses were carried out using Stata 14 (Statacorp, College Station, TX, US).

Mechanisms through which digital consulting may affect outcomes relevant to economic evaluation, for patients and services, was derived through the initial semistructured interviews of clinical staff. One health economist (SWK) read all interview transcriptions from staff to identify statements that reflected a consequence of digital consulting use that might lead to effects pertinent to economic evaluation. Relevant quotes were extracted to populate a thematic analysis of the purpose and content of digital consulting between clinical staff and patients; specific examples of how it was being used; the results of using digital consulting; the counterfactual showing what would have occurred if digital consulting was not available in this situation; and the incremental consequences in terms of costs and benefits to patients (whether health related). To ensure that extraction was consistent and appropriate, a sample of the transcripts from each site was independently reviewed by two other health economists (MD and JM) who then provided feedback to the primary health economic reviewer. Further information on the clinical implications of digital consulting use identified in the transcripts was provided by clinical experts in the LYNC study team.

XSL•FO
RenderX

It was rarely possible to determine all relevant information for the economic thematic analysis from the initial interviews. Therefore, once gaps in any of the domains had been identified, follow-up interviews were conducted by one health economist (MD) with clinical leads at each of the sites to expand on the information initially identified. These interviews were structured based on a bespoke interview guide developed for each site by the health economics team based on identified gaps in the thematic analysis.

Results

Quantitative Findings

A total of 115 staff (66.5%; 115/173) supplied health economic questionnaire data from 18 clinical sites. Table 1 provides a descriptive summary of these responses, broken down by channel, site, and role.

Email (mean=30 minutes; interquartile range [IQR]=0-45) and mobile (mean=20 minutes; IQR=0-30) were preferred methods for digital consulting. Staff reported that they use social media less (mean=3 minutes; IQR=0) than other channels.

Table 1. Breakdown of digital consulting use by channel, site, and staff role (minutes per day).

Breakdown of use	N	Mean	Interquartile range
Channel			
Email	115	30	0-45
Text	115	17	0-15
Social media	115	3	0-0
Mobile	115	20	0-30
Site			
Rheumatology	8	73	23-105
Mental health 2	6	8	0-15
Mental health 1	8	184	98-225
Renal	6	7	0-10
Dermatology	2	28	25-30
Diabetes 1	6	53	45-60
Sickle cell	6	133	45-180
Mental health 3	4	21	13-30
Liver	7	104	15-120
Inflammatory bowel disease 2	4	105	0-210
Cystic fibrosis 1	11	56	0-120
Diabetes 2	2	8	0-15
Cystic fibrosis 2	3	90	0-150
Sexual health	10	107	0-165
HIV	9	26	0-45
Cancer 2	11	97	21-150
Inflammatory bowel disease 1	7	64	0-120
Cancer 1	5	86	10-165
Role			
Physiotherapist	4	120	83-158
Psychologist	11	34	0-30
Dietician	8	14	0-15
Nurse	31	120	30-180
Consultant	31	28	0-45
Other	30	105	5-165

XSL•FO
RenderX

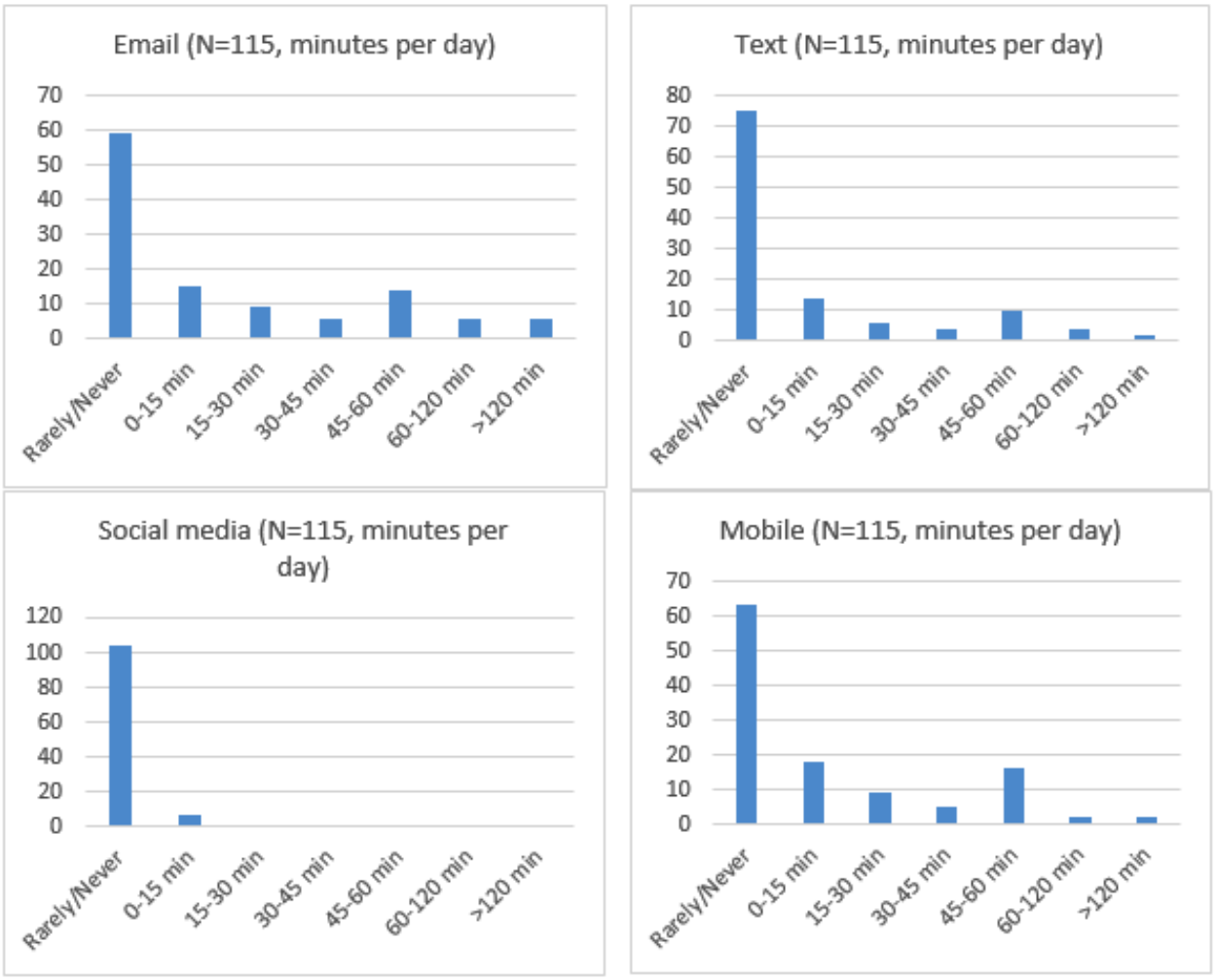
The time spent on using digital consulting was highest at the Mental health 1 site (mean=184 minutes; IQR=98-225) and lowest at the Renal site (mean=7 minutes; IQR=0-10). Staff in the Mental health 2 site (mean=8; IQR=0-15) and Diabetes 2 site (mean=8; IQR=0-15) also spent less time on digital consulting than staff in other sites.

Nurses (mean=120 minutes) and physiotherapists (mean=120 minutes; IQR=83-158) used digital consulting for approximately 2 hours; however, dieticians (mean=14 minutes; IQR=0-15) and psychologists (mean=34 minutes; IQR=0-30) used less. Psychologists reported that they use digital consulting less (mean=34 minutes; IQR=0-30) than nurses or physiotherapists.

The type of digital consulting used for clinical communication is illustrated in Figure 1. With respect to text, 34.8% (40/115) of the staff reported that they use a text, and 15% (6/40) of them mentioned that the time for using text as part of communicating with patients accounted for over 60 minutes per day. On the other hand, social media had rarely been used, that is, 9.6% (11/115) of staff replied that they use social media to communicate with patients. This result suggests that the use of digital consulting is mainly concentrated on email, SMS text messages and mobile phone than social media. One clinic (Rheumatology) did not use digital consulting at all, and a second clinic (Renal) reported just 7 minutes of use per day. The heaviest user consulted digitally with patients for more than 2 hours per day.

A clinic-level costing analysis was carried out to estimate the direct burden of digital consulting to the clinic, and the result is presented in the main clinical paper of the LYNC project [25]. The total cost was highest at one of the two Child and Adolescent Mental Health Services (£9560) and lowest at the Renal site (£161) except for the Rheumatology site (£0). The interviewed staff in the Rheumatology site reported that they never use digital consulting to communicate with patients. The total cost of using digital consulting per staff member ranges from £27 (Renal) to £1195 (Mental health 1) per month. The average cost per patient was particularly high at the Cystic fibrosis sites. The average cost per patient was £130 and £73 at the Cystic fibrosis 2 site and Cystic fibrosis 1 site, respectively. The next highest cost per patient was £16 per month (Renal clinic). The data suggest substantial variation in digital consulting use and therefore accounts for sites managing patients with the same condition. For instance, 2 Mental health sites showed a significant difference in the total cost between sites (£9560 vs £230); likewise, the total cost was different at the Cystic fibrosis (£5706 vs £1559) and Cancer (£3017 vs £6357) sites.

Figure 1. Breakdown of time spent by staff in digital consulting with young people (minutes per day) by channel.



XSL•FO
RenderX

Table 2. Factors associated with time spent on using digital consulting: Two-part model analysis (N=81).

Job role[a]	Coefficients from the two-part model (SE)	*P* value
Psychologist	31.67 (9.88)	<001
Physiotherapist	55.3 (14.66)	<001
Dietician	6.54 (6.39)	.31
Nurse	95.48 (9.85)	<001
Other	67.49 (10.20)	<001
Full time	38.49 (26.44)	.15
Site[b]		
Mental health 1	43.36 (24.16)	.07
Renal	−1.96 (8.25)	.81
Dermatology	−1.02 (2.51)	.68
Diabetes 1	11.18 (6.63)	.09
Sickle cell	37.91 (21.13)	.07
Mental health 3	28.77 (9.14)	<001
Liver	100.97 (10.77)	<001
Inflammatory bowel disease 2	74.89 (33.89)	.03
Cystic fibrosis 1	16.15 (10.67)	.13
Diabetes 2	18.84 (14.34)	.19
Cystic fibrosis 2	18.89 (24.34)	.44
Sexual health	65.91 (24.07)	.01
HIV	10.38 (9.95)	.30
Cancer 2	74.86 (9.56)	<001
Inflammatory bowel disease 1	74.92 (13.19)	<001
Cancer 1	80.98 (14.85)	<001

[a]Reference case: consultant.

[b]Reference case: Mental health 2 site.

Regression analysis (Table 2) showed that consultants used digital consulting less compared with other groups. All other groups spent more time using digital consulting (*P*<.001) compared with dieticians (*P*=.31) using a reference case of "Consultant." Nurses, physiotherapists, and psychologists used digital consulting more than consultants by 95.48, 55.3, and 31.67 minutes, respectively. Full-time staff spent more time using digital consulting than part-time staff, but the result was not statistically significant (*P*=.15). Time spent using digital consulting varied across sites, and no clear tendency in using digital consulting was found. Overall, these results show that being a consultant was a significant (negative) predictor of time spent on digital consulting.

Qualitative Findings

The qualitative component of the LYNC study [25] provided evidence of mechanisms through which digital communication could generate health benefits. The standard qualitative interviews, based on initial findings on health benefits of digital consultation to young people, identified some mechanisms through which these benefits could be interpreted as being cost-beneficial to the health care system, clinics, and individual patients. We also anticipated these potential economic benefits of digital consulting to signpost a number of important outcomes for health economic assessment in future cost-effectiveness trials.

Potential Routes to Economic & Clinical Impact

Prevention of Adverse Events

Staff and young people reported several examples of how the use of digital consulting enabled services to identify crises early by providing a channel for communication between appointments:

> *It saved my life basically. And it's made life more bearable for me not to do anything [self-harm], so...It's made me open up more to my therapist and stuff, and people on the team. I need to know that they're there and I can speak to them if I need it. Mental health 3 (outreach team) site.* [Young person 01]

This allowed for timelier management of adverse events such as self-harm associated with mental health, which could lead to improved health outcomes thus potentially preventing the

XSL•FO
RenderX

health care provider from spending more costs on patients' management.

Improved Well-Being

Young people reported psychological benefits from using digital consulting beyond its direct impact on their health. The use of digital consulting can reduce the stress of accessing clinical staff given that a SMS text message can be left at any time day or night, which was considered beneficial by some young people. For other young people, the use of digital consulting improved their access to social and emotional support or social services, which they thought contributed to reduced anxiety. The convenience of using email to contact a busy clinician is articulated by this young person:

> *So there's been emailing with the like dietician to see if it's okay to sort of start back on that. Within the week definitely, like it will never be more than like a couple of days to wait for a reply.* [Diabetes 1 Young person 13]

In addition to getting a quick reply is the importance of being able to ask questions between standard appointments, "...then I'm not, like, waiting six months to ask whatever questions I have. [Arthritis Young person 13]"

In other words, asynchronicity eased communication between health professionals and patients, which led to young people being able to undertake timely self-care activities which resulted in improved their well-being.

Time and Cost Savings for Users

Service users reported that digital consulting saved time and reduced their costs, for example, by reducing their visits to an accident and emergency department "I would probably still come over to A&E as I have done many years before if I did not have their mobile phones numbers. [Sickle cell Young person 08]" or through reducing the number of clinical visits, which was particularly important for those traveling long distances:

> *Yeah, it would mean that I don't have to come into clinic all the time. Save parking, save traffic, all that sort of thing.* [Liver Young person 18]

These statements provide specific examples of how digital consulting can be time- and cost-saving for young people.

Benefits of Using Digital Consulting for the Clinics

Improving Efficiency

It was identified that the use of digital consulting can improve efficiencies from the NHS perspective. Interviewees from a number of sites mentioned that digital consulting is commonly used to book an appointment, and interviewees from a few sites mentioned that the use of digital consulting can reduce missed appointments. The ability to replace some clinic visits, as described above, has benefits for services as well as users. A senior clinician at a diabetes clinic estimated that digital consulting saved approximately 8-12 visits per service user per year across a clinic of 686 patients.

Effect on Workload

Although many of the staff interviewed reported that using digital consulting increased their workload, this was not necessarily viewed negatively because they also felt that it improved patient care. Several sites reported that users were more likely to contact the service via digital consulting where they previously would have waited until the next clinic appointment. Digital communications can help young people who would otherwise be lost to the service. Some young people do not like synchronous communication but were willing to communicate asynchronously using digital technology.

Clinical staff saw the value of offering alternative communication pathways for some patients. It was seen as an important two-way communication mechanism to remind young persons that they had not been forgotten and that it is okay for them to communicate in any way they felt comfortable with.

> *So it's just a little, hi, you know, you're on our mind even though you're not on the ward, kind of thing, to keep that continuation of care going. They like that.* [Cancer 2 Support worker 05]

This was primarily a consequence of digital consulting enabling ease of access and therefore generating user expectations around staff availability and timeliness of response.

Discussion

To the best of our knowledge, this study is the first study to employ both qualitative and quantitative approaches to evaluate the benefits and costs of digital consulting in the UK NHS. Using this hybrid approach, it could capture the perceived benefits of using digital consulting for both clinics and patients. The findings of this study provide insights into the likely impact of digital consulting use across various long-term conditions in the UK NHS for young patients.

This study found significant variation in the time spent on digital consulting. Nursing staff were the greatest users of digital communication with the highest costs associated with their working time. It was also found that staff time is likely to be the key driver of the immediate cost impact of digital consulting. Staff time spent on digital consulting varied widely among health care workers with consultants spending less time on digital consulting to communicate with patients compared with other clinical staff members.

There are limitations that need to be noted in this study. First, the time spent on digital consulting is based on the self-reported questionnaires. In other words, this is based on self-estimation of the time spent by the clinical team members rather than the actual measurement of time spent at each site. As a result, the analysis result may be associated with measurement bias for the time spent on digital consulting. We suspect that clinical staff could have underestimated the actual time spent on digital consulting because they may not recall all their activities involving digital consulting methods. To overcome this issue, we would recommend the use of diaries or time logs for their daily activities; however, this is inherently challenging and was beyond the scope of the LYNC study.

XSL•FO
RenderX

Second, the study design did not permit a formal economic evaluation of any specific mode of digital consulting. The requirements of a formal economic evaluation are well established and include the need for clearly defined, mutually exclusive interventions with robust incremental estimates of cost and outcome data [32]. Given the design of the LYNC study, it was simply not possible to conduct a conventional economic analysis of this sort. As a result, although we were able to qualitatively establish the key mechanisms through which costs and benefits result from digital consulting, we were unable to measure the net impact of digital consulting. Specifically, we were unable to quantify the incremental impact of digital consulting use on clinic costs, adverse event rates, long-term disease progression, or "did not attend" rates. These limitations result from the breadth of the study and the fact that it was cross-sectional in nature. Nevertheless, we attempted to supplement these limitations by employing a qualitative approach. The qualitative approach identified staff reports showing that digital consulting increased workload. Also, it was found that digital consulting could lead to benefits and cost savings through qualitative mechanisms.

The qualitative evidence shows that young people value improved access to clinical staff afforded by digital communications and that health professionals viewed themselves to have kept better contact with young people in general and that it was of particular value to reach out to those who found engagement through traditional modes of communication more difficult. This is mainly because digital consulting enables young people to engage in ongoing care with health service providers between appointments, which could lead to improved outcomes for young people and prevent any potential side effects such as self-harm [25]. Mental health patients can control their desire to hurt themselves by contacting clinical staff in time with digital consulting when they feel the urge to inflict self-harm. Digital consulting was also able to offer detailed and personalized information so that young people could receive more responsive care. This suggests that evaluations of digital consulting need to carefully identify and measure its impact on both ongoing management of chronic conditions and the incidence of side effects.

These findings imply that specific digital consulting interventions have the potential to be highly cost-effective. Long-term benefits of using digital consulting may well outweigh the cost of using digital consulting per patient if it is possible to prevent complications and it is not unreasonable to hypothesize that digital consulting may be a cost-effective intervention for promoting patient activation [33] with self-management and reducing complications or secondary comorbidities.

However, two important points should be considered for digital consulting interventions to be cost-effective. First, it should be noted that costs and cost savings occur in different budgets; therefore, costs to the clinic, such as staff workload, may increase, whereas cost savings accrue elsewhere in NHS. Evaluations and service implementations need to carefully assess the impact of digital consultation on staff workload and consider how this impact can be managed and potentially minimized. Second, there are important benefits to digital consulting that go beyond health and into well-being, such as a sense of control over the condition [34]. Young people generally struggle in managing their chronic conditions, facing health and lifestyle impacts beyond their control [35]. For instance, our qualitative analysis revealed that the ease of access to clinical staff, which cannot be directly captured by health-related quality of life outcomes such as quality adjusted life years, can be improved by constant contacting via digital consulting with clinical staff. Moreover, young people find digital consulting less burdensome than face-to-face consultation or direct phone calling to the clinic and therefore contact clinical staff digitally. Self-determination theory suggests that individuals require competence, autonomy, and meaningful relationships to underpin the development of intrinsic motivation [36]. Self-determination may be supported through digital consulting, and this may be the mechanism through which it impacts well-being and quality of life [37]. Consequently, evaluations need to carefully consider whether psychological outcomes that cannot be captured by quality adjusted life years are an important part of the benefits created by digital consultations and how such outcomes might be measured and valued.

In conclusion, this study identified mechanisms through which digital consulting can lead to improved efficiencies for both clinics and patients. Qualitatively, digital consulting has positive potential economic impacts for NHS, such as preventing adverse events and improving efficiency and patient well-being despite increased staff workload. This study may be regarded as a preliminary study to inform the design of future economic evaluations and service implementation plans in this area.

Acknowledgments

This project was funded by the National Institute for Health Research, Health Services and Delivery Research Programme (project number 12/209/51). The views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding body or the Department of Health. We would like to thank Morna Woods and Rebecca Rogers for their help.

Authors' Contributions

SWK carried out the analysis and drafted this paper. JM contributed to the study design, drafted this paper, and supervised data analysis. MD collected data and drafted this paper. CB, VF, KH, and CH collected qualitative data and drafted this paper. JF drafted this paper. FG and JS supervised and helped draft the paper. All authors read and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

FG reports grants from the UK National Institute for Health Research. JF reports personal fees from NHS England, outside the submitted paper.

Multimedia Appendix 1

This is a health economics questionnaire.

[PDF File (Adobe PDF File), 124KB - medinform_v6i4e48_app1.pdf]

References

1. Culyer AJ. The bogus conflict between efficiency and vertical equity. Health Econ 2006 Nov;15(11):1155-1158. [doi: 10.1002/hec.1158] [Medline: 16929485]
2. Palmer S, Torgerson DJ. Definitions of efficiency. British Medical Journal 1999;318(7191):1136.
3. Blunt I, Bardsley M. Use of patient-level costing to increase efficiency in NHS trusts. Nuffield Trust 2012.
4. HM Government. 2014. personalised health and care 2020: a framework for action URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf [accessed 2018-09-26] [WebCite Cache ID 72iqLv8M6]
5. Richards H, King G, Reid M, Selvaraj S, McNicol I, Brebner E, et al. Remote working: survey of attitudes to eHealth of doctors and nurses in rural general practices in the United Kingdom. Fam Pract 2005 Feb;22(1):2-7. [doi: 10.1093/fampra/cmh716] [Medline: 15642724]
6. Wilson T, Buck D, Ham C. Rising to the challenge: will the NHS support people with long term conditions? BMJ 2005 Mar 19;330(7492):657-661 [FREE Full text] [doi: 10.1136/bmj.330.7492.657] [Medline: 15775000]
7. Henderson C, Knapp M, Fernández JL, Beecham J, Hirani SP, Cartwright M, Whole System Demonstrator evaluation team. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. BMJ 2013 Mar 20;346:f1035 [FREE Full text] [doi: 10.1136/bmj.f1035] [Medline: 23520339]
8. NHS England. Five year forward view. URL: https://www.england.nhs.uk/five-year-forward-view/ [accessed 2018-09-26] [WebCite Cache ID 72iqgU0Ys]
9. Wachter RM. Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England Report of the National Advisory Group on Health Information Technology in England. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [WebCite Cache ID 73X1THfIA]
10. Hobbs J, Wald J, Jagannath YS, Kittler A, Pizziferri L, Volk LA, et al. Opportunities to enhance patient and physician e-mail contact. Int J Med Inform 2003 Apr;70(1):1-9. [Medline: 12706177]
11. DuBenske L, Gustafson DH, Shaw BR, Cleary JF. Web-based cancer communication and decision making systems: connecting patients, caregivers, and clinicians for improved health outcomes. Med Decis Making 2010;30(6):732-744 [FREE Full text] [doi: 10.1177/0272989X10386382] [Medline: 21041539]
12. 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]
13. Patt M, Houston TK, Jenckes MW, Sands DZ, Ford DE. Doctors who are using e-mail with their patients: a qualitative exploration. J Med Internet Res 2003;5(2):e9 [FREE Full text] [doi: 10.2196/jmir.5.2.e9] [Medline: 12857665]
14. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: Integrating smartphone technology in behavioral healthcare. Professional Psychology: Research and Practice 2011;42(6):505-512. [doi: 10.1037/a0024485]
15. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
16. Bradford N, Young J, Armfield NR, Bensink ME, Pedersen L, Herbert A, et al. A pilot study of the effectiveness of home teleconsultations in paediatric palliative care. J Telemed Telecare 2012 Dec;18(8):438-442. [doi: 10.1258/jtt.2012.gth103] [Medline: 23155114]
17. Bergmo TS, Wangberg SC, Schopf TR, Solvoll T. Web-based consultations for parents of children with atopic dermatitis: results of a randomized controlled trial. Acta Paediatr 2009 Feb;98(2):316-320. [doi: 10.1111/j.1651-2227.2008.01033.x] [Medline: 18795905]
18. Koshy E, Car J, Majeed A. Effectiveness of mobile-phone short message service (SMS) reminders for ophthalmology outpatient appointments: observational study. BMC Ophthalmol 2008 May 31;8(1):9 [FREE Full text] [doi: 10.1186/1471-2415-8-9] [Medline: 18513438]
19. Menon-Johansson AS, McNaught F, Mandalia S, Sullivan AK. Texting decreases the time to treatment for genital Chlamydia trachomatis infection. Sex Transm Infect 2006 Feb;82(1):49-51 [FREE Full text] [doi: 10.1136/sti.2004.014381] [Medline: 16461603]

XSL•FO
RenderX

20. Lim M, Hocking JS, Aitken CK, Fairley CK, Jordan L, Lewis JA, et al. Impact of text and email messaging on the sexual health of young people: a randomised controlled trial. J Epidemiol Community Health 2012 Jan;66(1):69-74. [doi: 10.1136/jech.2009.100396] [Medline: 21415232]
21. Guerriero C, Cairns J, Roberts I, Rodgers A, Whittaker R, Free C. The cost-effectiveness of smoking cessation support delivered by mobile phone text messaging: Txt2stop. Eur J Health Econ 2013 Oct;14(5):789-797 [FREE Full text] [doi: 10.1007/s10198-012-0424-5] [Medline: 22961230]
22. Foley J, O'Neill M. Use of mobile telephone short message service (SMS) as a reminder: the effect on patient attendance. Eur Arch Paediatr Dent 2009 Jan;10(1):15-18. [Medline: 19254521]
23. Sutcliffe P, Martin S, Sturt J, Powell J, Griffiths F, Adams A, et al. Systematic review of communication technologies to promote access and engagement of young people with diabetes into healthcare. BMC Endocr Disord 2011 Jan 06;11:1 [FREE Full text] [doi: 10.1186/1472-6823-11-1] [Medline: 21210964]
24. Martin S, Sutcliffe P, Griffiths F, Sturt J, Powell J, Adams A, et al. Effectiveness and impact of networked communication interventions in young people with mental health conditions: a systematic review. Patient Educ Couns 2011 Nov;85(2):e108-e119. [doi: 10.1016/j.pec.2010.11.014] [Medline: 21239133]
25. Griffiths F, Bryce C, Cave J, Dritsaki M, Fraser J, Hamilton K, et al. Timely Digital Patient-Clinician Communication in Specialist Clinical Services for Young People: A Mixed-Methods Study (The LYNC Study). J Med Internet Res 2017 Dec 10;19(4):e102 [FREE Full text] [doi: 10.2196/jmir.7154] [Medline: 28396301]
26. Griffiths FE, Atherton H, Barker JR, Cave JA, Dennick K, Dowdall P, et al. Improving health outcomes for young people with long term conditions: The role of digital communication in current and future patient-clinician communication for NHS providers of specialist clinical services for young people - LYNC study protocol. Digit Health 2015;1:2055207615593698 [FREE Full text] [doi: 10.1177/2055207615593698] [Medline: 29942543]
27. NHS. 2015. Agenda for Change pay scales for 2014-2015 2014 URL: https://www2.rcn.org.uk/support/pay_and_conditions/pay_rates_2014-15 [accessed 2018-09-26] [WebCite Cache ID 72iqi9t8s]
28. Couchman, T. 2017. Analysis of Gender Pay Banding at the United Lincolnshire Hospitals NHS Trust URL: https://www.ulh.nhs.uk/content/uploads/2017/01/Gender_Pay_Banding_at_ULHT_2016_Analysis_FINAL.pdf [accessed 2018-09-26] [WebCite Cache ID 72iqjBOk8]
29. NICE. 2013. Update to the NICE technology appraisals methods guide (2013) URL: http://www.nicedsu.org.uk/NICE-Methods-Guide-updates(1985333).htm [accessed 2018-09-26] [WebCite Cache ID 72iqjivWJ]
30. Belotti F, Deb P, Manning WG, Norton EC. twopm: Two-part models. Stata Journal 2015;15(1):3-20 [FREE Full text]
31. Glick H, Doshi JA, Sonnad SS. In: Polsky D, editor. Economic Evaluation in Clinical Trials: Second Edition. Oxford, United Kingdom: Oxford University Press; 2014.
32. Widmer R, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
33. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
34. McCann TV, Clark E. Advancing self-determination with young adults who have schizophrenia. J Psychiatr Ment Health Nurs 2004 Feb;11(1):12-20. [Medline: 14723634]
35. Olsson C, Bond L, Johnson MW, Forer DL, Boyce MF, Sawyer SM. Adolescent chronic illness: a qualitative study of psychosocial adjustment. Ann Acad Med Singapore 2003 Jan;32(1):43-50. [Medline: 12625097]
36. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: Interventions based on self-determination theory. European Health Psychologist 2008;10(1):2-5 [FREE Full text]
37. Lachapelle Y, Wehmeyer ML, Haelewyck MC, Courbois Y, Keith KD, Schalock R, et al. The relationship between quality of life and self-determination: an international study. J Intellect Disabil Res 2005 Oct;49(Pt 10):740-744. [doi: 10.1111/j.1365-2788.2005.00743.x] [Medline: 16162119]

Abbreviations

IQR: interquartile range
LYNC: Long-term conditions Young people Networked Communication
NHS: National Health Service
SMS: short message service
UK: United Kingdom

XSL•FO
RenderX

Edited by G Eysenbach; submitted 09.01.18; peer-reviewed by J Pecina, K Machin; comments to author 24.03.18; revised version received 28.06.18; accepted 19.07.18; published 30.10.18.

Please cite as:
Kim SW, Madan J, Dritsaki M, Bryce C, Forjaz V, Fraser J, Griffiths F, Hamilton K, Huxley C, Sturt J
Benefits and Costs of Digital Consulting in Clinics Serving Young People With Long-Term Conditions: Mixed-Methods Approach
JMIR Med Inform 2018;6(4):e48
URL: http://medinform.jmir.org/2018/4/e48/
doi:10.2196/medinform.9577
PMID:30377145

©Sung Wook Kim, Jason Madan, Melina Dritsaki, Carol Bryce, Vera Forjaz, Joe Fraser, Frances Griffiths, Kathryn Hamilton, Caroline Huxley, Jackie Sturt. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 30.10.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using an Internet-Based Hospital to Address Maldistribution of Health Care Resources in Rural Areas of Guangdong Province, China: Retrospective and Descriptive Study

Cui He[1], PhD; Qiru Zhou[2], MBBS; Wei Chen[3], MD; Junzhang Tian[4], PhD; Lihua Zhou[5], MM; Hong Peng[6], PhD; Shuo Luan[1], MM; Shengfei Wang[2], MM

[1]Department of Rehabilitation, Guangdong Second Provincial General Hospital, Guangzhou, China

[2]Department of Guangdong Online Hospital, Guangdong Second Provincial General Hospital, Guangzhou, China

[3]Outpatient Department, Guangdong Second Provincial General Hospital, Guangzhou, China

[4]Guangdong Second Provincial General Hospital, Guangzhou, China

[5]Nursing Department, Guangdong Second Provincial General Hospital, Guangzhou, China

[6]Department of Otorhinolaryngology, Head and Neck Surgery, Guangdong Second Provincial General Hospital, Guangzhou, China

Corresponding Author:
Qiru Zhou, MBBS
Department of Guangdong Online Hospital
Guangdong Second Provincial General Hospital
466 Xingang Middle Road
Haizhu District
Guangzhou, 510317
China
Phone: 86 02089168777
Fax: 86 02089168017
Email: 562783227@qq.com

Abstract

Background: Health care maldistribution is a long-term problem in China. Telemedicine is an efficient way to deliver medical resources to remote areas; however, there are few studies on the effectiveness and challenges in providing health care from rural to urban areas in China.

Objective: The objective was to describe the effectiveness and challenges of telemedicine for providing health care from Guangzhou to rural areas in Guangdong Province.

Methods: We designed surveys and conducted them immediately after the consultation and 2-4 weeks later. Data were collected from June 2015 to May 2016 including patients' demographic features, patient satisfaction, medicine effect, patient compliance, acceptability of prescription expenses, patients' desire to revisit the department, the top 10 diseases, and self-reported difficulties in telemedicine experience. The monthly average prescription expense was described using a line chart. The monthly consultation and prescription, as well as monthly prescriptions of Western medicines and herbs, were described using a bar chart.

Results: Women comprised majority (45,386/67,740, 67.00%) of participants and men comprised the minority (22,354/67,740, 33.00%). The top 3 diseases were upper respiratory diseases (12,371/36,311, 34.07%), laryngopharyngitis (4938/36,311, 13.60%), and menstrual disorders (4669/36,311, 12.86%). The monthly prescription for Western medicine was much more than that for Chinese herbs. The annual average medicine expense per prescription was 62.9 ¥. The participants' perception of expense was acceptable (8775/12,450, 70.48%), mostly acceptable (2497/12,450, 20.01%), accepted but somewhat expensive (980/12,450, 7.9%), and unacceptable because of high cost (198/12,450, 1.6%). The surveys on patient satisfaction demonstrated very satisfied (55,687/67,740, 82.21%), satisfied (5891/67,740, 8.70%), basic satisfaction (3561/67,740, 5.26%), dissatisfaction (1454/67,740, 2.15%), and no comment (1147/67,740, 1.69%). Participants reported their treatment outcome as follows: full recovery (5008/12,450, 40.22%), recovering (4873/12,450, 39.14%), no effect (2153/12,450, 17.29%), or worsening (416/12,450, 3.3%). Approximately 89.01% (20,240/22,740) of participants will revisit the department, whereas 10.99% (2500/22,740) will not. Most patients complied with the doctors' advice completely (5430/10,290, 52.77%), whereas the rest reported partial compliance (3684/10,290, 35.80%) or no compliance at all (1176/10,290, 11.43%). The participants reported poor computer skills (4980/22,740, 21.90%), transportation inconvenience (4670/22,740, 20.50%), unstable internet connection (3820/22,740, 16.80%), language

XSL•FO
RenderX

barriers (3708/22,740, 16.30%), medication and medical hardware shortage (2459/22,740, 10.82%), tiring commute (2068/22,740, 9.08%), family care burdens (679/22,740, 3.0%), and other unclassified difficulties (356/22,740, 1.6%) as difficulties in using telemedicine.

Conclusions: Telemedicine has a wide disease spectrum, similar to ordinary medicine in China. It saves costs, has high patient satisfaction and price acceptability, and can relieve disease and syndromes. However, certain problems need to be resolved. Telemedicine could be a feasible approach to address the health care maldistribution in rural China. This study may provide useful information for policy making and guidance for further telemedicine practice in China and other developing countries.

(JMIR Med Inform 2018;6(4):e51) doi:10.2196/medinform.9495

KEYWORDS

telemedicine; health care delivery; prescription; cost; patient satisfaction; patient compliance

Introduction

China is regarded as a traditionally agricultural society comprised of urban and rural areas with populations of 6.6 billion and 6.7 billion, respectively [1]. The maldistribution of health care resources between rural and urban areas has been an obvious problem to date despite the launch of medical reforms by the government [2-4]. It has been reported that the nationwide hospital bed ratio in the urban population was 0.087% compared with 0.018% in the rural population. The ratio of medical professionals in the urban population was 0.1% compared with 0.016% in the rural population [5], indicating that substantially more people live in rural areas but have fewer medical resources. Poor access to medical care was so common that peasants have had to spend a whole day registering and another several hours waiting to see a doctor. Some were even ignorant about their disease until they became seriously ill.

Under these circumstances, the Chinese government proposed the development of telemedicine service as an approach to optimize and redistribute medical resources in remote places [6]. The World Health Organization definition of telemedicine is the delivery of health care services, wherein distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for the diagnosis, treatment, and prevention of diseases and injuries, research and evaluation, and for the continuing education of health care providers, all in the interest of advancing the health of individuals and their communities [7]. Previous research has confirmed the advantages of telemedicine, such as saving traveling time and reducing medical expenses [8,9], increasing patient satisfaction [9,10], effectiveness comparable to hospital face-to-face consultation [11], and improved management of chronic diseases such as diabetes and hypertension [8,12]. However, the effectiveness of telemedicine for the distribution of medical resources in rural China remains unclear. Guangdong Province is at the frontier of economic reform in China, and the Pearl River Delta region is more affluent than other rural areas.

In this study, we present the practice of telemedicine in a tertiary hospital in Guangdong Province that provides high-quality medical services to rural areas. We will describe the application, effectiveness, and difficulties of telemedicine practice for providing health care resources from Guangzhou to rural areas of Guangdong Province.

Methods

Study Design and Participants

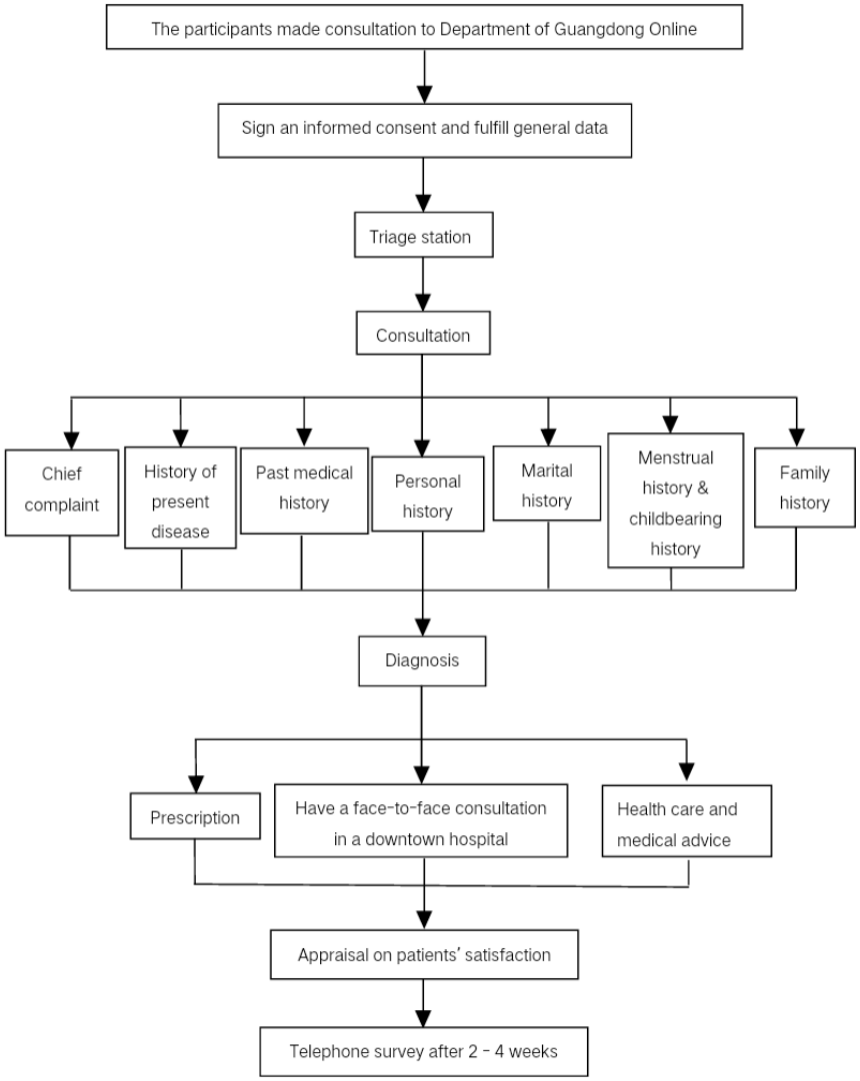
The study enrolled participants who requested telemedicine consultation and medication at rural sanitary stations connected to the Department of Guangdong Online Hospital, Guangdong Second Provincial General Hospital, from June 2015 to May 2016. Each participant was informed of what telemedicine was and of the nature of the research and voluntarily signed a consent form before consultation. The participants consulted the remote medicine appliance system and registered with their demographic data including name, sex, age, telephone number, identity card number, historical medical record, and allergic history. The triage staff at the Department of Guangdong Online Hospital referred them to a relevant doctor after assessing and recording their vital signs at the rural sanitary stations. A diagnosis was provided by the doctor, and medical care was delivered to the sanitary station. The doctors prescribed medicine to the participants (medication prescription group), gave professional advice (medical advice group), or suggested that they have a face-to-face consultation with the doctors at a downtown hospital (face-to-face consultation group) depending on the participants' diseases or syndromes. Prescribed medicines included Western medicines (Western medicine prescription group) and Chinese herbs (herb prescription group). For instance, a patient with a cough consulted with the Online Hospital at a sanitary station. The doctor diagnosed him with chronic bronchitis through inquiry, history medical record, and checking of his vital signs. The doctor prescribed medicine if he were confident of the diagnosis and considered the medicine necessary. The doctor gave professional suggestions such as the patient quit smoking and avoid allergens if he thought it was unnecessary for the patient to take medicine. However, if the doctor was unsure about the diagnosis or thought that the professional advice or medicine prescribed would not alleviate cough, he advised the patient to have a face-to-face consultation with a doctor at a downtown hospital. The communication was mainly conducted in Mandarin and Cantonese; some illiterate patients were assisted by staff at the rural sanitary stations. Satisfaction surveys were conducted immediately after the consultation, and telephone surveys on patient satisfaction, self-reported therapeutic effects, patients' desire to revisit the Department of Guangdong Online Hospital, patients' compliance with doctors' advice, and self-reported difficulties with their telemedicine experience were conducted 2 to 4 weeks after the

XSL•FO
RenderX

consultation (Figure 1). Two staff members from the department underwent advanced training to conduct telephone surveys. A full ethics review of the research protocol was conducted by the Institutional Review Board at the Guangdong Second Provincial General Hospital.

This study was launched by the Department of Guangdong Online Hospital, Guangdong Second Provincial General Hospital [13] the first telemedicine institution authorized by the Chinese government. The Department of Guangdong Online Hospital consisting of 24 clinics worked simultaneously with doctors from internal medicine, pediatrics, traditional Chinese medicine, dermatology, rehabilitation, and gynecology departments at the hospital. A medical technology company supported the Web-based consultation platform services including visual telephone, Web-based prescription services, electronic anatomic diagrams, triage, blood pressure and blood glucose measurement, Web-based auscultation, and geographic maps of the patients' locations. The telemedicine network covered approximately 300 rural sanitary stations in the rural areas of Maoming, Jieyang, Zhaoqing, Huizhou, Shantou, Zhanjiang, Zhongshan, Jiangmen, Meizhou, Yunfu, Heyuan, Yangjiang, Shaoguan, Chaozhou, and Shanwei (Figure 2 and Table 1).

Figure 1. Flowchart for telemedicine consultation with Department of Guangdong Online Hopsital, Guangdong Second Provincial General Hospital.



XSL•FO
RenderX

Figure 2. Guangdong administrative map.

XSL•FO
RenderX

Table 1. Distance to Guangzhou and population of rural regions.

Rural regions	Distance to Guangzhou (km)	Population, n
Mao ming	340.5	5,817,753
Jie yang	430.4	5,877,025
Zhao qing	97.1	3,918,085
Hui zhou	144.2	4,597,002
Shan tou	439.3	5,391,028
Zhan jiang	416.4	6,993,304
Zhong shan	83.4	3,120,884
Jiang men	91.3	4,448,871
Mei zhou	390	4,240,139
Yun fu	143.8	2,360,128
He yuan	199.6	2,953,019
Yang jiang	223.6	2,421,812
Shao guan	222.4	2,826,612
Chao zhou	498	2,669,844
Shan wei	274.2	2,935,717

Textbox 1. Participants' (medication prescription group) attitudes toward telemedicine.

Name: Sex: Age:

Identity number: Telephone number:

Residence place: Public clinic:

Please check the option that fits for you.

1. What kind of medicine did you obtain from the doctors?

Western medicine ( ); Chinese herbs ( )

2. What was the curative effect of the drug prescribed by the doctor?

Cured ( ); got better ( ); no use ( ); got worse ( )

3. What is your acceptance of the medical expense?

Completely acceptable ( ); mostly acceptable ( ); acceptable but somewhat expensive ( ); unacceptable due to high cost ( )

4. Do you think telemedicine is useful?

Very useful ( ); somewhat useful ( ); not very useful ( ); no use ( )

5. Will you visit the online hospital again?

Yes ( ); no ( )

6. What difficulties you have had in your telemedicine experience?

Survey Instruments

The survey instruments used in this study consisted of 3 evaluation charts. The charts were formulated according to the following 3 categories of service received by participants: medication prescriptions (Textbox 1), medical advice (Textbox 2), and referral to a doctor at a downtown hospital (Textbox 3). Each participant was interviewed and evaluated voluntarily according to the medical service received.

The questionnaires contained the basic information of the participants. Each chart contained 5 to 6 questions focusing on opinions and attitudes about telemedicine, including effectiveness of treatment, willingness to revisit, acceptance of medical expenses, and difficulties with the telemedicine experience.

XSL•FO
RenderX

Textbox 2. Participants' (medical advice group) attitudes toward telemedicine.

Name: Sex: Age:

Identity number: Telephone number:

Residence place: Public clinic:

Please check the option that fits for you.

1. Did you follow the doctor's advice completely?

Completely ( ); partially ( ); just listened to the advice ( )

2. What are your opinions on the usefulness of the doctor's advice?

Very useful ( ); useful ( ); no use ( ); got worse ( )

3. Do you think that telemedicine is useful?

Very useful ( ); useful ( ); not too useful ( ); no use ( )

4. Will you visit online hospital again?

Yes ( ); no ( )

5. What difficulties have you had in your telemedicine experience?

Textbox 3. Participants' (face-to-face consultation group) compliance with Web-based physician advice.

Name: Sex: Age:

Identity number: Telephone number:

Residence place: Public clinic:

Please tick the option that fit for you.

1. Did you follow the doctor's advice to have a face-to-face consultation in a downtown hospital?

Yes ( ); no ( )

2. What was the doctors' diagnosis?

3. What therapy did you receive at the downtown hospital?

Medicine ( ); surgery ( ); rehabilitation ( ); psychiatry ( ); others ( )

4. What was the outcome of the telemedicine?

Cured ( ); got better ( ); no curative effect ( ); got worse ( )

5. Do you think that telemedicine is useful?

Very useful ( ); useful ( ); no use ( ); got worse ( )

6. Would you visit the Department of Guangdong Online Hospital again?

Yes ( ); no ( )

Statistical Analysis

Five quantitative and qualitative variables were categorized and calculated according to data collected from the 15 rural regions shown in Figure 1. The demographic characteristic and top 10 diseases and syndrome were collected from annual surveys, whereas patient satisfaction, prescription amount (the number of times patients received prescription orders), consultation amount (the number of times they received consultations without prescriptions), and the average prescription expense were analyzed monthly. The questionnaire items about medication effectiveness, acceptance of prescription expense, attitudes toward remote medicine, and difficulties with the telemedicine experience were also investigated. The surveys were conducted by telephone. Data including patients' demographic features, patient satisfaction, medicine effect, patients' compliance with doctors' advices, patients' acceptance of prescription expense, patients' desire to revisit the Department of Guangdong Online Hospital, the top 10 diseases and self-reported difficulties in telemedicine experience that were described in terms of frequency and percentages (n, %). The monthly average prescription expense was described using a line chart. Monthly consultation and prescription amount, as well as monthly prescriptions of Western medicine and herbs, were described using a bar chart. The age distribution of the consultation population was described using a pie chart. Statistical analysis was performed with SPSS Version 22.0 statistic software package.

XSL•FO
RenderX

Results

Demographic Characteristics of Telemedicine Consultation

From June 2015 to May 2016, approximately 67,740 participants had consultations at the Department of Guangdong Online Hospital, Guangdong Second Provincial General Hospital. Approximately 22,740 of these participants were interviewed voluntarily. The number of participants in the medical prescription group was much greater than that in the medical and face-to-face consultation groups. Women comprised majority (45,386/67,740, 67.00%) of the participants and men comprised the minority (22,354/67,740, 33.00%, Table 2). Participants between 20 and 30 years old were more inclined to use telemedicine (24,269/67,740, 35.83%) compared with the others (Table 3).

Diseases, Consultations, Prescriptions, and Medicine Expenses

Disease Spectrum

The top 10 diseases included respiratory ailments, gynecological conditions, sleep disorders, and pain syndrome. The top 3 diseases and syndromes of telemedicine consultation in rural areas were upper respiratory ailments (12,371/36,311, 34.07%), laryngopharyngitis (4938/36,311, 13.60%), and menstrual disorder (4669/36,311, 12.86%), as seen in Table 4.

Table 2. Demographic characteristics of telemedicine users (N=67,740).

Demographic characteristics	Medical prescription group (n=35,021)	Medical advice group (n=21,887)	Face-to-face consultation group (n=10,832)	Total
Male, n (%)	10,059 (28.72)	8271 (37.79)	4024 (37.15)	22,354 (33.00)
Female, n (%)	24,962 (71.28)	13,616 (62.21)	6808 (62.85)	45,386 (67.00)
Age, mean (SD)	34.05 (15.45)	35.75 (20.40)	36.81 (14.72)	34.76 (17.06)

Table 3. Age distribution of consultation population (N=67,740).

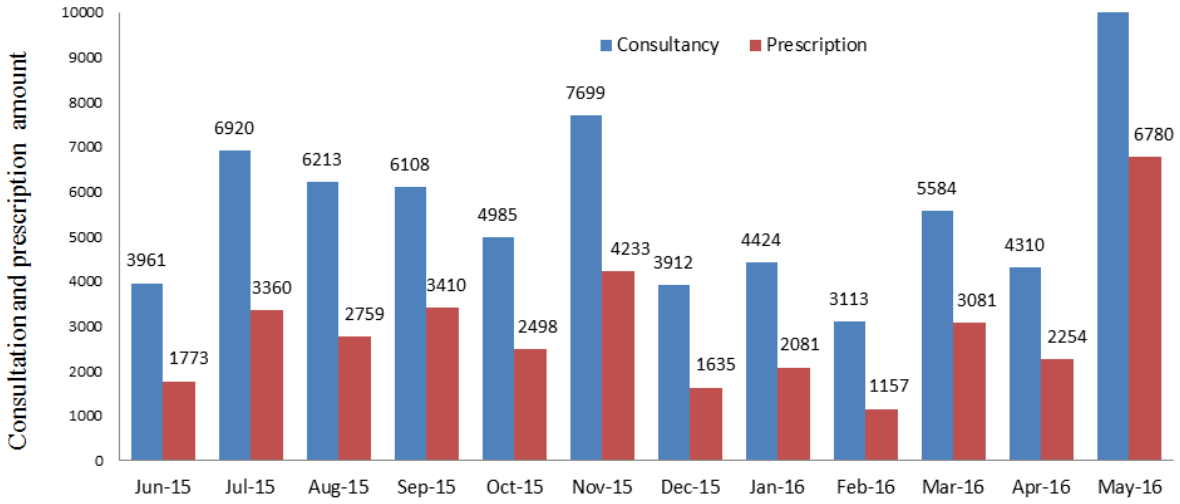
Age (years)	n (%)
<10	2618 (3.86)
11-20	3657 (5.40)
21-30	24,269 (35.83)
31-40	14,060 (20.76)
41-50	10,808 (15.95)
51-60	7816 (11.54)
61-70	3069 (4.53)
>70	1443 (2.13)

Table 4. Top 10 diseases and syndromes in telemedicine consultations in rural regions (N=36,311).

Ranking	Disease or syndromes	n (%)
1	Upper respiratory infection	12,371 (34.07)
2	Laryngopharyngitis	4938 (13.60)
3	menstrual disorder	4669 (12.86)
4	Cough	3728 (10.27)
5	sleep disorder	3132 (8.63)
6	Gastritis	1969 (5.42)
7	Dyspepsia	1812 (4.99)
8	Colpitis	1400 (3.86)
9	Dysmenorrheal	1258 (3.45)
10	Osphyalgia	1034 (2.85)

XSL•FO
RenderX

Figure 3. Monthly consultations and prescriptions.



Consultations and Prescriptions

The prescription amount (35,021) was lower than the total consultation amount (67,740) from June 2015 to May 2016, and the prescription amount was lower than the consultation amount every month (Figure 3). The total number of prescriptions for Western medicines (31,360) was much more than the number of prescriptions for Chinese herbs (3661) from June 2015 to May 2016, and the number of monthly prescriptions for Western medicines was much greater than that for Chinese herbs (Figure 4).

Monthly Average Medicine Expense Per Prescription

The monthly average medicine expense per prescription was lowest in June 2015 (54.84 ¥) and highest in May 2016 (77.29 ¥) but did not change substantially within a year (Figure 5). The annual average medicine expense per prescription was 62.9 ¥.

Results of Telephone Surveys

Patient Satisfaction

The surveys on patient satisfaction demonstrated that most patients were very satisfied (55,687/67,740, 82.20%), whereas few were satisfied (5891/67,740, 8.70%), basically satisfied (3561/67,740, 5.26%), dissatisfied (1454/67,740, 2.15%), or had no comment (1147/67,740, 1.69%). The percentages of those who were very satisfied, satisfied, and basic satisfied in the face-to-face consultation group (10,527/10,832, 97.10%) was higher than those in the medical prescription group (33,622/35,021, 96.00%) and the medical advice group (20,990/21,887, 95.90%), as seen in Table 5.

Medical Expense Acceptance

Most participants perceived the telemedicine expense to be acceptable (8775/12,450, 70.50%) or mostly acceptable (2497/12,450, 20.01%) with fewer perceiving it to be acceptable but somewhat expensive (980/12,450, 7.9%) or unacceptably expensive (198/12,450, 1.6%). The percentages of those who found their expenses completely acceptable, mostly acceptable, or acceptable but somewhat expensive in the Western medicine prescription group (11,103/11,271, 98.5%) were higher than in the herbal prescription group (1149/1179, 97.46%), as seen in Table 6.

Self-Reported Therapeutic Effects

A majority of participants reported positive treatment outcomes (full recovery: 5008/12,450, 40.22% and recovering: 4873/12,450, 39.14%); fewer reported no effect (2153/12,450, 17.29%) or worsening conditions (416/12,450, 3.3%). The percentages of full recovery and recovery in the Western medicine prescription group (8952/11,271, 79.43%) were the same as those in the herbal prescription group (929/1179, 78.80%), as seen in Table 7.

Patient Compliance

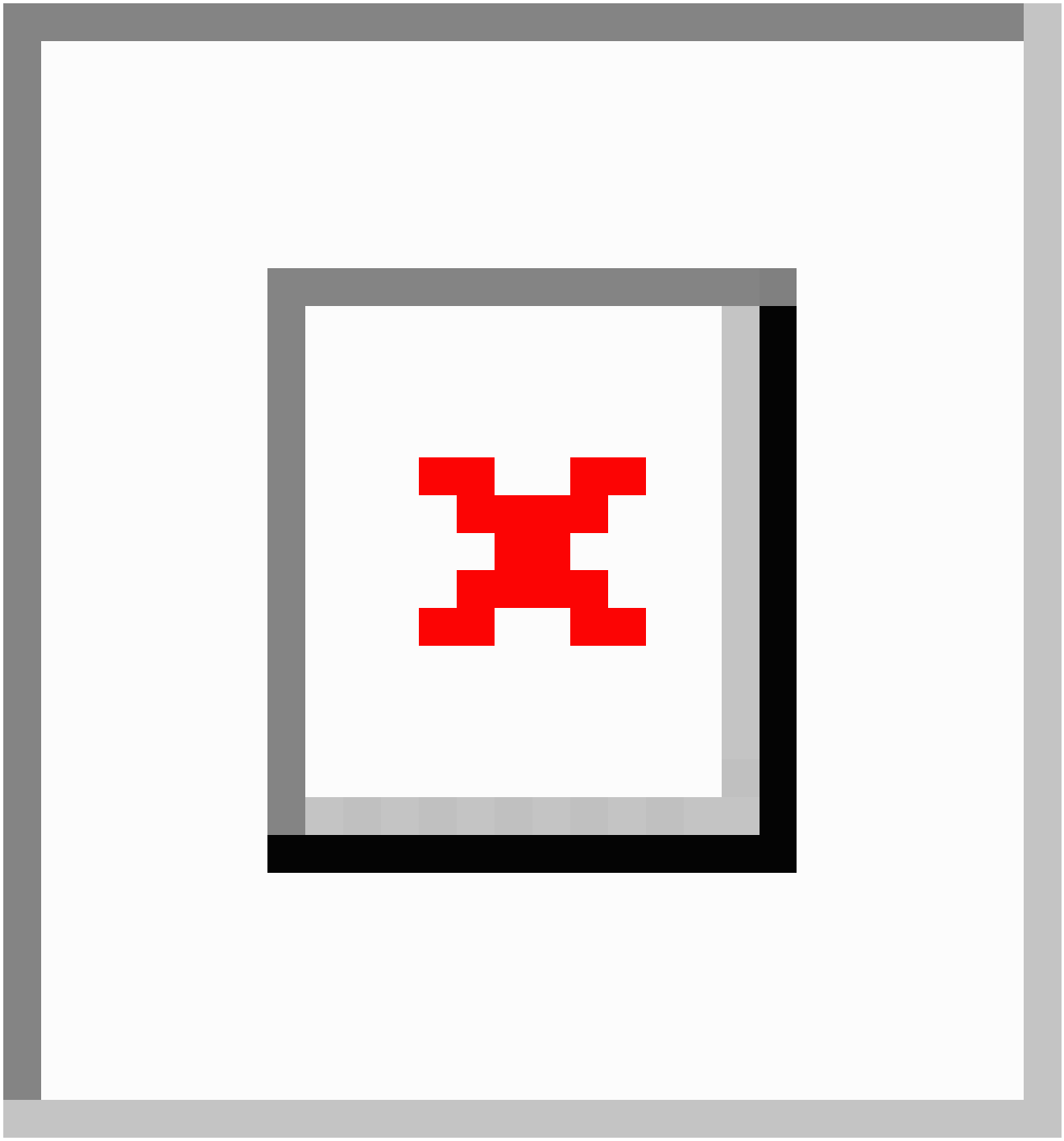
Patients' compliance with doctors' advice included complete compliance (5430/10,290, 52.77%), partial compliance (3684/10,290, 35.80%) with doctor advice and no compliance at all (1176/10,290, 11.43%). Approximately half reported complete compliance in both the medical advice group and the face-to-face consultation group. Nearly half complied with the doctors' advice partially and just listened to the advice without action in the medical advice group (3235/6907, 46.84%) and the face-to-face consultation group (1625/3383, 48.03%), as seen in Table 8.

Patients' Desire to Revisit the Online Hospital

Approximately 89.00% (20,240/22,740) of the participants were willing to revisit the Online Hospital, whereas 10.99% (2500/22,740) would not. Most of the participants in the medical prescription group (11,078/12,450, 88.98%) would revisit it, as would those in the medical advice group (6182/6907, 89.50%) and face-to-face consultation group (2982/3383, 88.10%), however, the rest would not revisit it, as seen in Table 9.

XSL•FO
RenderX

Figure 4. Monthly prescriptions for Western medicine and herbs.



XSL•FO
RenderX

Figure 5. Monthly average prescription expenditures.

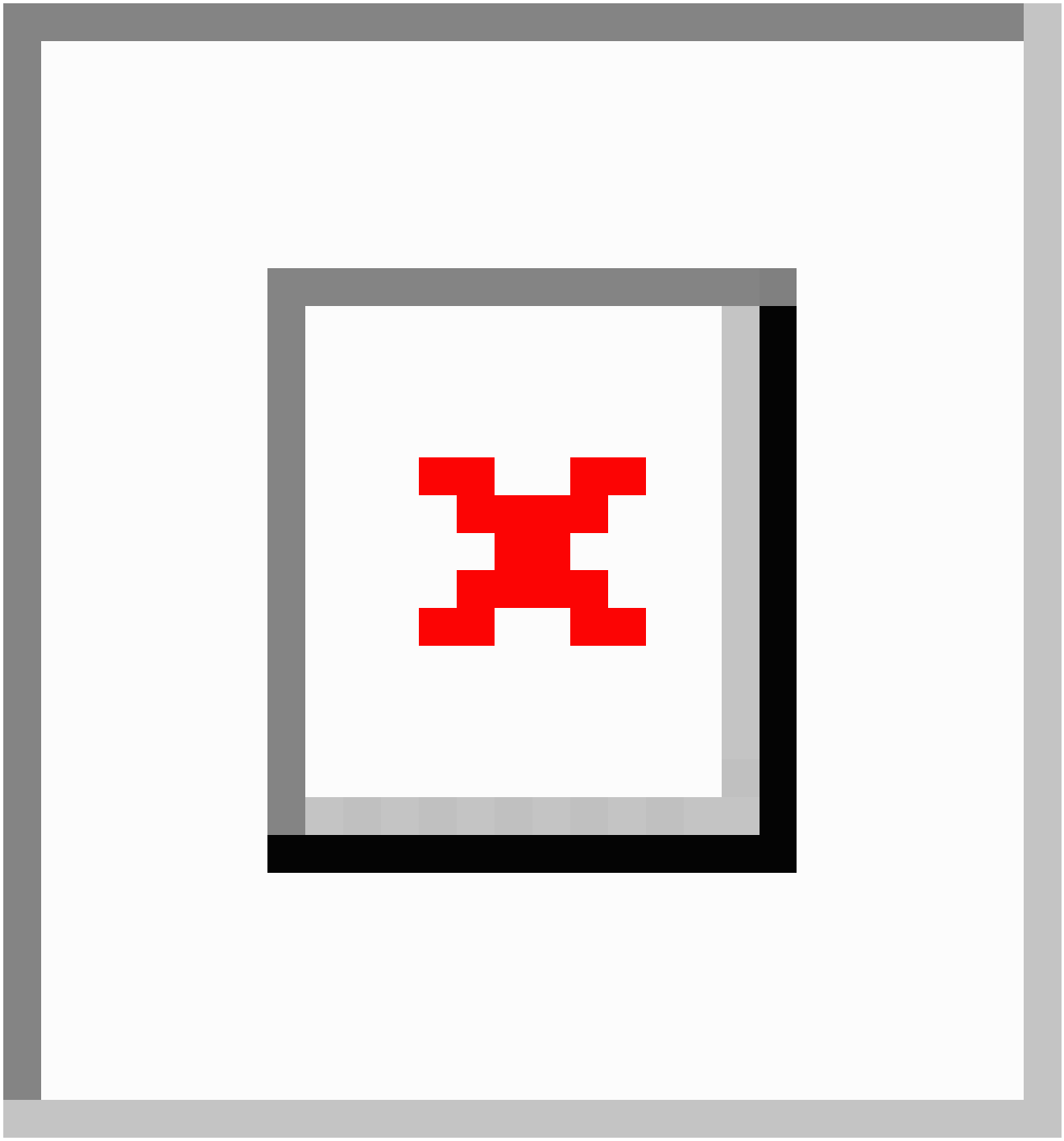


Table 5. Patient satisfaction rates in the 3 groups (N=67,740).

Patient satisfaction	Medical prescription group (n=35,021), n (%)	Medical advice group (n=21,887), n (%)	Face-to-face consultation group (n=10,832), n (%)	Total, n (%)
Very satisfied	28,718 (82.00)	17,838 (81.50)	9131 (84.30)	55,687 (82.21)
Satisfied	3152 (9.00)	1860 (8.50)	879 (8.11)	5891 (8.70)
Basic satisfied	1752 (5.00)	1292 (5.90)	517 (4.77)	3561 (5.26)
Unsatisfied	701 (2.0)	547 (2.5)	206 (1.90)	1454 (2.14)
No comment	698 (2.0)	350 (1.6)	99 (0.91)	1147 (1.69)

XSL•FO
RenderX

Table 6. Annual expense acceptance (N=12,450).

Expense acceptance	Western medicine prescription group (n=11,271), n (%)	Herb prescription group (n=1179), n (%)	Total, n (%)
Completely acceptable	7956 (70.59)	819 (69.47)	8775 (70.48)
Mostly acceptable	2262 (20.07)	235 (19.93)	2497 (20.02)
Acceptable but somewhat expensive	885 (7.85)	95 (8.06)	980 (7.9)
Unacceptable due to the high cost	168 (1.49)	30 (2.54)	198 (1.6)

Table 7. Self-reported therapeutic effects (N=12,450).

Treatment outcome	Western medicine prescription group (n=11,271), n (%)	Herb prescription group (n=1179), n (%)	Total, n (%)
Full recovery	4528 (40.17)	480 (40.71)	5008 (40.22)
Recovery	4424 (39.25)	449 (38.08)	4873 (39.19)
No effect	1933 (17.15)	220 (18.66)	2153 (17.29)
Worsening	386 (3.42)	30 (2.54)	416 (3.3)

Table 8. Patient compliance with doctors' advice (N=10,290).

Patient compliance	Medical advice group (n=6907), n (%)	Face-to-face consultation group (n=3383), n (%)	Total, n (%)
Completely	3672 (53.16)	1758 (51.97)	5430 (52.77)
Partially	2442 (35.36)	1242 (36.71)	3684 (35.80)
Just listen to the advice	793 (11.48)	383 (11.32)	1176 (11.43)

Table 9. Patients' desire to revisit the Department of Guangdong Online Hospital (N=22,740).

Whether to revisit	Medical prescription group (n=12,450), n (%)	Medical advice group (n=6905), n (%)	Face-to-face consultation group (n=3385), n (%)	Total, n (%)
Yes, I will revisit the online hospital	11,078 (88.98)	6180 (89.50)	2982 (88.)	20,240 (89.01)
No, I will not revisit the online hospital	1372 (11.02)	725 (10.50)	403 (11.90)	2500 (10.99)

Table 10. Self-reported difficulties with the telemedicine experience (N=22,740).

Difficulties in telemedicine experience	n (%)
Language barrier	3708 (16.30)
Tiring commute	2068 (9.08)
Transportation inconvenience	4670 (20.50)
Medication shortages	1282 (5.64)
Unstable internet connection	3820 (16.80)
Poor computer skills	4980 (21.90)
Family care burden	679 (3.0)
Medical hardware shortage	1177 (5.18)
Others unclassified difficulties	356 (1.6)

Self-Reported Difficulties With Telemedicine Experience

The top 3 self-reported difficulties by the participants included poor computer skills (4980/22,740, 21.90%), transportation inconvenience (4670/22,740, 20.50%), and an unstable internet connection (3820/22,740, 16.80%), whereas they also complained about medical hardware and medicine shortages, language barriers, and tiring commute (Table 10).

Discussion

Principal Findings

Data obtained in previous studies indicated that telemedicine could significantly improve patient satisfaction, save costs and time compared with conventional means, and increase access to health care resources for rural patients [9,14]. The results of our study were almost consistent with the findings of previous

XSL•FO
RenderX

research but presented more problem-oriented and detailed results. In this study, we found that telemedicine has a wide disease spectrum, similar to ordinary medicine in China. It saves costs, has high patient satisfaction and price acceptability, and can relieve disease and syndromes. However, challenges such as poor computer skills, transportation inconvenience, language barriers, unstable internet connection, and medication and hardware shortage at sanitary stations need to be resolved.

Demographic Characteristics of Telemedicine Consultation

We noted some demographic characteristics concerning the inclination to use telemedicine. Participants between 20 and 40 years old accounted for more than half of the consultation, and women were more inclined than men to use the internet to obtain medical services, which was not reported previously. These findings indicated that people in this age spectrum have more open attitudes toward new ways of receiving health care. They have better computer skills and language abilities than seniors and the young, which might be the reasons for the inclination. The rural migrant workers account for most of the migrant population in China. The percentages of rural female and male migrant workers were 47% and 53%, respectively [15]. This means that more female farmers stay at home compared with men. Women were the main users of telemedicine because their spouses were working in the cities which is a characteristic pattern in China.

Telemedicine Practice in Rural Areas

Disease Spectrum

Our research found that the top 3 diseases and syndromes treated by telemedicine were mainly upper respiratory diseases, Laryngopharyngitis and menstrual disorder, a pattern comparable with that reported in the health medical yearbook of China in 2009 [16]. This finding could guide disease prevention strategies in the rural regions of Guangdong Province.

Consultation and Prescription

Monthly consultation and prescription showed an abnormal distribution. The consultation services included prescriptions and advices. Some of the participants obtained prescriptions, whereas the others received professional advice from doctors. The number of prescriptions was lower than the number of visits every month. Both Western medicine and herbs were prescribed with the former prescribed much more than the latter, expanding the range of telemedicine treatment reported previously [17-19]. Traditional Chinese medicine (TCM) has had deep and wide grassroots following in rural China [20] and still plays an important role there. Statistics indicate that the number of consultations at rural sanitary stations for TCM service was 51,707,000 in 2012 [21]. Chinese herbs were undoubtedly popular in telemedicine practice in rural areas. The TCM services were delivered from a tertiary hospital to sanitary stations via telemedicine. This study, therefore, provided a good example of a combined treatment using both Western medicine and herbs for dealing with health problems by telemedicine in rural regions.

Prescription Expenses

We calculated that the annual average prescription price was 62.9 ¥, which was much lower than the average outpatient prescription expenditure of 291.3 ¥ in tertiary public hospitals and 189.5 ¥ in secondary public hospitals according to the latest data from January 2016 to October 2016 from the website of the National Health and Family Planning Commission of the People's Republic of China [22]. Therefore, the results indicated that telemedicine could save costs compared with conventional medicine and serve as a feasible solution to solve the problem of health care costs.

Effectiveness and Challenges

Effectiveness

Telemedicine had a high patient satisfaction rate. A survey of patient satisfaction was conducted immediately after the consultations. Approximately vast majority (55,687/67,740, 82.21%) of the participants were very satisfied with their telemedicine experience, whereas minority (1454/67,740, 2.14%) was not satisfied. The results of this research were comparable with those of a previous study that found a good patient satisfaction rate with telemedicine in rural regions [10]. A survey on patient satisfaction with conventional medicine among peasants by the statistics center of the National Health and Family Planning Commission demonstrated that a high proportion of peasants were satisfied (80%), whereas a small fraction was unsatisfied (1.5%), which was lower than the patient satisfaction rate in telemedicine practice [23]. This study indicated that the telemedicine prescription expense was highly acceptable by rural patients with a rate of acceptance (11,272/12,450, 90.50%) that was higher than that for conventional medicine in rural areas (40%) [23]. The consultation was free, and patients only had to pay for medication, which was different from face-to-face consultations in hospitals in China. Our survey results implied that a large proportion of participants believed that telemedicine could save costs. Because the annual average prescription expense was much lower than the outpatient prescription expense assessed by the National Bureau of Statistics of China [22], we determined that telemedicine could save costs, which was consistent with Thaker's findings [24], but diverged from those of Upatising [25]. Our study indicated that telemedicine could help alleviate or even cure diseases or syndromes, thus improving treatment outcomes, which were consistent with the findings of Tso [26] and Muller [27]. We also found that approximately 89% of the participants would like to revisit the Department of Guangdong Online Hospital. The number of participants who reported complete adherence to physician's advice was almost equal to that of those reporting partial compliance and no compliance, which was consistent with Bateman's research [28]. The findings of our study indicated that telemedicine was popular but did not have a good compliance. Therefore, more research and efforts are needed to improve patient compliance.

Challenges

We also assessed the difficulties associated with the telemedicine experience. Opinions from 22,740 participants on

XSL•FO
RenderX

the problems they encountered and included poor patient computer skills, transportation problems, an unstable internet connection, and language communication barriers. The percentage of the peasants who were illiterate and had less than a middle school education was 91.7% [15]; therefore, acquisition of computer skills and Mandarin or Cantonese would undoubtedly be hardship for them, especially for the seniors. The older participants tended to consider poor computer skills and poor oral expression and understanding of Mandarin, which were quite different from their local dialects, as the biggest barriers in their telemedicine experiences; as a result, in most cases, they needed assistance from the sanitary station. The participants also complained of the inconvenience of traveling several miles without public transportation to the downtown hospitals. Some complained of an unstable internet connection with intermittent visual and auditory signals hindering smooth communication with doctors. Some complained of exhaustive journeys from their home to sanitary station without public transportation. The participants, especially women, mentioned that no one could take care of their children, parents, farmland, and livestock if they traveled downtown for telemedicine services. Some complained that the sanitary stations lacked the medicines prescribed by the doctors, requiring them to seek substitute medications or return home empty handed. Some also complained that the pharmacies lacked the usual medical devices, such as a hematomanometer and glucose meter. This research was the first study to our knowledge to investigate the difficulties with telemedicine experience for people in rural regions in China; it could thus provide useful guidance for future telemedicine practice in rural areas.

Background on Telemedicine Practice in Rural China

The maldistribution of medicine and hygiene resources between rural and urban regions has been evident despite several stages of medical reforms initiated in 1985. The rural population accounted for a large portion of the population but had fewer medical resources of both medical institutions and staffs than people in cities. The statistics reveal that the total number of people diagnosed and treated in medical and hygiene institutions in China was 64.2 billion, an increase of 2%, and the number of people in hospitals was 26.3 billion, an increase of 5%, but in grass-root medical institutes, the number was 35.6 billion, a decrease of 1% compared with the same period in 2016 [29]. This implies that a large number of rural people would seek health care from urban hospitals but not in their places of residential areas. As a result, the Chinese government launched official documents requiring and advocating telemedicine as an important approach for solving medical resource inequality between urban and rural regions and facilitate the ability of rural people to obtain medical services in their areas through local medicine and hygiene administration departments to resolve difficulties in access to better and sufficient medical services for rural residents [6]. The Department of Guangdong Online Hospital, Guangdong Second Provincial General Hospital, the first remote medical institution authorized by the National Health and Family Planning Commission, has endeavored to deliver health care to rural regions through telemedicine practice on which our research was based, indicating that adding telemedicine to rural sanitary stations was a feasible way to solve difficulties in access to high-quality medical services in tertiary hospitals for rural residents.

Limitations

The results of this study could guide further promotion and advocacy of telemedicine for delivering better medical services from larger cities to remote and rural regions in China and other developing countries. However, some limitations are worth noting. Some of the participants could not be contacted for follow-up telephone interviews because of wrong telephone numbers or their refusal to answer questions.

Conclusions and Future Directions

Telemedicine has a wide disease spectrum similar to that addressed by ordinary medicine in China. It could save costs, provide high patient satisfaction, popularity and price acceptance, and help cure or relieve diseases and syndromes. However, problems remain that must be resolved, including poor computer skills, transportation inconvenience, language communication barriers, unstable internet connection, and medicine and hardware shortage at sanitary stations. Telemedicine could serve as a feasible approach to addressing the maldistribution of health care resources in rural China. The results could guide further promotion and advocacy of telemedicine practice in rural regions of China and other developing countries. Future research should focus on providing a smooth and stable telemedicine network, resolving transportation inconvenience, improving patient compliance and helping the elderly obtain telemedicine services in rural regions.

Acknowledgments

We thank the Department of Guangdong Online Hospital, Guangdong Second Provincial General Hospital for their support of this research.

Conflicts of Interest

None declared.

References

1. National Bureau of Statistics of the People's Republic of China. The population census of China in 2010. 2011. URL: http://www.stats.gov.cn/tjsj/pcsj/ [accessed 2018-09-24] [WebCite Cache ID 72gYPSY10]

XSL•FO
RenderX

2. Zhou K, Zhang X, Ding Y, Wang D, Lu Z, Yu M. Inequality trends of health workforce in different stages of medical system reform (1985-2011) in China. Hum Resour Health 2015 Dec 08;13:94 [FREE Full text] [doi: 10.1186/s12960-015-0089-0] [Medline: 26645960]
3. Liu W, Liu Y, Twum P, Li S. National equity of health resource allocation in China: data from 2009 to 2013. Int J Equity Health 2016 Apr 19;15:68 [FREE Full text] [doi: 10.1186/s12939-016-0357-1] [Medline: 27093960]
4. Chen Z. Launch of the health-care reform plan in China. The Lancet 2009 Apr;373(9672):1322-1324. [doi: 10.1016/S0140-6736(09)60753-4]
5. National Health and Family Planning Commission of the People's Republic of China. Statistical bulletin on the development of health and family planning in China in 2015. The Central Government of People's Republic of China; 2016. URL: http://www.nhfpc.gov.cn/guihuaxxs/s10748/201607/da7575d64fa04670b5f375c87b6229b0.shtml [accessed 2018-09-24] [WebCite Cache ID 72geI97V8]
6. National Health and Family Planning Commission of the People's Republic of China. The opinion about promotion of remote medical service by medical institutions of national Health and Family Planning Commission of the People?s Republic of China. URL: http://www.nhc.gov.cn/yzygj/s3593g/201408/f7cbfe331e78410fb43d9b4c61c4e4bd.shtml [accessed 2018-09-24] [WebCite Cache ID 72gedmNNv]
7. World Health Organization. Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth 2009. Geneva Switzerland: World Health Organization; 2011. URL: https://www.researchgate.net/publication/277020369 [accessed 2018-09-24] [WebCite Cache ID 72gfO4r2M]
8. Kim H, Ahn M, Choi Y, Choi E, Kim D, Shin S, et al. Fifteen-year Experience with Telemedicine Services in Gangwon Province in Korea. Healthc Inform Res 2015 Oct;21(4):283-291 [FREE Full text] [doi: 10.4258/hir.2015.21.4.283] [Medline: 26618035]
9. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. Pediatr Res 2015 Oct 14;79(1-2):169-176 [FREE Full text] [doi: 10.1038/pr.2015.192]
10. López C, Valenzuela JI, Calderón JE, Velasco AF, Fajardo R. A telephone survey of patient satisfaction with realtime telemedicine in a rural community in Colombia. J Telemed Telecare 2011 Dec;17(2):83-87. [doi: 10.1258/jtt.2010.100611] [Medline: 21139016]
11. Ferrer-Roca O, Garcia-Nogales A, Pelaez C. The impact of telemedicine on quality of life in rural areas: the Extremadura model of specialized care delivery. Telemed J E Health 2010 Mar;16(2):233-243. [doi: 10.1089/tmj.2009.0107] [Medline: 20302385]
12. Jaglal SB, Haroun VA, Salbach NM, Hawker G, Voth J, Lou W, et al. Increasing access to chronic disease self-management programs in rural and remote communities using telehealth. Telemed J E Health 2013 Jun;19(6):467-473 [FREE Full text] [doi: 10.1089/tmj.2012.0197] [Medline: 23570277]
13. Jiong T, Chun X, Shao L. The internet hospital: an emerging innovation in China. Lancet Glob Health 2015;3(8):e445-e446. [doi: 10.1016/S2214-109X]
14. Sorwar G, Rahamn MM, Uddin R, Hoque MR. Cost and Time Effectiveness Analysis of a Telemedicine Service in Bangladesh. Stud Health Technol Inform 2016;231:127-134. [Medline: 27782024]
15. National Bureau of Statistics of the People's Republic of China. Bulletin of main data of the third national agricultural census (no. 5). The personnel of agricultural production and operation URL: http://www.stats.gov.cn/tjsj/tjgb/nypcgb/qgnypcgb/201712/t20171215_1563599.html [accessed 2018-09-25] [WebCite Cache ID 72ggPgwiE]
16. The Statistics Center of National HealthFamily Planning Commission of the People's Republic of China. Health Medical Yearbook of China in 2009. 2009. The inquiry of disease spectrum and morbidity within two weeks for regional citizens URL: http://www.nhc.gov.cn/htmlfiles/zwgkzt/ptjnj/year2009/t-8.htm [accessed 2018-09-25] [WebCite Cache ID 72ggYF8cm]
17. Mehrotra A, Jena AB, Busch AB, Souza J, Uscher-Pines L, Landon BE. Utilization of Telemedicine Among Rural Medicare Beneficiaries. JAMA 2016 May 10;315(18):2015-2016 [FREE Full text] [doi: 10.1001/jama.2016.2186] [Medline: 27163991]
18. Das T, Pappuru RR. Telemedicine in diabetic retinopathy: Access to rural India. Indian J Ophthalmol 2016 Jan;64(1):84-86 [FREE Full text] [doi: 10.4103/0301-4738.178151] [Medline: 26953029]
19. Kruse CS, Bouffard S, Dougherty M, Parro JS. Telemedicine Use in Rural Native American Communities in the Era of the ACA: a Systematic Literature Review. J Med Syst 2016 Jun;40(6):145 [FREE Full text] [doi: 10.1007/s10916-016-0503-8] [Medline: 27118011]
20. Zhang D, Unschuld PU. China's barefoot doctor: past, present, and future. The Lancet 2008 Nov;372(9653):1865-1867. [doi: 10.1016/S0140-6736(08)61355-0]
21. The Statistics Center of National Health and Family Planning Commission of the People's Republic of China. 2013 China Health Statistics Yearbook. URL: http://www.nhfpc.gov.cn/htmlfiles/zwgkzt/ptjnj/year2013/index2013.html [accessed 2018-09-25] [WebCite Cache ID 72h235ZEU]
22. National Health and Family Planning Commission of the People's Republic of China. The patients' hospitalization costs for secondary and tertiary public hospitals from January to October in 2016. URL: http://www.nhc.gov.cn/mohwsbwstjxxzx/s7967/201612/25fa8a3386d745219cfcab27d7ffc5ad.shtml [accessed 2018-09-25] [WebCite Cache ID 72ggeC13M]

XSL•FO
RenderX

23. The Statistics Center of National Health and Family Planning Commission of the People's Republic of China. Results of the Fifth National Health Service Survey - Residents' Satisfaction. URL: http://med.wanfangdata.com.cn/ [accessed 2018-09-25] [WebCite Cache ID 72ggkAcdT]
24. Thaker DA, Monypenny R, Olver I, Sabesan S. Cost savings from a telemedicine model of care in northern Queensland, Australia. Med J Aust 2013 Sep 16;199(6):414-417. [Medline: 24033216]
25. Upatising B, Wood DL, Kremers WK, Christ SL, Yih Y, Hanson GJ, et al. Cost comparison between home telemonitoring and usual care of older adults: a randomized trial (Tele-ERA). Telemed J E Health 2015 Jan;21(1):3-8 [FREE Full text] [doi: 10.1089/tmj.2014.0021] [Medline: 25453392]
26. Tso JV, Farinpour R, Chui HC, Liu CY. A Multidisciplinary Model of Dementia Care in an Underserved Retirement Community, Made Possible by Telemedicine. Front Neurol 2016 Dec;7:225 [FREE Full text] [doi: 10.3389/fneur.2016.00225] [Medline: 28066313]
27. Müller KI, Alstadhaug KB, Bekkelund SI. Telemedicine in the management of non-acute headaches: A prospective, open-labelled non-inferiority, randomised clinical trial. Cephalalgia 2017 Aug;37(9):855-863. [doi: 10.1177/0333102416654885] [Medline: 27301460]
28. Bateman EH, Keef DM. How can eHealth enhance adherence to cancer therapy and supportive care? Srp Arh Celok Lek 2016(144):116-121. [Medline: 27276871]
29. The Statistics Center of National Health and Family Planning Commission of the People's Republic of China. The statistical data of the national medicine and hygiene institution in the end of October, 2016. URL: http://www.nhc.gov.cn/mohwsbwstjxxzx/s7967/201612/6300bd3cfa6148adb2b15a76dfdf4b48.shtml [accessed 2018-09-25] [WebCite Cache ID 72ghT7poW]

Abbreviations

TCM: traditional Chinese medicine

Edited by G Eysenbach; submitted 09.12.17; peer-reviewed by B Xie, L Sheets; comments to author 16.03.18; revised version received 30.04.18; accepted 20.07.18; published 21.12.18.

Please cite as:
He C, Zhou Q, Chen W, Tian J, Zhou L, Peng H, Luan S, Wang S
Using an Internet-Based Hospital to Address Maldistribution of Health Care Resources in Rural Areas of Guangdong Province, China: Retrospective and Descriptive Study
JMIR Med Inform 2018;6(4):e51
URL: http://medinform.jmir.org/2018/4/e51/
doi:10.2196/medinform.9495
PMID:30578195

©Cui He, Qiru Zhou, Wei Chen, Junzhang Tian, Lihua Zhou, Hong Peng, Shuo Luan, Shengfei Wang. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 21.12.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Computerized Method for Measuring Computed Tomography Pulmonary Angiography Yield in the Emergency Department: Validation Study

Safiya Richardson[1], MD, MPH; Philip Solomon[1], MD; Alexander O'Connell[1], MD; Sundas Khan[1], MD; Jonathan Gong[1], MD; Alex Makhnevich[1], MD; Guang Qiu[1], MD, PhD; Meng Zhang[1], PhD; Thomas McGinn[1], MD, MPH

Department of Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, United States

Corresponding Author:
Safiya Richardson, MD, MPH
Department of Medicine
Donald and Barbara Zucker School of Medicine at Hofstra/Northwell
500 Hofstra University
Hempstead, NY, 11549
United States
Phone: 1 4046957883
Email: srichard12@northwell.edu

Abstract

Background: Use of computed tomography pulmonary angiography (CTPA) in the assessment of pulmonary embolism (PE) has markedly increased over the past two decades. While this technology has improved the accuracy of radiological testing for PE, CTPA also carries the risk of substantial iatrogenic harm. Each CTPA carries a 14% risk of contrast-induced nephropathy and a lifetime malignancy risk that can be as high as 2.76%. The appropriate use of CTPA can be estimated by monitoring the CTPA yield, the percentage of tests positive for PE. This is the first study to propose and validate a computerized method for measuring the CTPA yield in the emergency department (ED).

Objective: The objective of our study was to assess the validity of a novel computerized method of calculating the CTPA yield in the ED.

Methods: The electronic health record databases at two tertiary care academic hospitals were queried for CTPA orders completed in the ED over 1-month periods. These visits were linked with an inpatient admission with a discharge diagnosis of PE based on the International Classification of Diseases codes. The computerized the CTPA yield was calculated as the number of CTPA orders with an associated inpatient discharge diagnosis of PE divided by the total number of orders for completed CTPA. This computerized method was then validated by 2 independent reviewers performing a manual chart review, which included reading the free-text radiology reports for each CTPA.

Results: A total of 349 CTPA orders were completed during the 1-month periods at the two institutions. Of them, acute PE was diagnosed on CTPA in 28 studies, with a CTPA yield of 7.7%. The computerized method correctly identified 27 of 28 scans positive for PE. The one discordant scan was tied to a patient who was discharged directly from the ED and, as a result, never received an inpatient discharge diagnosis.

Conclusions: This is the first successful validation study of a computerized method for calculating the CTPA yield in the ED. This method for data extraction allows for an accurate determination of the CTPA yield and is more efficient than manual chart review. With this ability, health care systems can monitor the appropriate use of CTPA and the effect of interventions to reduce overuse and decrease preventable iatrogenic harm.

(JMIR Med Inform 2018;6(4):e44) doi:10.2196/medinform.9957

KEYWORDS

health informatics; pulmonary embolism; electronic health record; quality improvement; clinical decision support systems

XSL•FO
RenderX

Introduction

The ability of computed tomography (CT) to diagnose pulmonary embolism (PE) was demonstrated in 1980 [1]. The introduction of multidetector row CT pulmonary angiography (CTPA) revolutionized the diagnostic approach to PE in 1998 [2]. The availability and use of this new technology rapidly increased in the following years, and by 2001, CT overtook the ventilation/perfusion lung (V/Q) scan as the most common method for diagnosing PE [3].

In 2006, results from the landmark prospective investigation of pulmonary embolism diagnosis (PIOPED) II trial established CTPA as the first-choice diagnostic imaging modality, with a sensitivity of >90% for patients with high clinical suspicion of PE and a specificity of 96% [4,5]. Over the next 5 years, there was a 4-fold increase in CTPA use and a 33% decrease in V/Q scanning [6]. However, CTPA is associated with a nearly 7-fold higher radiation burden than V/Q scanning [6], with attributable lifetime malignancy risk of up to 2.76% in young female patients [7]. Moreover, in a recent prospective study, it was found that up to 14% of patients who underwent CTPA developed contrast-induced nephropathy [8].

Increased rates of CTPA use and improved understanding of the associated adverse effects have prompted researchers to measure the CTPA yield [9-11]. The CTPA yield is a measure of the appropriateness of use, defined as the percentage of tests completed to evaluate for PE that are positive for PE. The majority of these studies have used manual chart abstraction to calculate the CTPA yield [9-11]. Furthermore, a form of artificial intelligence, natural language processing, has been shown to reliably calculate the CTPA yield in a few recent studies [12-14].

These methods have demonstrated reliability but are time consuming or require technology not available at most health care institutions. To date, a simple, standardized method of electronically calculating the CTPA yield has not been described. The objective of this study is to propose and validate a computerized method for calculating the CTPA yield in the emergency department (ED).

Methods

Procedure

We performed a multicenter observational study to validate a computerized method of calculating CTPA yield. The study was conducted at two tertiary care hospitals, the North Shore University Hospital and the Long Island Jewish Medical Center in New York, in April and November, 2016, respectively. The two hospitals are supported by the Sunrise Clinical Manager electronic health record (EHR), a subsidiary of Allscripts Healthcare Solutions (Chicago, Illinois, United States). This study was approved by the Northwell Health's Institutional Review Board.

The EHR databases at the two institutions are the replicated copies of the Sunrise Clinical Manager application. The database is replicated near real time with a <2-hour latency. Of note, this process is monitored by dedicated database administrators and analytics support team members to ensure fidelity. The databases were queried for CTPA orders completed in the ED over a 1-month period for each hospital. Patients' visits were extracted from the EHR if they had a "completed" CTPA order during their ED course. However, patients with "cancelled" or "discontinued" CTPA orders were not included. Furthermore, patients with CTPAs ordered on the same day as CT angiography of the abdomen and pelvis were excluded, as these were under the protocol to rule out aortic dissection and not PE (Figure 1).

CTPA orders from the ED were then linked to inpatient visits. PE diagnosis was measured on the basis of an inpatient discharge diagnosis of the International Classification of Diseases, Clinical Modification codes, versions 9 and 10 (ICD-9-CM and ICD-10-CM), provided by the Centers for Medicare and Medicaid Services and the National Center for Health Statistics. We included both primary and secondary diagnoses in the analysis. Furthermore, the full range of PE diagnosis codes was used: 415.0, 415.11, 415.12, 415.13, and 415.19 for ICD-9-CM; and I26.0, I26.01, I26.02, I26.09, I26.9, I26.90, I26.92, and I26.99 for ICD-10-CM.

The CTPA yield was calculated as the number of ED CTPA orders linked to an inpatient discharge diagnosis of PE divided by the total number of CTPAs completed in the ED that month. This calculated yield was then validated by performing a manual chart review. In the manual chart review, the free-text radiology read of each completed ED CTPA order was reviewed to classify the CTPA as positive or negative for PE. In addition, both the ED provider note and inpatient discharge note were reviewed to ensure that the CTPA was done to evaluate for PE and the diagnosis was not revised during the inpatient visit.

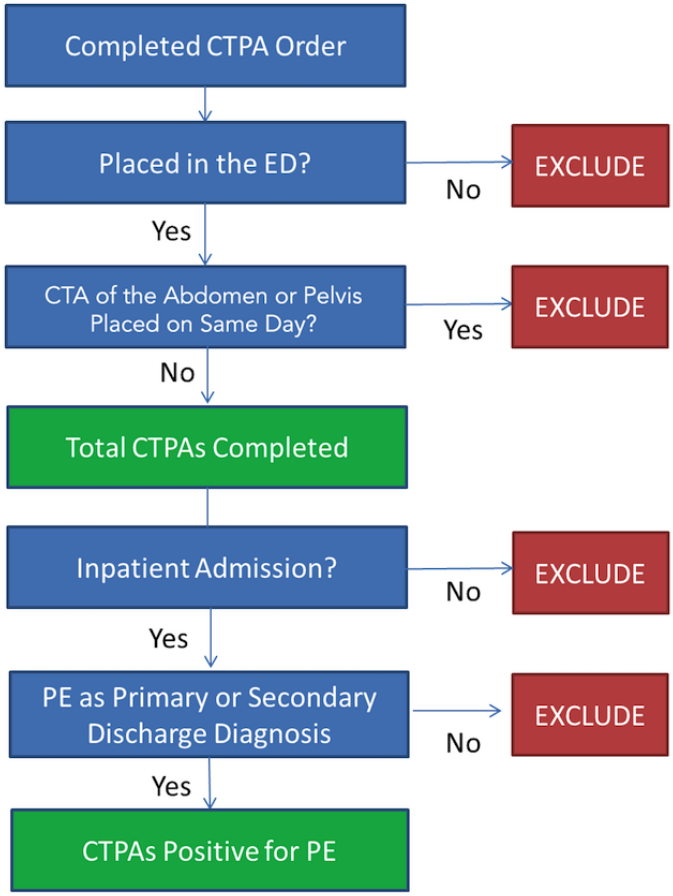
The computerized calculated yield for each month and institution was compared with the yield generated from manual chart reviews by 2 independent reviewers. The reviewers were trained internal medicine physicians with experience in reading radiology reports. Of note, the reviewers were blinded to each other but not to the computerized results and had full access to discharge documentation and the entire medical chart.

Data Analysis

We used McNemar's test to determine whether the CTPA yields were different between the computerized calculated yield and the manual chart reviews. The kappa coefficient and the corresponding 95% CI were calculated to measure the agreement between the computerized calculated yield and the manual chart reviews.

XSL•FO
RenderX

Figure 1. Algorithm of the computerized method for measuring computed tomography pulmonary angiography (CTPA) yield. ED: emergency department; CTA: computed tomography angiography; PE: pulmonary embolism.



Results

In total, 375 CTPAs were completed during the allotted period for review. Of them, 6.9% (26/375) orders were completed on the same day as a CT angiography of the abdomen and pelvis and were, thus, excluded. Furthermore, manual chart reviews confirmed that each of these 26 omitted CTPA orders and none of the remaining 349 were ordered with the intention to rule out an aortic dissection. There were no cases where a chart review of the ED provider note and inpatient discharge note changed the interpretation of the CTPA results. In addition, the independent reviewers reported the same results, with no disagreement.

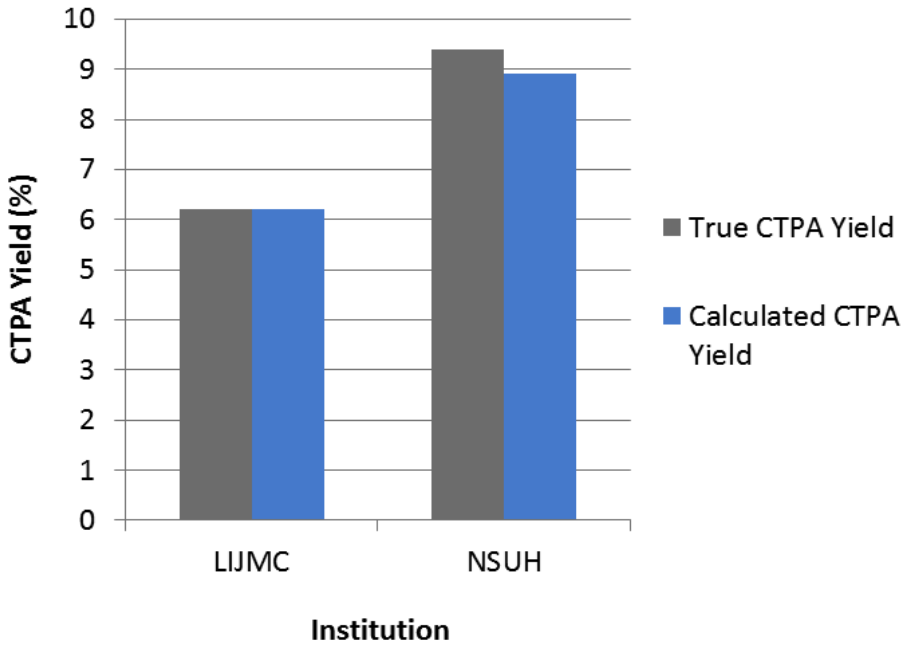
At the North Shore University Hospital site, of 203 CTPA orders completed for the evaluation of PE, 18 orders were found to have an associated inpatient discharge diagnosis of PE. The calculated yield was 8.9% (18/203). Manual chart reviews revealed 19 positive scans for a true CTPA yield of 9.4% (19/203). Notably, one discordant scan was found in a patient directly discharged from the ED, and as a result, the patient never received an inpatient discharge diagnosis.

At the Long Island Jewish Medical Center site, 146 CTPA orders were completed and 9 were found to have an associated inpatient discharge diagnosis of PE. The calculated CTPA yield was 6.2% (9/146). The manual chart reviews produced identical results, confirming 9 positive CTPA scans.

Overall, the computerized method captured 27 of 28 scans positive for PE, with an accuracy of 96.4% (27/28; Figure 2). The overall CTPA yield for both institutions was 7.7% (27/349). In this study, the P=.32, indicating that the proportions were not significantly different between the two groups. Furthermore, the kappa coefficient was .98, with 95% CI (0.94-1.00) also indicating an agreement between the two groups.

XSL•FO
RenderX

Figure 2. True computed tomography pulmonary angiography (CTPA) yield vs calculated CTPA Yield by computerized method. LIJMC: Long Island Jewish Medical Center; NSUH: North Shore University Hospital.



Discussion

Principal Findings

To the best of our knowledge, this is the first study to propose and validate a simple, standardized method of electronically calculating the CTPA yield. This method has wide applicability to address increasing concerns about both overtesting and overdiagnosis of PE. The increase in the incidence of PE accompanying the increased use of CTPA [15] has been associated with a decrease in the PE case mortality [16,17]. Physicians are testing more for PE and seem to be finding and treating clinically insignificant PEs. The ability of health care systems to computerize the monitoring of the CTPA yield allows them to address overtesting and overdiagnosis using systems interventions.

In addition, clinical decision support tools, built to estimate the pretest probability of PE and discourage the CTPA use in low-risk patients, have been shown to improve the CTPA yield. These tools reduce testing by 25%, without any missed PEs [9,18,19]. However, these studies are limited by the time required for manual chart reviews. Studies of interventions designed to reduce unnecessary CTPA use decrease exposure to both contrast and radiation and avoid costly "incidentalomas." Furthermore, incidental findings requiring clinical or radiological follow-up were found in 24% of patients without PE [16].

In this study, we describe a validated method to measure the CTPA yield that allows the data collection process to be computerized and does not require artificial intelligence. We utilized both ICD-9 and ICD-10 codes to fully encompass PE coding at the time of discharge. This method can be applied to allow for the comparison of the CTPA yield in different health care systems and different types of acute care facilities. Improved data collection will allow for more targeted interventions, with an ultimate goal of increased CTPA yields and decreased CTPA utilization.

Limitations

The one discordant scan in our study points to a limitation that will likely become more relevant in future studies. One positive CTPA was missed by our computerized method because the patient was discharged directly from the ED, meaning there was no linkable inpatient visit or potential discharge ICD code. With a push toward cost-conscious care and away from inpatient medicine, there will likely be more patients with acute PE diagnosed in the ED who are treated as outpatients. While the safety of this practice was unclear and controversial just a few years ago [20,21], it has recently become more common with the increased use of direct-acting oral anticoagulants [22,23] and safety research in the field [24,25]. This is particularly true in large health care systems with tertiary EDs that can safely assess patients' risk with bedside echocardiography and lower-extremity ultrasound [26]. Future studies will link CTPA scans to both inpatient and ED visits to improve the accuracy.

In addition, future studies may specify the type of PE and consider the discovery of a subsegmental PE as a negative study. This was not addressed in this study as currently, these are not treated differently and ICD codes do not distinguish these types of PEs. Notably, although this was not observed in our study, this method will likely count studies conducted in patients with chronic PE as positive. Finally, this study was conducted at two hospitals under one health care system, Northwell Health. Hence, future directions include studying this method at other institutions to ensure its accuracy.

Conclusions

This is the first successful validation study of a simple computerized method for calculating the CTPA yield in the ED.

XSL•FO
RenderX

This method for data extraction allows for an accurate and efficient determination of the CTPA yield and represents a significant improvement from the manual chart review. With this ability, health care systems can monitor the appropriate use of CTPA and the effect of interventions to reduce overuse and decrease preventable iatrogenic harm.

Conflicts of Interest

None declared.

References

1. Godwin JD, Webb WR, Gamsu G, Ovenfors CO. Computed tomography of pulmonary embolism. AJR Am J Roentgenol 1980 Oct;135(4):691-695. [doi: 10.2214/ajr.135.4.691] [Medline: 6778100]
2. Wiener RS, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. Arch Intern Med 2011 May 09;171(9):831-837 [FREE Full text] [doi: 10.1001/archinternmed.2011.178] [Medline: 21555660]
3. Stein PD, Kayali F, Olson RE. Trends in the use of diagnostic imaging in patients hospitalized with acute pulmonary embolism. Am J Cardiol 2004 May 15;93(10):1316-1317. [doi: 10.1016/j.amjcard.2004.02.022] [Medline: 15135716]
4. Stein PD, Woodard PK, Weg JG, Wakefield TW, Tapson VF, Sostman HD, PIOPED II Investigators. Diagnostic pathways in acute pulmonary embolism: recommendations of the PIOPED II Investigators. Radiology 2007 Jan;242(1):15-21. [doi: 10.1148/radiol.2421060971] [Medline: 17185658]
5. Remy-Jardin M, Pistolesi M, Goodman LR, Gefter WB, Gottschalk A, Mayo JR, et al. Management of suspected acute pulmonary embolism in the era of CT angiography: a statement from the Fleischner Society. Radiology 2007 Nov;245(2):315-329. [doi: 10.1148/radiol.2452070397] [Medline: 17848685]
6. Mettler FA, Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. Radiology 2008 Jul;248(1):254-263. [doi: 10.1148/radiol.2481071451] [Medline: 18566177]
7. Niemann T, Zbinden I, Roser HW, Bremerich J, Remy-Jardin M, Bongartz G. Computed tomography for pulmonary embolism: assessment of a 1-year cohort and estimated cancer risk associated with diagnostic irradiation. Acta Radiol 2013 Sep;54(7):778-784. [doi: 10.1177/0284185113485069] [Medline: 23761544]
8. Mitchell AM, Jones AE, Tumlin JA, Kline JA. Prospective study of the incidence of contrast-induced nephropathy among patients evaluated for pulmonary embolism by contrast-enhanced computed tomography. Acad Emerg Med 2012 Jun;19(6):618-625 [FREE Full text] [doi: 10.1111/j.1553-2712.2012.01374.x] [Medline: 22687176]
9. Wang RC, Bent S, Weber E, Neilson J, Smith-Bindman R, Fahimi J. The Impact of Clinical Decision Rules on Computed Tomography Use and Yield for Pulmonary Embolism: A Systematic Review and Meta-analysis. Ann Emerg Med 2016 Jun;67(6):693-701.e3. [doi: 10.1016/j.annemergmed.2015.11.005] [Medline: 26747217]
10. van Belle A, Büller HR, Huisman MV, Huisman PM, Kaasjager K, Kamphuisen PW, Christopher Study Investigators. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. JAMA 2006 Jan 11;295(2):172-179. [doi: 10.1001/jama.295.2.172] [Medline: 16403929]
11. Dunne RM, Ip IK, Abbett S, Gershanik EF, Raja AS, Hunsaker A, et al. Effect of Evidence-based Clinical Decision Support on the Use and Yield of CT Pulmonary Angiographic Imaging in Hospitalized Patients. Radiology 2015 Jul;276(1):167-174 [FREE Full text] [doi: 10.1148/radiol.15141208] [Medline: 25686367]
12. Swartz J, Koziatek C, Theobald J, Smith S, Iturrate E. Creation of a simple natural language processing tool to support an imaging utilization quality dashboard. Int J Med Inform 2017 Dec;101:93-99. [doi: 10.1016/j.ijmedinf.2017.02.011] [Medline: 28347453]
13. Tian Z, Sun S, Eguale T, Rochefort C. 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]
14. Rochefort CM, Verma AD, Eguale T, Lee TC, Buckeridge DL. A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data. J Am Med Inform Assoc 2015 Jan;22(1):155-165 [FREE Full text] [doi: 10.1136/amiajnl-2014-002768] [Medline: 25332356]
15. Hall WB, Truitt SG, Scheunemann LP, Shah SA, Rivera MP, Parker LA, et al. The prevalence of clinically relevant incidental findings on chest computed tomographic angiograms ordered to diagnose pulmonary embolism. Arch Intern Med 2009 Nov 23;169(21):1961-1965. [doi: 10.1001/archinternmed.2009.360] [Medline: 19933956]
16. Perrier A, Bounameaux H. Accuracy or outcome in suspected pulmonary embolism. N Engl J Med 2006 Jun 01;354(22):2383-2385. [doi: 10.1056/NEJMe068076] [Medline: 16738276]
17. Glassroth J. Imaging of pulmonary embolism: too much of a good thing? JAMA 2007 Dec 19;298(23):2788-2789. [doi: 10.1001/jama.298.23.2788] [Medline: 18165674]
18. Yan Z, Ip IK, Raja AS, Gupta A, Kosowsky JM, Khorasani R. Yield of CT Pulmonary Angiography in the Emergency Department When Providers Override Evidence-based Clinical Decision Support. Radiology 2017 Dec;282(3):717-725 [FREE Full text] [doi: 10.1148/radiol.2016151985] [Medline: 27689922]

XSL•FO
RenderX

19. Costantino MM, Randall G, Gosselin M, Brandt M, Spinning K, Vegas CD. CT angiography in the evaluation of acute pulmonary embolus. AJR Am J Roentgenol 2008 Aug;191(2):471-474. [doi: 10.2214/AJR.07.2552] [Medline: 18647919]
20. Piran S, Le GG, Wells P, Gandara E, Righini M, Rodger M, et al. Outpatient treatment of symptomatic pulmonary embolism: a systematic review and meta-analysis. Thromb Res 2013 Nov;132(5):515-519. [doi: 10.1016/j.thromres.2013.08.012] [Medline: 24035045]
21. Yoo H, Queluz T, El DR. Outpatient versus inpatient treatment for acute pulmonary embolism. Cochrane Database Syst Rev 2014 Nov 20(11):CD010019. [doi: 10.1002/14651858.CD010019.pub2] [Medline: 25411774]
22. Ghazvinian R, Gottsäter A, Elf JL. Efficacy and safety of outpatient treatment with direct oral anticoagulation in pulmonary embolism. J Thromb Thrombolysis 2018 Feb;45(2):319-324 [FREE Full text] [doi: 10.1007/s11239-017-1607-9] [Medline: 29305675]
23. Medina A, Raskob G, Ageno W, Cohen AT, Brekelmans MPA, Chen CZ, et al. Outpatient Management in Patients with Venous Thromboembolism with Edoxaban: A Post Hoc Analysis of the Hokusai-VTE Study. Thromb Haemost 2017 Dec;117(12):2406-2414 [FREE Full text] [doi: 10.1160/TH17-05-0523] [Medline: 29212128]
24. Abusibah H, Abdelaziz MM, Standen P, Bhatia P, Hamad MM. Ambulatory management of pulmonary embolism. Br J Hosp Med (Lond) 2018 Jan 02;79(1):18-25. [doi: 10.12968/hmed.2018.79.1.18] [Medline: 29315054]
25. Jiménez D, Yusen R. Outpatient therapy for acute symptomatic pulmonary embolism diagnosed in the emergency department: Time to improve the evidence base. Thromb Res 2018 Dec;161:117-118. [doi: 10.1016/j.thromres.2017.11.022] [Medline: 29198735]
26. Bledsoe JR, Woller SC, Stevens SM, Aston V, Patten R, Allen T, et al. Management of Low-Risk Pulmonary Embolism Patients Without Hospitalization: The Low-Risk Pulmonary Embolism Prospective Management Study. Chest 2018 Feb 02. [doi: 10.1016/j.chest.2018.01.035] [Medline: 29410163]

Abbreviations

CTPA: Computed Tomography Pulmonary Angiography
ED: emergency department
EHR: electronic health record
ICD-9-CM: International Classification of Diseases, Clinical Modification codes, version 9
ICD-10-CM: International Classification of Diseases, Clinical Modification codes, version 10
PE: pulmonary embolism
V/Q: ventilation/perfusion lung

Edited by G Eysenbach; submitted 29.01.18; peer-reviewed by A Sheikhtaheri, M Stuckey, E Lang, A Press; comments to author 21.03.18; revised version received 16.05.18; accepted 06.07.18; published 25.10.18.

Please cite as:
Richardson S, Solomon P, O'Connell A, Khan S, Gong J, Makhnevich A, Qiu G, Zhang M, McGinn T
A Computerized Method for Measuring Computed Tomography Pulmonary Angiography Yield in the Emergency Department: Validation Study
JMIR Med Inform 2018;6(4):e44
URL: http://medinform.jmir.org/2018/4/e44/
doi:10.2196/medinform.9957
PMID:30361200

©Safiya Richardson, Philip Solomon, Alexander O'Connell, Sundas Khan, Jonathan Gong, Alex Makhnevich, Guang Qiu, Meng Zhang, Thomas McGinn. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 25.10.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Appropriateness of Hospital Admission for Emergency Department Patients with Bronchiolitis: Secondary Analysis

Gang Luo[1], PhD; Michael D Johnson[2], MD; Flory L Nkoy[2], MD, MS, MPH; Shan He[3], PhD; Bryan L Stone[2], MD, MS

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

[2]Department of Pediatrics, University of Utah, Salt Lake City, UT, United States

[3]Homer Warner Research Center, Intermountain Healthcare, Murray, UT, United States

Corresponding Author:
Gang Luo, PhD
Department of Biomedical Informatics and Medical Education
University of Washington
University of Washington Medicine South Lake Union
850 Republican Street, Building C, Box 358047
Seattle, WA, 98195
United States
Phone: 1 206 221 4596
Fax: 1 206 221 2671
Email: gangluo@cs.wisc.edu

Abstract

Background: Bronchiolitis is the leading cause of hospitalization in children under 2 years of age. Each year in the United States, bronchiolitis results in 287,000 emergency department visits, 32%-40% of which end in hospitalization. Frequently, emergency department disposition decisions (to discharge or hospitalize) are made subjectively because of the lack of evidence and objective criteria for bronchiolitis management, leading to significant practice variation, wasted health care use, and suboptimal outcomes. At present, no operational definition of appropriate hospital admission for emergency department patients with bronchiolitis exists. Yet, such a definition is essential for assessing care quality and building a predictive model to guide and standardize disposition decisions. Our prior work provided a framework of such a definition using 2 concepts, one on safe versus unsafe discharge and another on necessary versus unnecessary hospitalization.

Objective: The goal of this study was to determine the 2 threshold values used in the 2 concepts, with 1 value per concept.

Methods: Using Intermountain Healthcare data from 2005-2014, we examined distributions of several relevant attributes of emergency department visits by children under 2 years of age for bronchiolitis. Via a data-driven approach, we determined the 2 threshold values.

Results: We completed the first operational definition of appropriate hospital admission for emergency department patients with bronchiolitis. Appropriate hospital admissions include actual admissions with exposure to major medical interventions for more than 6 hours, as well as actual emergency department discharges, followed by an emergency department return within 12 hours ending in admission for bronchiolitis. Based on the definition, 0.96% (221/23,125) of the emergency department discharges were deemed unsafe. Moreover, 14.36% (432/3008) of the hospital admissions from the emergency department were deemed unnecessary.

Conclusions: Our operational definition can define the prediction target for building a predictive model to guide and improve emergency department disposition decisions for bronchiolitis in the future.

(JMIR Med Inform 2018;6(4):e10498) doi:10.2196/10498

KEYWORDS

appropriate hospital admission; bronchiolitis; emergency department; operational definition

XSL•FO
RenderX

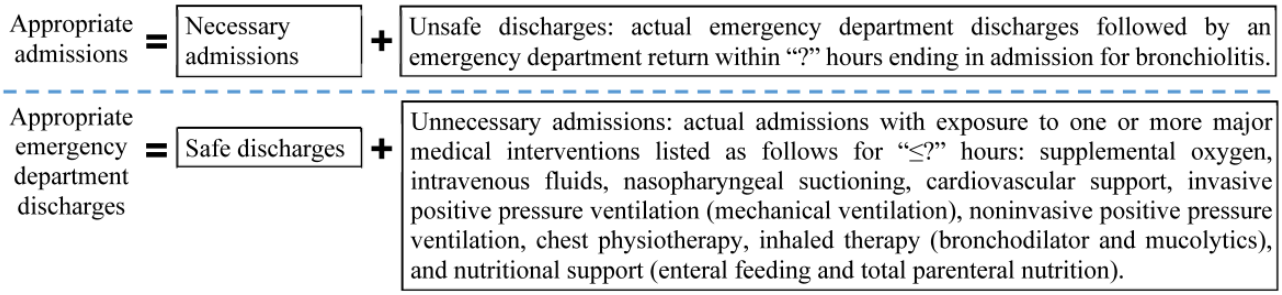
Introduction

Bronchiolitis is the inflammation of the bronchioles, the smallest air passages in the lungs, mainly seen in children under 2 years of age in response to viral respiratory infection. More than one-third of the children are diagnosed with bronchiolitis before 2 years of age [1]. Bronchiolitis is the leading cause of hospitalization in children under 2 years of age, and it is responsible for 16% of hospitalizations in this age group [2-5]. Each year in the United States, bronchiolitis incurs about 287,000 emergency department (ED) visits [6], 128,000 hospitalizations [2], and US $1.73 billion of total inpatient cost (2009) [2].

Around 32%-40% of the ED visits for bronchiolitis end in hospitalization [7-9]. As acknowledged in the current clinical guidelines for bronchiolitis [10,11], ED disposition decisions (to discharge or hospitalize) are often made subjectively because of the lack of evidence and objective criteria for bronchiolitis management [4,12]. This causes large practice variation [3,12-23], wasted health care use, increased iatrogenic risk, and suboptimal outcomes due to unnecessary admissions and unsafe discharges [15,21,24]. About 10% of the infants with bronchiolitis experience adverse events during hospitalization [25]. At present, no operational definition of appropriate hospital admission for ED patients with bronchiolitis exists [26]. Yet, such an operational definition is essential for assessing ED care quality and building a predictive model to guide and standardize disposition decisions [26].

Our prior work [26] has provided a framework of such an operational definition using 2 concepts: one on safe versus unsafe discharge and another on necessary versus unnecessary hospitalization (Figure 1). Each concept uses a threshold value to be determined. Appropriate admissions include both necessary admissions (actual admissions that are necessary) and unsafe discharges. Appropriate ED discharges include both safe discharges and unnecessary admissions. This study aims to determine the 2 threshold values in a data-driven way, to complete the first operational definition of appropriate hospital admission for ED patients with bronchiolitis, and to report the corresponding percentages of unnecessary admissions and unsafe discharges.

Figure 1. Definition framework of appropriate admission. ?: threshold value.



Methods

Study Design and Ethics Approval

In this study, we performed secondary analysis of retrospective data. The Institutional Review Boards of the University of Washington Medicine, University of Utah, and Intermountain Healthcare reviewed and approved this study and waived the need for informed consent for all patients.

Patient Population

Our patient cohort included children under 2 years of age who had ED encounters at 22 Intermountain Healthcare hospitals for bronchiolitis in 2005-2014. Intermountain Healthcare is the largest health care system in Utah, with 185 clinics and 22 hospitals providing ~85% of pediatric care available in Utah [27]. We used an approach similar to that used by Flaherman et al [28-30] to identify as many ED encounters for bronchiolitis as possible.

Several International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) discharge diagnosis codes, rather than only the discharge diagnosis code of bronchiolitis, can be possibly assigned to an ED encounter for bronchiolitis. Using methods used in prior studies [28-30], we included patients with an ED or hospital ICD-9-CM primary discharge diagnosis of bronchiolitis or bronchitis (466.x), viral pneumonia (480.x), adenoviral infection (079.0), rhinovirus infection (079.3), respiratory infection due to influenza (487.0, 487.1), respiratory syncytial virus (079.6), H1N1 influenza (488.1, 488.11, 488.12), influenza due to identified avian influenza virus (488, 488.0, 488.01, 488.02), or influenza due to novel influenza A (488.81, 488.82).

We also included all patients with any of the above as a nonprimary diagnosis, as long as the ICD-9-CM primary diagnosis was any of the following: apnea (786.03), shortness of breath (786.05), tachypnea (786.06), wheezing (786.07), other respiratory abnormalities (786.09), cough (786.2), fever (780.60, 780.61), acute nasopharyngitis (460), acute upper respiratory infections (465.x), other specified viral infection (079.89), urinary tract infection (599.0), pneumonia unspecified organism (486), unspecified viral infection (079.99), volume depletion (276.5x), or respiratory failure (518.81, 518.82).

Dataset

We extracted a clinical and administrative data set from Intermountain Healthcare's enterprise data warehouse. The data set included ED visit and hospitalization information of our patient cohort.

Data Analysis

To determine the threshold value used for defining unsafe discharges (Figure 1), we examined the length distribution of

XSL•FO
RenderX

the interval between an ED discharge and a return ED visit within 2 weeks ending in hospitalization for bronchiolitis [31,32]. In children under 2 years of age, bronchiolitis lasting longer than 2 weeks tends to result from new infection with a differing virus strain instead of persistent infection by the same virus strain [33].

To determine the threshold value used for defining unnecessary admissions (Figure 1), we examined the patients who were hospitalized for ≤12 hours and discharged with no readmission for bronchiolitis within 2 weeks. These patients tended to have been admitted unnecessarily. We used their median duration of using major medical interventions as a conservative threshold for using major medical interventions in all admissions. As shown in Figure 1, major medical interventions include supplemental oxygen, intravenous fluids, nasopharyngeal suctioning, cardiovascular support, invasive positive pressure ventilation (mechanical ventilation), noninvasive positive pressure ventilation, chest physiotherapy, inhaled therapy (bronchodilator and mucolytics), and nutritional support (enteral feeding and total parenteral nutrition) [26]. Every hospital admission with exposure to major medical interventions for no longer than the threshold was deemed unnecessary. During 2005-2012, Intermountain Healthcare iteratively modified its internal guidelines for bronchiolitis management in the ED and hospital several times, with an associated change in the distribution of the duration of using major medical interventions. After the beginning of 2013, significant changes in internal guidelines did not occur. Duration of using major medical interventions became stabilized. To compute the threshold value, we used 2013-2014 data with a stable distribution of duration of using major medical interventions. Both durations of hospitalization and using major medical interventions included only time in the hospital after the patient left the ED.

Results

Table 1 shows the demographic and clinical characteristics of our patient cohort: children under 2 years of age who had ED encounters for bronchiolitis. About 38.20% (14,292/37,417) of the ED visits for bronchiolitis ended in hospitalization.

Table 1. Demographic and clinical characteristics of our patient cohort.

Characteristic	Emergency department visits (n=37,417)	Emergency department discharges (n=23,125)	Emergency department visits ending in hospitalization (n=14,292)
Age, n (%)			
<2 months	4769 (12.75)	1646 (7.12)	3123 (21.85)
2 to <12 months	22,101 (59.07)	14,569 (63.00)	7532 (52.70)
12-24 months	10,547 (28.19)	6910 (29.88)	3637 (25.45)
Gender, n (%)			
Male	21,536 (57.56)	13,399 (57.94)	8137 (56.93)
Female	15,881 (42.44)	9733 (42.06)	6155 (43.07)
Race, n (%)			
American Indian or Alaska native	458 (1.22)	295 (1.28)	163 (1.14)
Asian	395 (1.06)	222 (0.96)	173 (1.21)
Black or African American	1017 (2.72)	664 (2.87)	353 (2.47)
Native Hawaiian or other Pacific islander	2209 (5.90)	1243 (5.38)	966 (6.76)
White	28,510 (76.20)	17,660 (76.37)	10,850 (75.92)
Unknown or not reported	4828 (12.90)	3041 (13.15)	1787 (12.50)
Ethnicity, n (%)			
Hispanic	9011 (24.08)	5975 (25.84)	3036 (21.24)
Non-Hispanic	18,823 (50.31)	11,278 (48.77)	7545 (52.79)
Unknown or not reported	9583 (25.61)	5872 (25.39)	3711 (25.97)
Insurance, n (%)			
Private	22,162 (59.23)	13,052 (56.44)	9110 (63.74)
Public	13,448 (35.94)	8729 (37.75)	4719 (33.02)
Self-paid or charity	1807 (4.82)	1344 (5.81)	463 (3.24)
Asthma, n (%)	2246 (6.00)	883 (3.82)	1363 (9.54)
Chronic complex condition [34], n (%)	2040 (5.45)	365 (1.58)	1675 (11.72)

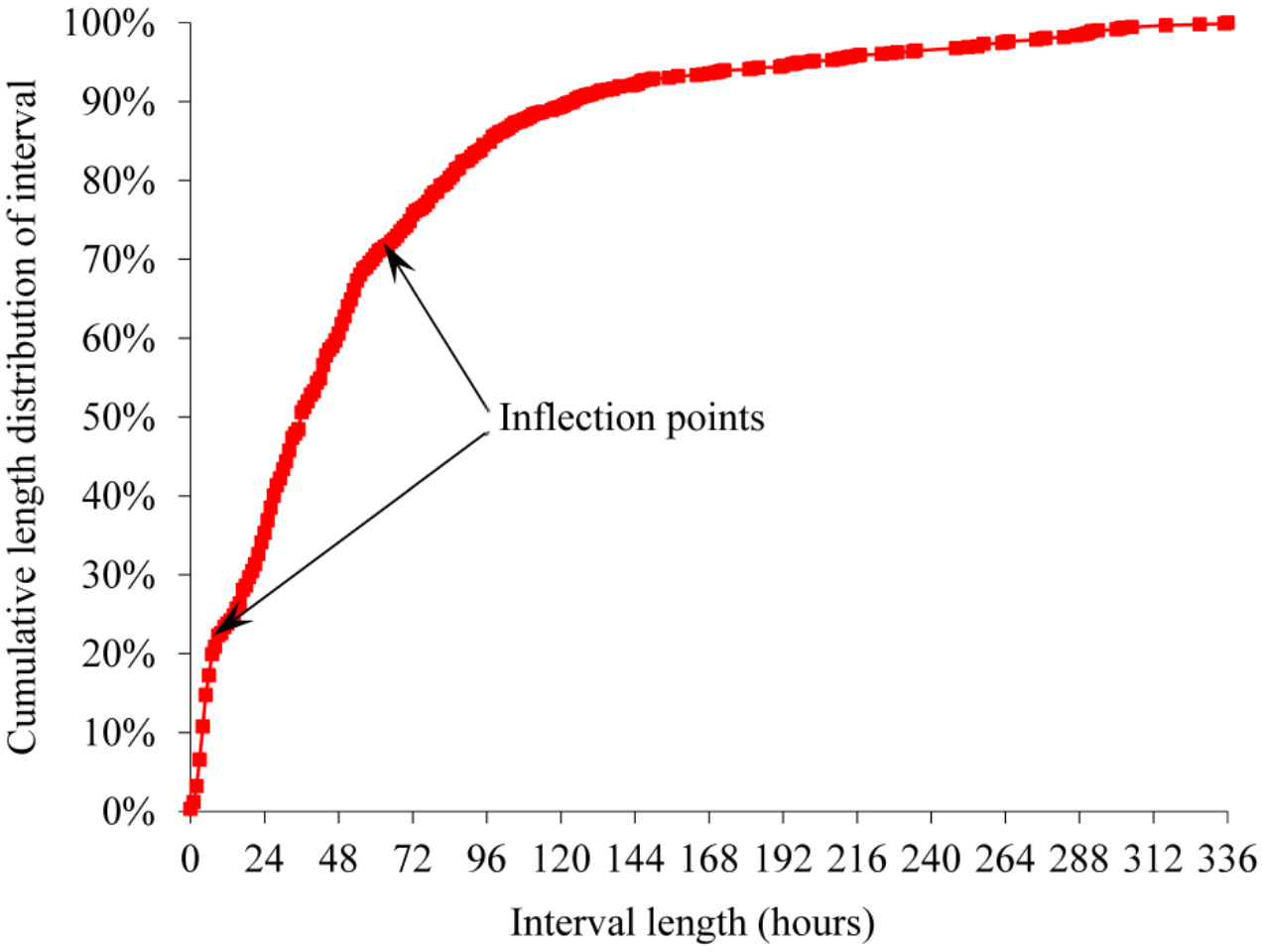
XSL•FO
RenderX

Figures 2 and 3 show cumulative length distributions in hours of the interval between an ED discharge and a return ED visit within 2 weeks ending in hospitalization for bronchiolitis. Figure 4 shows probability density function of the interval length. The probability density function was relatively large until the interval length reached the cumulative distribution curve's inflection point at about 10 to 12 hours and became smaller afterward. The cumulative distribution curve seemed to have 2 inflection points, suggesting 3 underlying distributions. As indicated by the dotted curve in Figure 4, the 3 distributions are postulated to represent an early ED return after an inappropriate ED discharge, natural disease progression in a subgroup of appropriate ED discharges, and an even later ED return due to a new viral infection after an appropriate ED discharge, respectively. When selecting the threshold value for defining unsafe discharges (Figure 1), we wanted our choice to capture the majority of unsafe discharges while avoiding contamination with ED returns not due to unsafe discharges. To help make the choice, we used the probability density function that has a local minimum at the interval length of 10 to 12 hours. We chose 12 hours, which fulfilled our selection criteria. Accordingly, 0.96% (221/23,125) of the ED discharges were followed by an ED return within 12 hours resulting in hospital admission for bronchiolitis and were deemed unsafe ED discharges.

In 2013-2014, no major medical intervention was applied in 6.45% (194/3008) of the cases of hospitalization from the ED for bronchiolitis. In another 7.91% (238/3008) of the cases, 1 or more major medical interventions were applied, but the duration of using them was ≤6 hours. Among the patients hospitalized in 2013-2014, 8.31% (250/3008) were hospitalized for ≤12 hours and discharged with no readmission for bronchiolitis within 2 weeks. Figure 5 shows the distribution of duration of using major medical interventions in these patients. Median duration of using major medical interventions was 6 hours, which we used as the threshold value for defining unnecessary admissions (Figure 1). Accordingly, 14.36% (432/3008) of the hospital admissions from the ED in 2013-2014 incurred exposure to major medical interventions for no longer than this threshold and were deemed unnecessary.

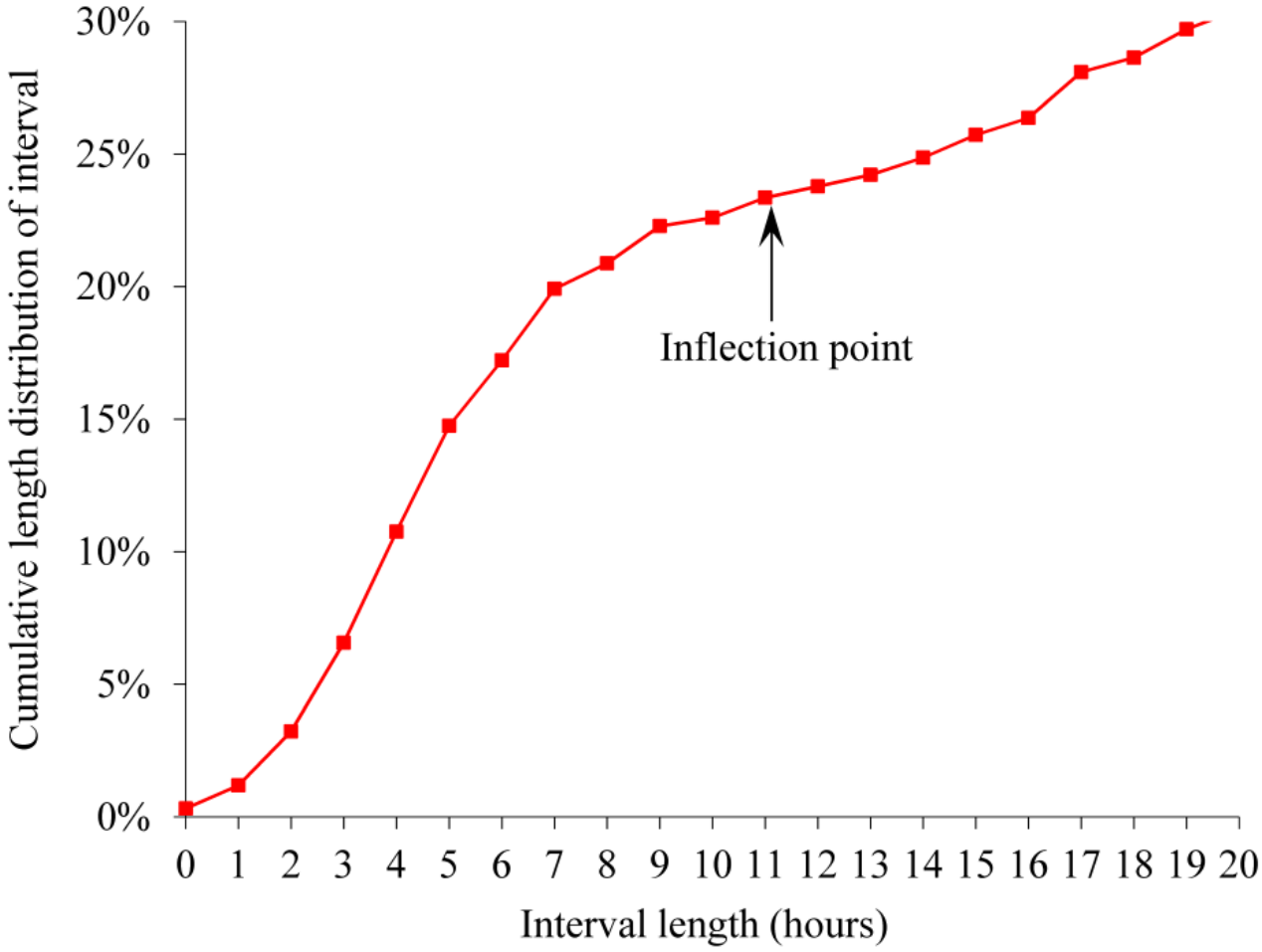
By filling in the 2 threshold values in our definition framework (Figure 1) [26], we completed the first operational definition of appropriate hospital admission for ED patients with bronchiolitis. Appropriate hospital admissions include actual admissions with exposure to major medical interventions for more than 6 hours, as well as actual ED discharges followed by ED return within 12 hours ending in admission for bronchiolitis. Putting unsafe ED discharges and unnecessary admissions together, 6.08% of the ED disposition decisions for bronchiolitis were deemed inappropriate.

Figure 2. Cumulative length distribution of interval between emergency department discharge and return visit within 2 weeks ending in hospitalization for bronchiolitis.



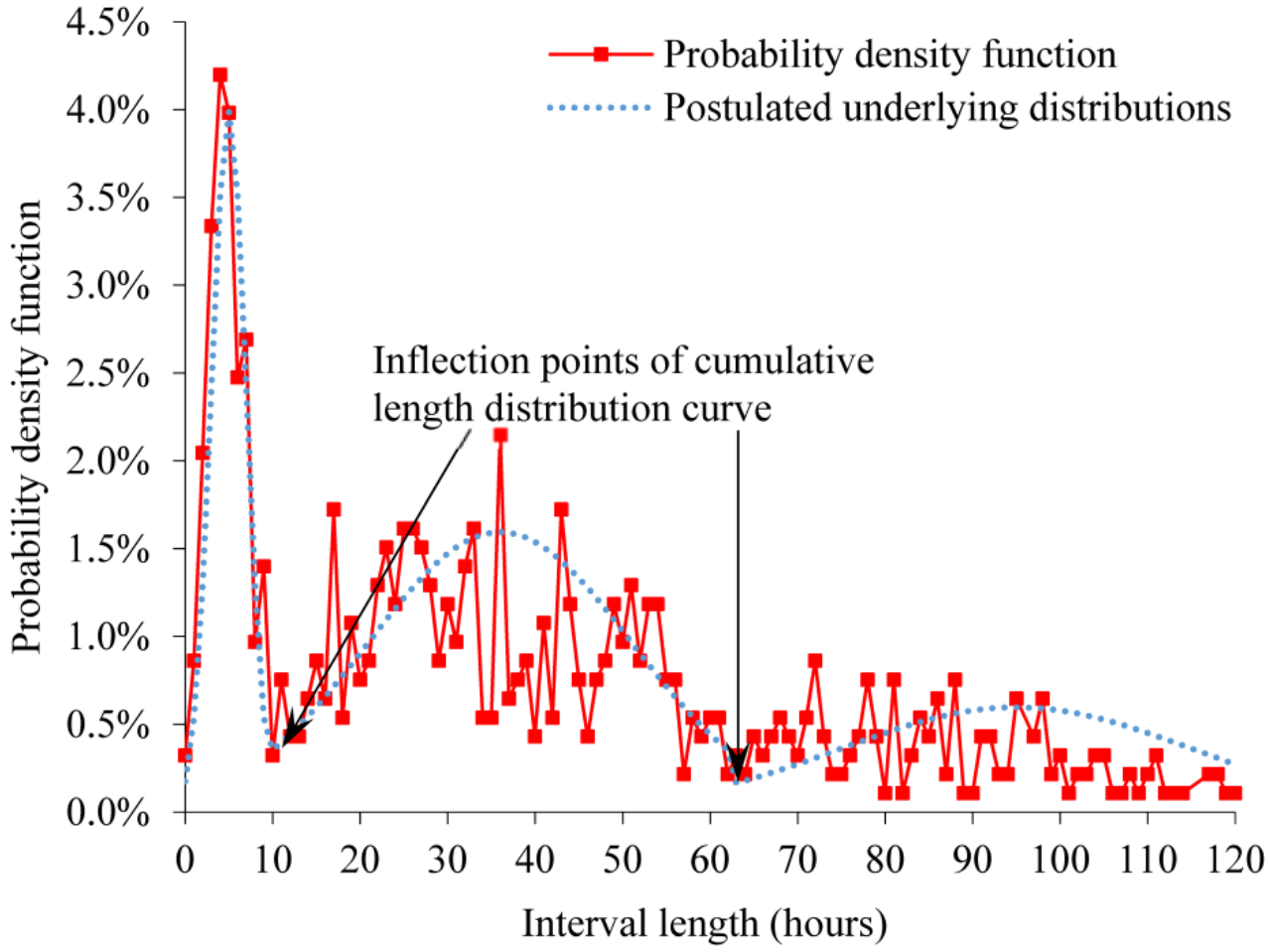
XSL•FO
RenderX

Figure 3. Cumulative length distribution of interval between emergency department discharge and return visit within 2 weeks ending in hospitalization for bronchiolitis, when the interval length is ≤20 hours.



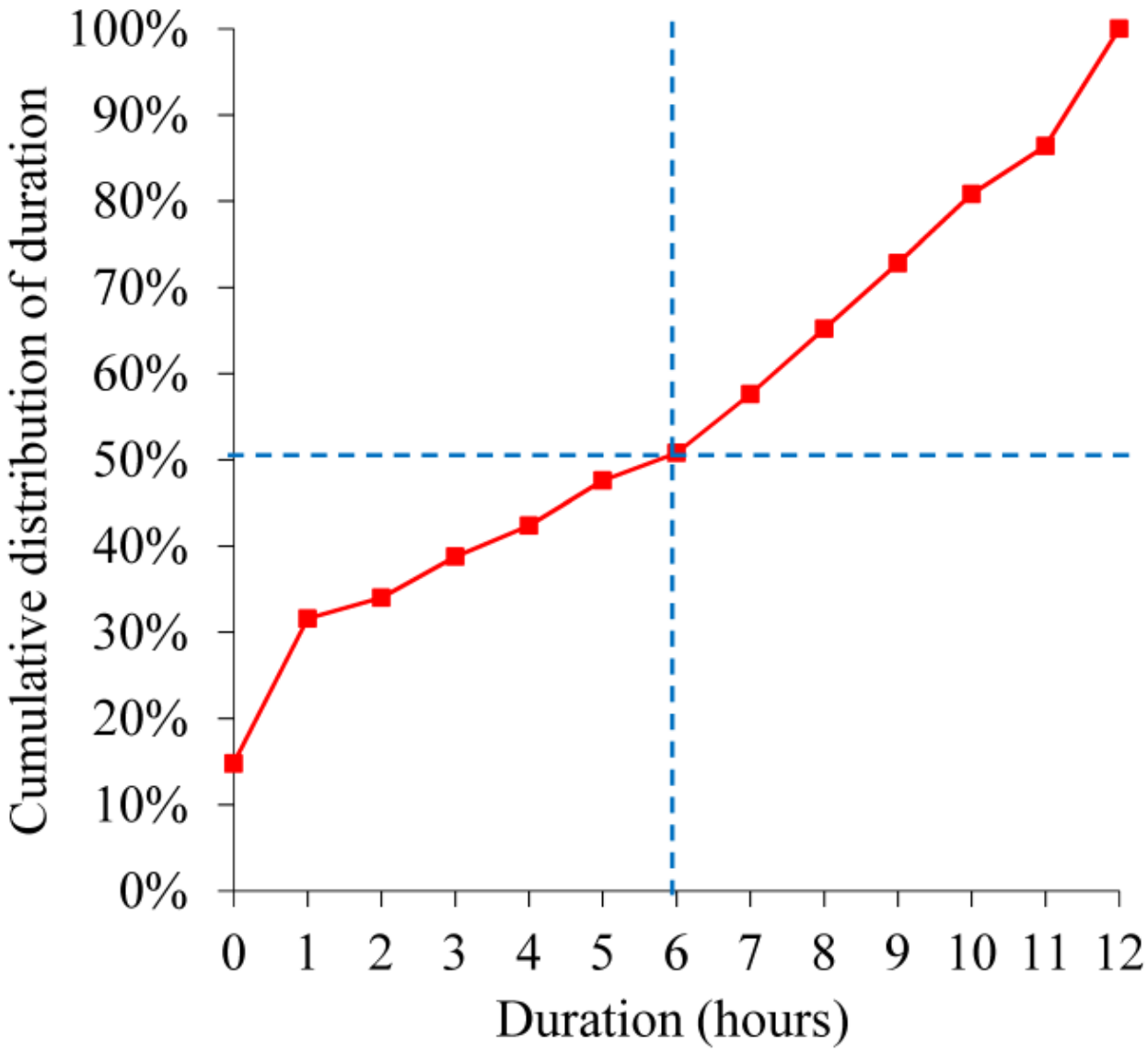
XSL•FO
RenderX

Figure 4. Probability density function of interval between emergency department discharge and return visit within 2 weeks ending in hospitalization for bronchiolitis, when the interval length is ≤120 hours.



XSL•FO
RenderX

Figure 5. Distribution of duration of using major medical interventions among patients who were hospitalized for ≤12 hours in 2013-2014 and discharged with no readmission for bronchiolitis within 2 weeks.



Discussion

Principal Findings

We completed the first operational definition of appropriate hospital admission for ED patients with bronchiolitis. The definition uses 2 concepts, one on safe versus unsafe discharge and another on necessary versus unnecessary hospitalization. Based on the definition, we found that many ED disposition decisions for bronchiolitis were deemed inappropriate. Our findings highlight opportunities for improving ED disposition decisions and the need to build a model to predict appropriate admission. The model could become the foundation of a decision support tool to help make appropriate ED disposition decisions for bronchiolitis, improve bronchiolitis outcomes, and cut health care costs [26]. Although the model could be built without using the ED physician's initial, tentative disposition decision as an input variable, the model would likely be more accurate if this variable is included. In either case, physicians can use the model's output to give a second thought on their initial, tentative disposition decision.

Comparison With Prior Work

Some aspects of our findings are similar to those in previous studies. In our data set, about 38.20% (14,292/37,417) of the ED visits for bronchiolitis ended in hospitalization. This percentage is close to the corresponding percentages (32%-40%) reported in the literature [7-9]. For 30 EDs in 15 US states, Norwood et al [35] have presented the length distribution of interval between an ED discharge and a return ED or clinic visit within 2 weeks for bronchiolitis. That distribution is similar to the one we have shown in Figure 2, which presents the length distribution of interval between an ED discharge and a return ED visit within 2 weeks ending in hospitalization for bronchiolitis.

Some of our findings are different from those in previous studies. In our dataset, 14.36% (432/3008) of the hospital admissions from the ED in 2013-2014 were deemed

XSL•FO
RenderX

unnecessary. This percentage is smaller than the corresponding percentages (20%-29%) suggested in the literature [36,37]. Intermountain Healthcare has multiple collaborative partnerships among its EDs and hospitals to ensure that pediatric specialty care is co-ordinated and not focused just in a tertiary pediatric hospital. Several quality improvement projects for bronchiolitis management were completed during 2005-2012, impacting the ED and hospital care of children in multiple hospitals within Intermountain Healthcare. The average quality of ED disposition decisions for bronchiolitis made at Intermountain Healthcare could be higher than that of ED disposition decisions for bronchiolitis made at some other health care systems, particularly if those health care systems employ few pediatricians in their EDs.

Limitations

This study has several limitations. One limitation is that the study used data from a single health care system, Intermountain Healthcare, and our results may not be generalized to other health care systems. Notably, most Intermountain Healthcare hospitals are at a high elevation (more than 4000 feet above sea level). This may result in increased incidence of hypoxia. About 46% of the patients hospitalized with bronchiolitis at Intermountain Healthcare are discharged on home oxygen for outpatient management. Protocols are in place to facilitate brief hospitalizations if oxygen is the only intervention a patient needs in the hospital [30]. In the future, it would be desirable to use data from other health care systems to validate our operational definition of appropriate hospital admission for ED patients with bronchiolitis. As indicated by similarities between our findings and those in previous studies, we do not expect such validation to significantly change our results. Intermountain Healthcare is a large health care system with EDs at 22 heterogeneous hospitals spread over a large geographic area, ranging from community metropolitan and rural hospitals attended by general practitioners and family doctors with constrained pediatric resources to tertiary care children's and general hospitals in urban areas attended by subspecialists [26]. Each hospital has a different patient population, geographic location, staff composition, scope of services, and cultural background. This variation provides a realistic situation for finding factors generalizable to other hospitals across the United States.

Another limitation of this study is that Intermountain Healthcare does not have complete clinical and administrative data on all of its patients although it is an integrated health care system. Within 2 weeks of a visit to an Intermountain Healthcare ED for bronchiolitis, a patient could use a non-Intermountain Healthcare hospital for bronchiolitis again. If this occurred, our data set would miss the information on health care use that occurred at a non-Intermountain Healthcare hospital. Including data from non-Intermountain Healthcare hospitals may lead to different results. Nevertheless, we do not expect this to greatly change the accuracy of our results. Intermountain Healthcare provides ~85% of the pediatric care available in Utah [27]. Thus, our dataset is reasonably, although not 100%, complete in terms of capturing bronchiolitis patients' use of hospitals at Utah.

A third limitation is that this study does not consider factors, such as preference of the patient's parents, patient transportation availability, and time of day, while defining appropriate hospital admission. Many of these factors are often undocumented in patient records. For some hospital admissions from the ED that were deemed unnecessary based on our operational definition of appropriate hospital admission, the original admission decisions could be made due to these factors.

Conclusions

We provided the first operational definition of appropriate hospital admission for ED patients with bronchiolitis. Our operational definition can define the prediction target for building a predictive model in the future with the goal of standardizing and improving ED disposition decisions for bronchiolitis.

Acknowledgments

GL, BLS, MDJ, FLN, and SH were partially supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number R21HL128875. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

GL was mainly responsible for the paper. GL conceptualized and designed the study, performed literature review and data analysis, and wrote the paper. BLS, MDJ, and FLN provided feedback on various medical issues, contributed to conceptualizing the presentation, and revised the paper. SH took part in retrieving the Intermountain Healthcare data set and interpreting its detected peculiarities. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Zorc JJ, Hall CB. Bronchiolitis: recent evidence on diagnosis and management. Pediatrics 2010 Feb;125(2):342-349. [doi: 10.1542/peds.2009-2092] [Medline: 20100768]

XSL•FO
RenderX

2. Hasegawa K, Tsugawa Y, Brown DFM, Mansbach JM, Camargo CA. Trends in bronchiolitis hospitalizations in the United States, 2000-2009. Pediatrics 2013 Jul;132(1):28-36 [FREE Full text] [doi: 10.1542/peds.2012-3877] [Medline: 23733801]
3. Mansbach JM, Emond JA, Camargo CA. Bronchiolitis in US emergency departments 1992 to 2000: epidemiology and practice variation. Pediatr Emerg Care 2005 Apr;21(4):242-247. [Medline: 15824683]
4. Parker MJ, Allen U, Stephens D, Lalani A, Schuh S. Predictors of major intervention in infants with bronchiolitis. Pediatr Pulmonol 2009 Apr;44(4):358-363. [doi: 10.1002/ppul.21010] [Medline: 19283838]
5. Shay DK, Holman RC, Newman RD, Liu LL, Stout JW, Anderson LJ. Bronchiolitis-associated hospitalizations among US children, 1980-1996. JAMA 1999 Oct 20;282(15):1440-1446. [Medline: 10535434]
6. Hasegawa K, Tsugawa Y, Brown DFM, Mansbach JM, Camargo CA. Temporal trends in emergency department visits for bronchiolitis in the United States, 2006 to 2010. Pediatr Infect Dis J 2014 Jan;33(1):11-18 [FREE Full text] [doi: 10.1097/INF.0b013e3182a5f324] [Medline: 23934206]
7. Marlais M, Evans J, Abrahamson E. Clinical predictors of admission in infants with acute bronchiolitis. Arch Dis Child 2011 Jul;96(7):648-652. [doi: 10.1136/adc.2010.201079] [Medline: 21339199]
8. Laham FR, Trott AA, Bennett BL, Kozinetz CA, Jewell AM, Garofalo RP, et al. LDH concentration in nasal-wash fluid as a biochemical predictor of bronchiolitis severity. Pediatrics 2010 Feb;125(2):e225-e233 [FREE Full text] [doi: 10.1542/peds.2009-0411] [Medline: 20100751]
9. Corneli HM, Zorc JJ, Holubkov R, Bregstein JS, Brown KM, Mahajan P, Bronchiolitis Study Group for the Pediatric Emergency Care Applied Research Network. Bronchiolitis: clinical characteristics associated with hospitalization and length of stay. Pediatr Emerg Care 2012 Feb;28(2):99-103. [doi: 10.1097/PEC.0b013e3182440b9b] [Medline: 22270499]
10. Scottish IGN. Bronchiolitis in children. A national clinical guideline. URL: https://www.guidelinecentral.com/summaries/bronchiolitis-in-children-a-national-clinical-guideline/ [accessed 2015-09-21] [WebCite Cache ID 6bhQvWHzA]
11. Bronchiolitis Guideline Team, Cincinnati Children's Hospital Medical Center. Evidence-based care guideline for management of first time episode bronchiolitis in infants less than 1 year of age. URL: http://www.guideline.gov/content.aspx?id=34411 [accessed 2015-09-20] [WebCite Cache ID 6bhOyjITC]
12. Brand PL, Vaessen-Verberne AA. Differences in management of bronchiolitis between hospitals in The Netherlands. Dutch Paediatric Respiratory Society. Eur J Pediatr 2000 May;159(5):343-347. [Medline: 10834519]
13. Ducharme FM. Management of acute bronchiolitis. BMJ 2011 Apr 06;342:d1658. [Medline: 21471173]
14. Behrendt CE, Decker MD, Burch DJ, Watson PH. International variation in the management of infants hospitalized with respiratory syncytial virus. International RSV Study Group. Eur J Pediatr 1998 Mar;157(3):215-220. [Medline: 9537488]
15. Chamberlain JM, Patel KM, Pollack MM. Association of emergency department care factors with admission and discharge decisions for pediatric patients. J Pediatr 2006 Nov;149(5):644-649. [doi: 10.1016/j.jpeds.2006.05.047] [Medline: 17095336]
16. Johnson DW, Adair C, Brant R, Holmwood J, Mitchell I. Differences in admission rates of children with bronchiolitis by pediatric and general emergency departments. Pediatrics 2002 Oct;110(4):e49. [Medline: 12359822]
17. Mallory MD, Shay DK, Garrett J, Bordley WC. Bronchiolitis management preferences and the influence of pulse oximetry and respiratory rate on the decision to admit. Pediatrics 2003 Jan;111(1):e45-e51. [Medline: 12509594]
18. Plint AC, Johnson DW, Wiebe N, Bulloch B, Pusic M, Joubert G, et al. Practice variation among pediatric emergency departments in the treatment of bronchiolitis. Acad Emerg Med 2004 Apr;11(4):353-360 [FREE Full text] [Medline: 15064208]
19. Vogel AM, Lennon DR, Harding JE, Pinnock RE, Graham DA, Grimwood K, et al. Variations in bronchiolitis management between five New Zealand hospitals: can we do better? J Paediatr Child Health 2003;39(1):40-45. [Medline: 12542811]
20. Willson DF, Horn SD, Hendley JO, Smout R, Gassaway J. Effect of practice variation on resource utilization in infants hospitalized for viral lower respiratory illness. Pediatrics 2001 Oct;108(4):851-855. [Medline: 11581435]
21. Willson DF, Jiao JH, Hendley JO, Donowitz L. Invasive monitoring in infants with respiratory syncytial virus infection. J Pediatr 1996 Mar;128(3):357-362. [Medline: 8774504]
22. Wang EE, Law BJ, Boucher FD, Stephens D, Robinson JL, Dobson S, et al. Pediatric Investigators Collaborative Network on Infections in Canada (PICNIC) study of admission and management variation in patients hospitalized with respiratory syncytial viral lower respiratory tract infection. J Pediatr 1996 Sep;129(3):390-395. [Medline: 8804328]
23. Christakis DA, Cowan CA, Garrison MM, Molteni R, Marcuse E, Zerr DM. Variation in inpatient diagnostic testing and management of bronchiolitis. Pediatrics 2005 Apr;115(4):878-884. [doi: 10.1542/peds.2004-1299] [Medline: 15805359]
24. Mansbach JM, Clark S, Christopher NC, LoVecchio F, Kunz S, Acholonu U, et al. Prospective multicenter study of bronchiolitis: predicting safe discharges from the emergency department. Pediatrics 2008 Apr;121(4):680-688. [doi: 10.1542/peds.2007-1418] [Medline: 18381531]
25. McBride SC, Chiang VW, Goldmann DA, Landrigan CP. Preventable adverse events in infants hospitalized with bronchiolitis. Pediatrics 2005 Sep;116(3):603-608. [doi: 10.1542/peds.2004-2387] [Medline: 16140699]
26. Luo G, Stone BL, Johnson MD, Nkoy FL. Predicting Appropriate Admission of Bronchiolitis Patients in the Emergency Department: Rationale and Methods. JMIR Res Protoc 2016 Mar 07;5(1):e41 [FREE Full text] [doi: 10.2196/resprot.5155] [Medline: 26952700]

XSL•FO
RenderX

27. Byington CL, Reynolds CC, Korgenski K, Sheng X, Valentine KJ, Nelson RE, et al. Costs and infant outcomes after implementation of a care process model for febrile infants. Pediatrics 2012 Jul;130(1):e16-e24 [FREE Full text] [doi: 10.1542/peds.2012-0127] [Medline: 22732178]
28. Flaherman VJ, Ragins AI, Li SX, Kipnis P, Masaquel A, Escobar GJ. Frequency, duration and predictors of bronchiolitis episodes of care among infants ≥32 weeks gestation in a large integrated healthcare system: a retrospective cohort study. BMC Health Serv Res 2012 Jun 08;12:144 [FREE Full text] [doi: 10.1186/1472-6963-12-144] [Medline: 22682080]
29. Mittal V, Darnell C, Walsh B, Mehta A, Badawy M, Morse R, et al. Inpatient bronchiolitis guideline implementation and resource utilization. Pediatrics 2014 Mar;133(3):e730-e737 [FREE Full text] [doi: 10.1542/peds.2013-2881] [Medline: 24534398]
30. Sandweiss DR, Mundorff MB, Hill T, Wolfe D, Greene T, Andrews S, et al. Decreasing hospital length of stay for bronchiolitis by using an observation unit and home oxygen therapy. JAMA Pediatr 2013 May;167(5):422-428. [doi: 10.1001/jamapediatrics.2013.1435] [Medline: 23479000]
31. Halfon P, Eggli Y, van Melle G, Chevalier J, Wasserfallen JB, Burnand B. Measuring potentially avoidable hospital readmissions. J Clin Epidemiol 2002 Jun;55(6):573-587. [Medline: 12063099]
32. Heggestad T, Lilleeng SE. Measuring readmissions: focus on the time factor. Int J Qual Health Care 2003 Apr;15(2):147-154. [Medline: 12705708]
33. Jartti T, Lee W, Pappas T, Evans M, Lemanske RF, Gern JE. Serial viral infections in infants with recurrent respiratory illnesses. Eur Respir J 2008 Aug;32(2):314-320 [FREE Full text] [doi: 10.1183/09031936.00161907] [Medline: 18448489]
34. Feudtner C, Feinstein JA, Zhong W, Hall M, Dai D. Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. BMC Pediatr 2014 Aug 08;14:199 [FREE Full text] [doi: 10.1186/1471-2431-14-199] [Medline: 25102958]
35. Norwood A, Mansbach JM, Clark S, Waseem M, Camargo CA. Prospective multicenter study of bronchiolitis: predictors of an unscheduled visit after discharge from the emergency department. Acad Emerg Med 2010 Apr;17(4):376-382 [FREE Full text] [doi: 10.1111/j.1553-2712.2010.00699.x] [Medline: 20370776]
36. Shaw KN, Bell LM, Sherman NH. Outpatient assessment of infants with bronchiolitis. Am J Dis Child 1991 Feb;145(2):151-155. [Medline: 1994678]
37. Shafik MH, Seoudi TMM, Raway TS, Al Harbash NZ, Ahmad MMA, Al Mutairi HF. Appropriateness of pediatric hospitalization in a general hospital in Kuwait. Med Princ Pract 2012;21(6):516-521 [FREE Full text] [doi: 10.1159/000339084] [Medline: 22678120]

Abbreviations

ED: emergency department
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification

Edited by G Eysenbach; submitted 24.03.18; peer-reviewed by A Ialynytchev; accepted 08.10.18; published 05.11.18.

Please cite as:
Luo G, Johnson MD, Nkoy FL, He S, Stone BL
Appropriateness of Hospital Admission for Emergency Department Patients with Bronchiolitis: Secondary Analysis
JMIR Med Inform 2018;6(4):e10498
URL: http://medinform.jmir.org/2018/4/e10498/
doi:10.2196/10498
PMID:30401659

©Gang Luo, Michael D Johnson, Flory L Nkoy, Shan He, Bryan L Stone. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 05.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Data for Research Through a Nationwide Privacy-Proof System in Belgium: Design and Implementation

Nicolas Delvaux[1], MD; Bert Aertgeerts[1], MD, PhD; Johan CH van Bussel[2], PhD; Geert Goderis[1], MD, PhD; Bert Vaes[1], MD, PhD; Mieke Vermandere[1], MD, PhD

[1]Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[2]Sciensano, Brussels, Belgium

Corresponding Author:
Nicolas Delvaux, MD
Department of Public Health and Primary Care
KU Leuven
Kapucijnenvoer 33
Blok J, PB 7001
Leuven, B-3000
Belgium
Phone: 32 16379885
Fax: 32 16337480
Email: nicolas.delvaux@kuleuven.be

Abstract

Background: Health data collected during routine care have important potential for reuse for other purposes, especially as part of a learning health system to advance the quality of care. Many sources of bias have been identified through the lifecycle of health data that could compromise the scientific integrity of these data. New data protection legislation requires research facilities to improve safety measures and, thus, ensure privacy.

Objective: This study aims to address the question on how health data can be transferred from various sources and using multiple systems to a centralized platform, called Healthdata.be, while ensuring the accuracy, validity, safety, and privacy. In addition, the study demonstrates how these processes can be used in various research designs relevant for learning health systems.

Methods: The Healthdata.be platform urges uniformity of the data registration at the primary source through the use of detailed clinical models. Data retrieval and transfer are organized through end-to-end encrypted electronic health channels, and data are encoded using token keys. In addition, patient identifiers are pseudonymized so that health data from the same patient collected across various sources can still be linked without compromising the deidentification.

Results: The Healthdata.be platform currently collects data for >150 clinical registries in Belgium. We demonstrated how the data collection for the Belgian primary care morbidity register INTEGO is organized and how the Healthdata.be platform can be used for a cluster randomized trial.

Conclusions: Collecting health data in various sources and linking these data to a single patient is a promising feature that can potentially address important concerns on the validity and quality of health data. Safe methods of data transfer without compromising privacy are capable of transporting these data from the primary data provider or clinician to a research facility. More research is required to demonstrate that these methods improve the quality of data collection, allowing researchers to rely on electronic health records as a valid source for scientific data.

(JMIR Med Inform 2018;6(4):e11428) doi:10.2196/11428

KEYWORDS

electronic health records; health information exchange; health information interoperability; learning health systems; medical record linkage

XSL•FO
RenderX

Introduction

More than a decade ago, the Institute of Medicine introduced the "learning health system" (LHS) in response to the challenges on how to generate and apply the best evidence to guide health care choices [1]. An important aim of the LHS is to enable the use of routinely collected health data for knowledge generation not only to ensure innovation in health care but also for the quality, safety, and value. In the LHS cycle, the analysis of routine health data creates new insights, which are then introduced back to health care providers through quality improvement tools such as clinical decision support, feedback, or audit reports. Even though the gold standard for measuring the effectiveness in health care has always been the randomized clinical trial, increasing attention is being given to registries and health data to contribute to evidence-based practice [2]. The use of health data recorded in the electronic health record (EHR) for research could help bridge the gap between evidence generated in controlled experiments and its application in daily clinical practice [3].

Even though the type of data recorded for research and the data stored in EHRs are similar, the use of health data poses some important problems. Concerns regarding the data quality and validity, completeness of data capture, and lack of interoperability have been identified as important barriers to the use of EHRs for clinical research [4]. Experiences from European and American efforts in the use or reuse of health data from EHRs have identified several important challenges [5,6]. Sources of bias included health care system bias, variations in EHR system functionalities and layout, and data extraction tools. These concerns have prompted opinions that the reuse of data for purposes other than that for which they were originally collected may be inappropriate [7]. Moreover, the European General Data Protection Regulation has further restricted the reuse of health data for research purposes [8,9]. Recent recommendations and guidance have proposed solutions to these concerns, but many challenges remain unanswered [10,11].

Perhaps, one of the most important challenges to the use of health data in clinical research is the persistent divide between clinicians (data providers) and researchers (data scientists) [12]. The incapacity to communicate between engineers and researchers, on the one hand, and clinicians, on the other hand, has resulted in disconnect between the world of research in medical informatics and the true problems in health care. In the evaluation and management of the health data for specific research questions, it is important that researchers understand the accuracy, comprehensiveness, retrievability, and specificity of health data recorded during routine care. On the other hand, clinicians need to be aware of the potential of the health data they record or manage and the implications that inconsistent or missing recordings may have on the reusability. Important facilitators to bridging this divide are enabling semantic interoperability, creating an environment of safe and reliable data transfer, ensuring privacy and security, and incentivizing valid and complete data capture [10]. In addition, experiences in the use of diverse sources of health data have led to a better understanding of data provenance (understanding of the authoritative source of a given data element of interest). For instance, if the measure of interest is whether a patient took a certain drug, then the best source for this outcome may not be the clinician's order entry data but instead the nursing medication administration record. In this sense, data collection tools that allow the aggregation of health data across sources are important enablers of the LHS.

In Belgium, >150 clinical registries actively collect health data from multiple sources such as primary care facilities, laboratories, hospitals, and radiology centers. Moreover, there are multiple information systems or EHRs available for each of these sources. For example, for primary care practices alone, at least, 8 different EHRs are available. In 2012, the Scientific Institute of Public Health was charged with centralizing and improving these clinical registries as part of the national electronic health (eHealth) action plan in a new platform named Healthdata.be. The challenge for this task was to develop a system that allows the integration of data from diverse sources and collects them through multiple systems by clinicians during routine care, while ensuring the accuracy, validity, safety, and privacy of the data. This study addresses the following questions:

- How can health data be transferred from various original sources of entry to a centralized platform for reuse and what efforts can be done to limit sources of bias?
- How can health data within the LHS be used for various research designs?

This study will describe elements of the Healthdata.be project designed for data extraction, data transfer, and data processing. Subsequently, we will demonstrate how Healthdata.be was used in the INTEGO primary care morbidity registry [13] and in a cluster randomized trial in primary care [14].

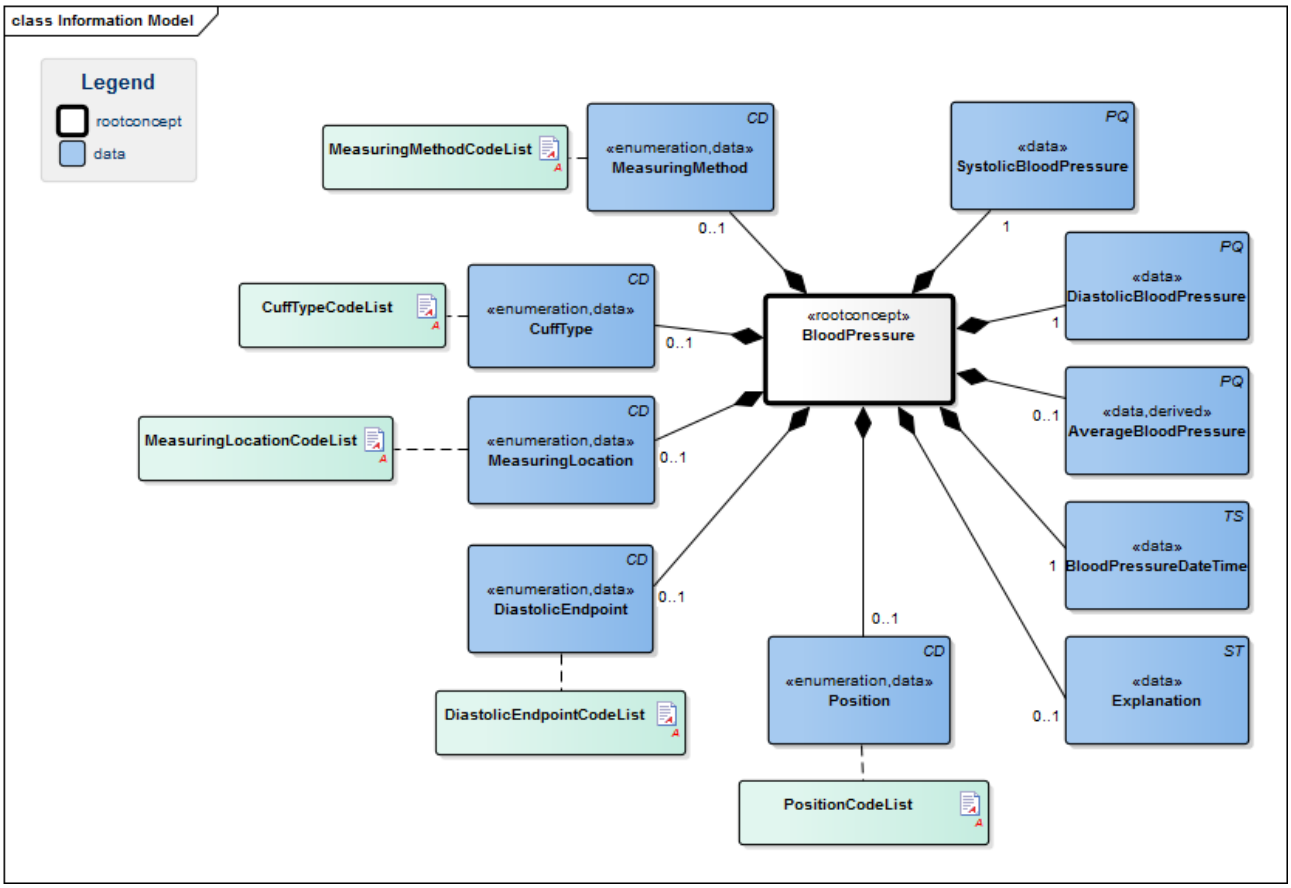
Methods

Data Structure and Semantic Interoperability

Health data are at the core of both EHRs and clinical research registries. However, to collect these data in a meaningful manner, these must have the same structure, use interoperable terminologies, and be documented using a detailed clinical model (DCM) [15,16]. DCMs provide detailed specifications of medical concepts in a given context and specify precisely the terminology to be used in terms of technical standards, reference models, and platforms [17,18]. They define all structured elements and attributes of a concept, including their relationships to the root concept, their data types, and the code lists that can be used. Where possible, code lists include internationally accepted coding such as logical observation identifiers names and codes (LOINC), Systematized Nomenclature of Medicine--Clinical Terms, International Classification of Diseases, international classification for primary care, etc. Figure 1 illustrates the DCM for the concept *blood pressure*, including all the associated data elements and their code lists as designed by the Netherlands Federation of University Medical Centres.

XSL•FO
RenderX

Figure 1. The detailed clinical model for the root concept blood pressure. CD: coded descriptor; PQ: physical quantity; TS: timestamp; ST: string (free) text. Source: https://www.healthdata.be/doc/cbb/index.php5/Be.en.hd.BloodPressure.



The content standardization of scientific data collections, using a DCM, contributes to the enhanced data quality and correct interpretation of data for research [19]. In addition, a substantial part of the information collected in the context of clinical registries is the continuity of care concepts (eg, diagnoses, medication, and laboratory test results), implying a certain overlap in content between different registries. Aligning all registries with DCMs, therefore, enables the harmonization across projects, allowing for the maximal reuse of existing data. As they are independent of technical aspects, such as message format, system, and network, DCMs can be considered technically neutral. Therefore, they can be used as a semantic layer between communication standards on the side of the primary source, and the parameters defined within different registries on the other side. Given the technological complexity and the variety of legacy systems, a stepwise approach for the registry standardization is proposed. In the first phase, existing registries are simply mapped to the DCMs. To maximize the degree of semantic interoperability, the registry structure can then be adjusted to comply with the logic defined in the DCMs. Registry variables that do not align with the DCM logic can still be completed using the manual interface. In the last phase, registries will be fully DCM-based, allowing the data provider to complete all registrations at the point of care.

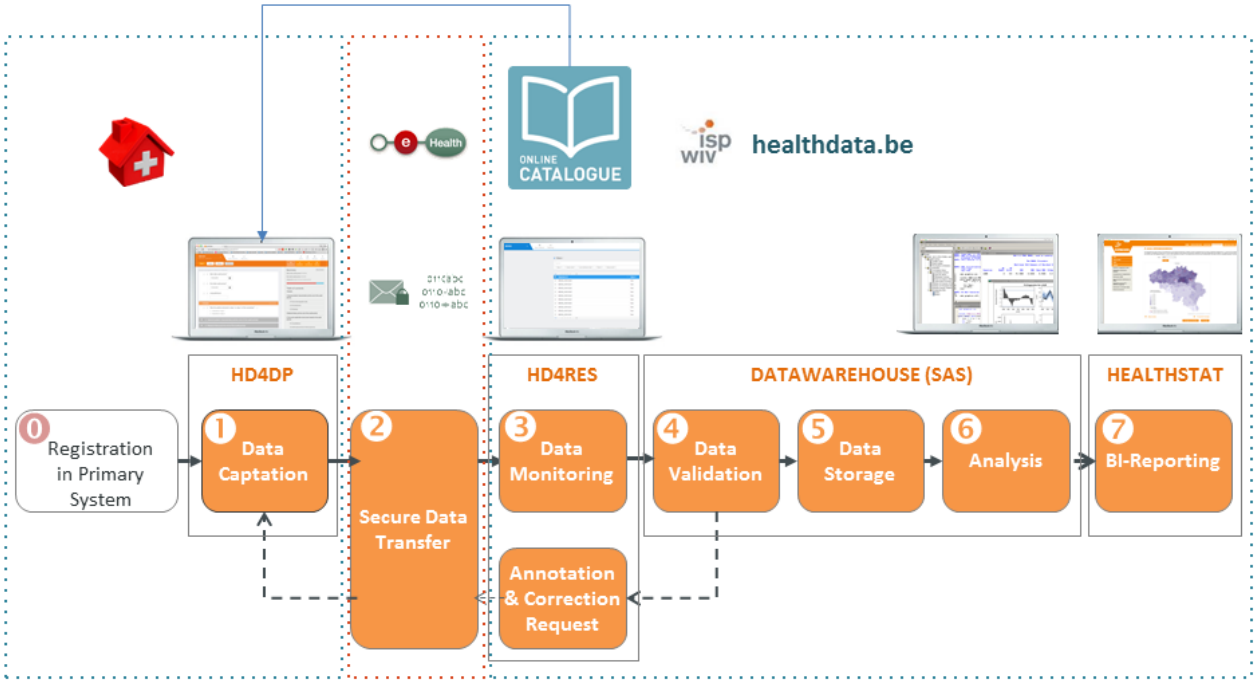
Systems used by data providers are being urged to comply with these DCMs, and these elements are being included in local certification standards. When data providers or researchers are confronted with a concept for which no DCM exists, they can apply for the development of one to enable the automated provisioning of registries.

Data Extraction

When shaping the principles of LHSs, the Institute of Medicine reiterated the need to reflect on the burden data collection can be on health care professionals and the importance of limiting this burden to the issues most important to patient care and knowledge generation [20]. Healthdata.be has developed an electronic data capture (EDC) system called HD4DP (Healthdata for data providers), which enables data extraction at the primary source of clinicians. The logic defined in the DCMs is used in an application programming interface that serves as an interface between data providers' primary source systems and the EDC. Supported input messages, based on health exchange standards, are mapped to the DCMs, upon which structured and coded information from primary source systems can be automatically prefilled in the EDC; this reflects the principle that health data should be recorded only once, and is expected to reduce the administrative burden for health professionals markedly. When data are entered through this data collection form (which can be Web-based but also installed locally), a comma separated value file is generated that can then be transferred on a patient-by-patient basis. Furthermore, data can be transferred in batch if the data provider's system can generate this dataset with the variables automatically prefilled; this method for data capture is illustrated in Figure 2 in the box titled HD4DP.

XSL•FO
RenderX

Figure 2. Data capture, data transfer, data encryption, and data reception through the Healthdata.be platform. HD4DP: Healthdata for data providers; HD4RES: Healthdata for research. Source: https://healthdata.wiv-isp.be/en/services.



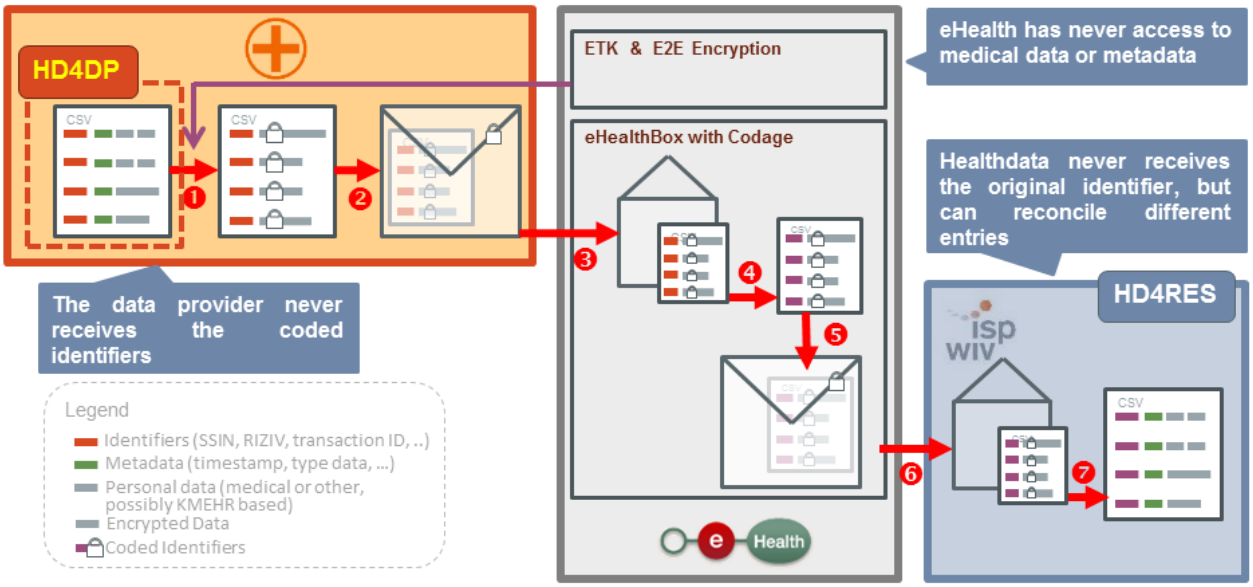
Data Transfer, Safety, and Privacy

The transfer of sensitive health data is challenging with regards to technicality, safety, and privacy. In Belgium, data transfer between health care professionals is organized through existing eHealth channels by an end-to-end (E2E) encryption [21]. For this data transfer, patient and provider identification is ensured through their unique social security identification number (SSIN), and data are encrypted with an algorithm that uses the SSIN of the sender and the receiver. This implies that only the sender or receiver can decrypt the message using their electronic identification card and pin code. The message is sent through an app called the eHealthBox [22], which can only be accessed using the same electronic identification card and pin code. Hence, this eHealthBox functions as a mailbox for encrypted health data. Institutions such as Healthdata.be also have an eHealthBox, which can be only accessed by individuals with the necessary security clearances. The transfer of health data from primary data providers to the Healthdata.be platform uses these channels of encrypted data transfer.

An important feature in enabling the linkage of health data extracted from different systems or settings is the ability to identify data from the same patient. When health data are being sent from one health care provider to another in the context of clinical care, the content of the message is encrypted, but the identity of the patient remains known. However, for research purposes, the content of the message is encrypted, but the identity of the patient must be blinded; this poses an important challenge when health data for a single patient are collected across sources. To enable this linkage without unblinding the coded data, an extra step is introduced in the data transfer. Where masking of identifiers is required, the national eHealth services act as a trusted third party and use an algorithm to pseudonymize this data element [23]. The algorithm used will always code the same patient identifier in the same way, ensuring that data from multiple sources from a single person will always be linkable but still coded. Healthdata.be is, therefore, not responsible for the coding and pseudonymization of sensitive data but uses existing eHealth services that have been technically tested and validated for this purpose. A critical moment in this chain of data transfer is the point where the data are received by the eHealth service and decrypted, before the identifier is coded. To ensure that a breach in this chain does not result in the loss of sensitive data, the whole message, except for the patient identifier, is encrypted the second time using an eHealth token key encryption known only by Healthdata.be [21]. This double encryption (eHealth token key encryption and E2E encryption) ensures safe and blinded data while allowing for the linkage of health data from multiple sources from a single person. All these processes, including the encryption and coding, happen automated and require no human input. Figure 3 illustrates each of the encryption, coding, transfer, and decryption steps. Step 1: data encryption (with exception of the identifiers); Step 2: encryption of whole message for data transfer through eHealth; Step 3: data transfer to trusted third party; Step 4: pseudonymization by trusted third party; Step 5: data encryption for transfer through eHealth; Step 6: data transfer to Healthdata.be eHealth:electronic health; Step 7: decryption of health data

XSL•FO
RenderX

Figure 3. Illustration of the data encryption, coding and decryption steps; SSIN: social security identification number; HD4DP: Healthdata for data providers; HD4RES: Healthdata for researchers; CSV: comma separated value; ETK: eHealth token key; E2E: end-to-end. Source: https://healthdata.wiv-isp.be/en/services.



Healthdata.be can process the data collection for very diverse specialties or research facilities in health care. To ensure that the requested data are in accordance with the research question or aim of the project, a thorough screening of the project is organized. Each project submits a research protocol, including a list of specific data variables being collected. An internal steering committee, an ethics committee associated with a research center, and the National Privacy Commission's Sector Committee for eHealth review this submission. Only when all authorities have approved the project, can the data collection commence.

Data Analysis

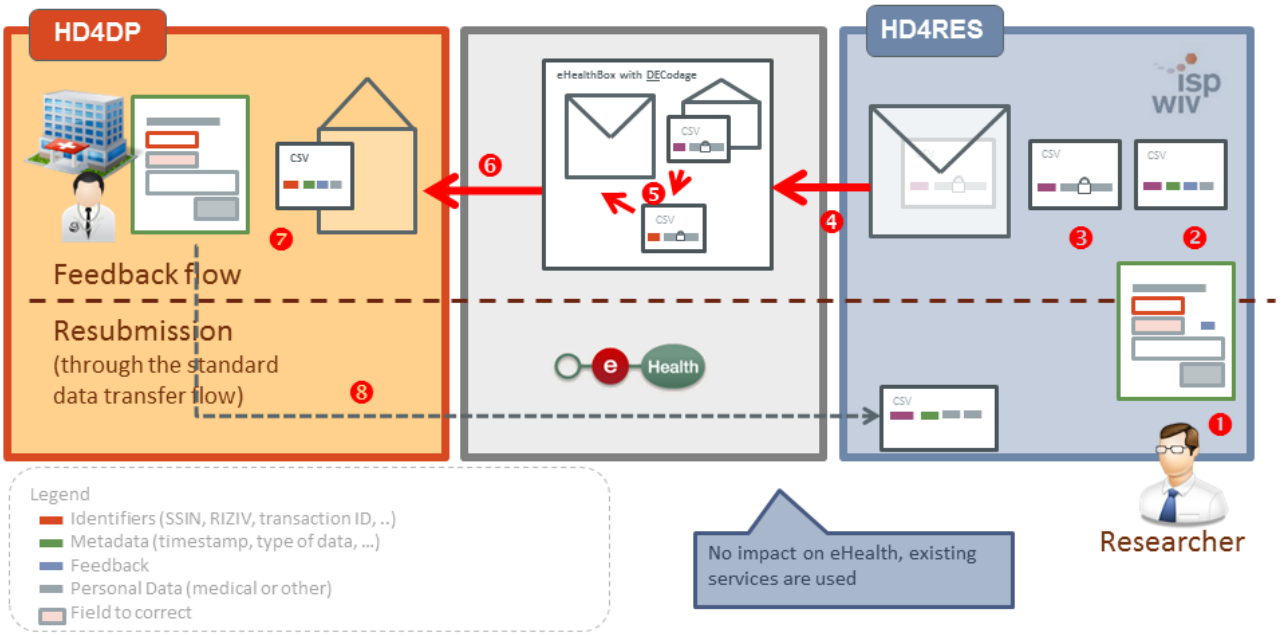
The Healthdata.be platform not only enables safe data transfer but also provides a secure environment for data handling and data analysis for research purposes. Coded data are received by the HD4RES (Healthdata for research) service, which shows the data as sent by the data provider. The interface of the HD4RES is almost identical to that of the HD4DP, except that identification details are coded. Upon arrival in the HD4RES, the data are not yet stored in the datawarehouse (DWH) of Healthdata.be. The DWH has 3 separate entities—the validation environment, the analysis environment, and the reporting environment—and uses SAS Enterprise Guide (SAS Institute Inc) to visualize and process the data. It is first stored in a validation table where data quality is controlled. Healthdata.be allows for semiautomated processes so that the validation of continuous data capture can be operationalized. Once validated, data are then promoted to the analysis environment of Healthdata.be. Access to the HD4RES and the separate environments of the DWH are secured through a 2-factor authentication and can be restricted depending on the needs of the researcher. Furthermore, data processing and reporting can be operationalized to accommodate a continuous data flow in ongoing registers.

Feedback

A pitfall to accepting data from various sources is the possibility of missing or erroneous data. Erroneous data can be prevented by introducing restricted possibilities, ranges, or syntaxes for the data transferred through the HD4DP. For example, validation rules that detect out-of-range data, missing data, or alphanumeric results for a numeric value can already prevent the transfer of these errors at the site of the data provider. However, it may still be possible that an aberrant value is transferred to the HD4RES that needs correction. To allow for this correction by the data provider, a feedback loop has been designed. This feedback loop uses the same channels and encryption methods for data transfer and includes decoding of the SSIN by the trusted third party of eHealth so that the primary data provider can identify the person for whom a corrected data variable is requested. Figure 4 presents this feedback loop.

XSL•FO
RenderX

Figure 4. The illustration of the feedback loop in case of missing or erroneous data using the eHealth services. HD4DP: Healthdata for data providers; CSV: comma separated value; HD4RES: Healthdata for researchers. Source: https://healthdata.wiv-isp.be/en/services.



Results

Case for INTEGO

INTEGO is a primary care morbidity registry, which was founded over 20 years ago [13]. The INTEGO registry has collected data from >100 general practitioners (GPs) and 400,000 patients since its start. Participating GPs all used Medidoc (Corilus, Aalter, Belgium) for documenting their clinical practice and were skilled in structured registration. Eligibility for participation in the INTEGO network required GPs to record on average, at least, one new diagnosis per patient per year, <5% noncoded diagnoses, and these 2 previous requirements should remain stable for, at least, 3 years. In 2017, further information technology development for Medidoc was terminated, and all its users were urged by the vendor to migrate to a new EHR software, called CareConnect (Corilus), a cloud-based EHR. This transition marked the moment to redesign the data collection for this registry, which had not changed since its start. To comply with the new General Data Protection Regulation legislation, the Healthdata.be platform was identified as a partner for this task.

INTEGO does not require any data collection besides that being collected for daily clinical practice. However, the quality of the collected data is expected to be of high quality. The data transferred to INTEGO not only include basic concepts, such as diagnoses or problems, procedures, prescriptions, laboratory tests, parameters, or vital signs and personal information, but also include intricate attributes such as longitudinal care for the same problem (problem-oriented medical registration), causal relationships between diagnoses and prescriptions, or the evolution of a health issue from a symptom into a diagnosis over time. Although many aspects of this registry were already described in existing DCMs, many of these attributes required additional coding and mapping to maintain their meaningfulness. The validity of the recorded data from the original INTEGO database has been studied through comparison with other existing continuous morbidity registries and found to be comparable [13,24]. There is little reason to assume that the data quality would change with the migration to the Healthdata.be platform, but continuous internal validity controls are ongoing.

The INTEGO procedures were approved by the KU Leuven Ethics Committee (nr. ML1723) and by the National Privacy Commission's Sector Committee for eHealth (decision nr. 13.026 of March 19, 2013). The procedures to collect data by Healthdata.be were approved by the Belgian Privacy Commission on April 17, 2018.

To date, almost all GPs have migrated to CareConnect, and the first data export is being prepared and tested. On the one hand, a "core INTEGO" will be constructed, based on the original eligibility criteria to participate in the INTEGO network, to perform epidemiological research. On the other hand, an "extended INTEGO" will be constructed, without eligibility criteria to participate, to perform research on the quality of registration, quality of care, and impact of audit and feedback.

Case for the Electronic Laboratory Medicine Ordering With Evidence-Based Order Sets in Primary Care Trial

The Electronic Laboratory Medicine ordering with evidence-based Order sets in primary care (ELMO) trial is a practical cluster randomized trial investigating the effects of decision support on the quantity and quality of laboratory test ordering behavior by GPs [14]. Data are collected from 3 separate sources on around 11,500 patients. Data on laboratory tests are provided by 3 private laboratories, each with a separate internal coding system for laboratory tests. The Belgian Terminology Center has introduced a national subset of LOINC codes, which are increasingly in use [25]. For the tests being investigated, mappings to LOINC codes were realized before

XSL•FO
RenderX

the start of the trial to ensure interoperability. Data collection includes laboratory tests (type of test, value for test, and units), indications for laboratory test ordering, total cost of the test, identification of the physician, and identification of patients.

In addition, we collected patient-specific data directly from the GPs. GP investigators used several different EHR software for the registration of clinical practice. To ensure uniformity in the data collection, we designed a clinical report form (CRF), detailing the exact information we wished to extract from the EHR and which data would need to be added manually. To facilitate data extraction, we designed the CRF so that >60% of data would be automatically extracted from the EHR, meaning that these data already complied with one or more existing DCMs as defined by Healthdata.be and in use by most EHRs. The CRF was programmed and distributed to all EHRs through an app named Healthdata for Primary Care (HD4PrC), which is a tool that extracts the requested data directly from the EHR and populates the CRF with these data. Only data requests that could not be mapped to a DCM needed to be added manually. Examples of the requested patient-specific data were diagnoses or problems (including international classification for primary care codes and date of diagnosis), procedures performed or ordered, referrals to specialist care, pre- and posttest probabilities of disease, and diagnostic error. These data were then sent to the Healthdata.be platform through the described eHealth channels.

Finally, for a subset of patients, data were obtained directly from patients. A similar CRF was designed, which surveyed patients on data similar to the data requested from investigating GPs. Additional information on the socioeconomic status was requested. To ensure uniformity and avoid technical issues, the CRF was not sent directly to participating patients, but a telephonic interview was conducted by a research assistant who completed the CRF based on patients' responses.

All these data were collected in separate SAS datasets on the Healthdata.be platform, which was accessible through a secured server. Access to various parts of the datasets was dependent on the role of the investigator, where data managers had access to the staging datasets, and statisticians had access to the analytics datasets. The chief investigator had access to all datasets and managed the authorizations of the entire team.

Discussion

Principal Findings

Healthdata.be has successfully connected a myriad of data providers on a centralized platform through a secure and private method of interoperable data transfer across settings and systems. This was done by enabling interoperable data collection, encrypted data transfer, and coded data collection while still allowing to connect data from the same patient collected from multiple sources through a system of pseudonymization. Healthdata.be has largely been able to bridge the disconnect between clinicians and researchers. In addition, Healthdata.be has been able to centralize >100 clinical registries governed by various research facilities, most of which are continuously collecting new data. A list of current registries being hosted on the Healthdata.be platform is available from the website (www.healthdata.be). Alongside the centralization of clinical registries, the platform can also be used in clinical trials or studies using routinely collected data at the point of care. Additional data, which are not defined through a DCM and specific to the trial or study, can be added manually to the data collection tool. These features make Healthdata.be an important facilitator of the LHS and help drive quality improvement in health care.

Facilitating access to reliable health data may be crucial to LHSs, but several situations have illustrated that there may be boundaries to this easier access. The Danish General Practice Database [26] was long considered an outstanding example of a clinical registry but was suspended because of concerns on privacy and security. Similarly, a large data-sharing database linking GP records with hospital data by the British National Health Service was terminated because of the lack of public confidence [27]. It does not appear to be a coincidence that public concerns on privacy coincide with patients' increasing access to their own medical data [28]. In response to these concerns, Healthdata.be requires that all clinical registries or trials that wish to use its platform, receive approval from the National Privacy Commission's Sector Committee for eHealth. This authority scrutinizes each app on security and privacy and determines whether the data collection is appropriate for the project. Projects that have not obtained approval are not operationalized within Healthdata.be.

Limitations

An important limitation to the Healthdata.be platform is rooted in the decentralized data provision. Despite the efforts to standardize data collection using DCMs, variability in data collection at the point of care is inevitable. Even though DCMs may clearly define a clinical concept, there may still be variations in its use in documenting daily clinical practice. To be fully interoperable, DCMs must also be integrated into a conceptual model such as problem-oriented medical registration. These conceptual models not only include interoperable standards for individual concepts but also the relations between concepts. Moreover, even when a concept is well defined and documented within the same conceptual model, interrater differences persist; this is a feature that is common to the way the narrative of a patient is translated into an EHR documentation. Recording guidelines are required, and training on how to put these into practice is imperative, but harmonizing system designs and user interfaces may prove to be crucial. Many of the robust registries, such as the British General Practice Research Database [29] or the Dutch Sentinel General Practice Network [30], import their data from a minimal number of different systems to ensure validity. Many of the registries using the Healthdata.be service are also based on data from a single system, but it is also possible to use authentic sources for several data elements to ensure that the most accurate data are available. For instance, administrative data can be fetched from social security services in real time, ensuring that data, such as the date of birth, are correct. Similarly, like in the ELMO study, it is possible to access multiple systems and collect data where results are most likely to be accurate. In this sense, it seems logical to collect data on laboratory tests from the

XSL•FO
RenderX

laboratory information system rather than from the GP EHR. However, no comparisons of the validity between registries collecting their data from a single source versus multiple sources have been done.

Conclusions

The reuse of health data collected as part of routine clinical care can further research and improve health care. By ensuring semantic interoperability, safe data transfer, and trustworthy data handling, important sources of bias can be avoided. Concerns on data quality and validity can be addressed by collecting data from those sources where the data capture is bound to be most complete and linking these data from multiple sources through pseudonymization. Further research is required to assess whether these methods truly address concerns on the data quality. To date, patients have only limited access and cannot add or change health data in their own patient record. When these features become more widespread, it would be interesting to evaluate how this may influence the data quality and validity.

Acknowledgments

Healtdata.be is a project within the Scientific Institute of Public Health, currently known as Sciensano. Healthdata.be is funded by the National Institute for Health and Disability Insurance. The INTEGO continuous morbidity register is funded by the Flemish Health Ministry. The ELMO study is funded by the Belgian Health Care Knowledge Centre Clinical Trials Programme under research agreement KCE16011.

Conflicts of Interest

None declared.

References

1. Institute of Medicine (IOM). The Learning Healthcare System: Workshop Summary. Washington DC: National Academies Press; 2007.
2. Lauer MS, D'Agostino RB. The randomized registry trial--the next disruptive technology in clinical research? N Engl J Med 2013 Oct 24;369(17):1579-1581. [doi: 10.1056/NEJMp1310102] [Medline: 23991657]
3. Angus DC. Fusing Randomized Trials With Big Data: The Key to Self-learning Health Care Systems? JAMA 2015 Aug 25;314(8):767-768. [doi: 10.1001/jama.2015.7762] [Medline: 26305643]
4. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol 2017 Jan;106(1):1-9 [FREE Full text] [doi: 10.1007/s00392-016-1025-6] [Medline: 27557678]
5. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res 2018 May 29;20(5):e185 [FREE Full text] [doi: 10.2196/jmir.9134] [Medline: 29844010]
6. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, 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]
7. van der Lei J. Use and abuse of computer-stored medical records. Methods Inf Med 1991 Apr;30(2):79-80. [Medline: 1857252]
8. Coppen R, van Veen EB, Groenewegen PP, Hazes JMW, de Jong JD, Kievit J, et al. Will the trilogue on the EU Data Protection Regulation recognise the importance of health research? Eur J Public Health 2015 Oct;25(5):757-758 [FREE Full text] [doi: 10.1093/eurpub/ckv149] [Medline: 26265364]
9. Regulation (EU) 2016/679 of the European Parliamentof the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal dataon the free movement of such data repealing Directive 95/46/EC (General Data Protection Regulation). In: Off J Eur Union. Luxembourg: Publications Office of the European Union; May 04, 2016:1-88.
10. Hersh WR, Cimino J, Payne PRO, Embi P, Logan J, Weiner M, et al. Recommendations for the use of operational electronic health record data in comparative effectiveness research. EGEMS (Wash DC) 2013;1(1):1018 [FREE Full text] [doi: 10.13063/2327-9214.1018] [Medline: 25848563]
11. Institute of Medicine (IOM). Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington DC: National Academies Press; 2013.
12. Celi LA, Davidzon G, Johnson AE, Komorowski M, Marshall DC, Nair SS, et al. Bridging the Health Data Divide. J Med Internet Res 2016 Dec 20;18(12):e325 [FREE Full text] [doi: 10.2196/jmir.6400] [Medline: 27998877]
13. Truyers C, Goderis G, Dewitte H, Akker MV, Buntinx F. The Intego database: background, methods and basic results of a Flemish general practice-based continuous morbidity registration project. BMC Med Inform Decis Mak 2014 Jun 06;14:48 [FREE Full text] [doi: 10.1186/1472-6947-14-48] [Medline: 24906941]
14. Delvaux N, De Sutter A, Van de Velde S, Ramaekers D, Fieuws S, Aertgeerts B. Electronic Laboratory Medicine ordering with evidence-based Order sets in primary care (ELMO study): protocol for a cluster randomised trial. Implement Sci 2017 Dec 06;12(1):147 [FREE Full text] [doi: 10.1186/s13012-017-0685-6] [Medline: 29212546]

XSL•FO
RenderX

15. Institute of Medicine (IOM). Digital Data Improvement Priorities for Continuous Learning in Health and Health Care: Workshop Summary. Washington DC: National Academies Press; 2013.
16. Goossen W, Goossen-Baremans A, van der Zel DZM. Detailed clinical models: a review. Healthc Inform Res 2010 Dec;16(4):201-214 [FREE Full text] [doi: 10.4258/hir.2010.16.4.201] [Medline: 21818440]
17. Moreno-Conde A, Moner D, Cruz WDD, Santos MR, Maldonado JA, Robles M, et al. Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. J Am Med Inform Assoc 2015 Jul;22(4):925-934. [doi: 10.1093/jamia/ocv008] [Medline: 25796595]
18. Health informatics - Detailed clinical models, characteristics and processes. Geneva: ISO; 2015. URL: https://www.iso.org/obp/ui/ [accessed 2018-10-09] [WebCite Cache ID 732dxOClf]
19. Vawdrey DK, 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]
20. Institute of Medicine (IOM). Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in HealthHealth Care: Workshop Series Summary Internet. Grossmann C, Powers B, McGinnis JM. editors. Washington DC: National Academies Press (US); 2011.
21. Cookbook End-to-end Encryption Known recipient Version 2.4. Brussels: eHealth Platform; 2018 Jul 18. URL: https://www.ehealth.fgov.be/ehealthplatform/file/view/b49edc5ce656876fba0b4005a6ecffdb?filename=cookbook_destinataire_connu_new_contact.pdf [accessed 2018-10-09] [WebCite Cache ID 732eImLEG]
22. eHealthbox v3.0 Consultation Web Service Cookbook Version 3.2. Brussels: eHealth Platform; 2018 Apr 30. URL: https://www.ehealth.fgov.be/ehealthplatform/file/view/3a05ba51107dabd7b98e12398c1aad7d?filename=eHBox%20v3%20Consultation%20WS%20-%20Cookbook%20v3.2%20dd%2030042018.pdf [accessed 2018-10-09] [WebCite Cache ID 732eROFgS]
23. Seals Web Service v1.0 Cookbook version 2.1.: eHealth Platform; 2018 Mar 16. URL: https://www.ehealth.fgov.be/ehealthplatform/file/view/242acb087a818d3b2e3e66046015c231?filename=Seals%20WS%20Cookbook%20v2-1%20dd%2016032018.pdf [accessed 2018-10-09] [WebCite Cache ID 732eh5ny4]
24. Fleming D, Elliott C, Pringle M. Electronic Health Indicator Data (eHID): Report from the programme Public Health and Risk Assessment, Health and Consumer Protection.: Public Health and Risk Assessment, Health and Consumer Protection. Strand I: Health Information Priority 2.2.5: eHealth. Directorate General SANCO. European Commission; 2008 Apr 09. URL: http://ec.europa.eu/health/ph_projects/2003/action1/docs/2003_1_19_frep_en.pdf [accessed 2018-10-09] [WebCite Cache ID 732f7nPXB]
25. Logical observation identifiers names and code (LOINC).: ReTaM LOINC URL: https://www.health.belgium.be/en/node/27379#article [accessed 2018-10-09] [WebCite Cache ID 732fGmTwQ]
26. Pedersen KM, Andersen JS, Søndergaard J. General practice and primary health care in Denmark. J Am Board Fam Med 2012 Mar;25 Suppl 1:S34-S38 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110216] [Medline: 22403249]
27. Limb M. Controversial database of medical records is scrapped over security concerns. BMJ 2016 Jul 07;354:i3804. [Medline: 27387682]
28. Mold F, de Lusignan S. Patients' Online Access to Their Primary Care Electronic Health Records and Linked Online Services: Implications for Research and Practice. J Pers Med 2015 Dec 04;5(4):452-469 [FREE Full text] [doi: 10.3390/jpm5040452] [Medline: 26690225]
29. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. Ther Adv Drug Saf 2012 Apr;3(2):89-99 [FREE Full text] [doi: 10.1177/2042098611435911] [Medline: 25083228]
30. Schweikardt C, Verheij RA, Donker GA, Coppieters Y. The historical development of the Dutch Sentinel General Practice Network from a paper-based into a digital primary care monitoring system. J Public Health 2016 Aug 15;24(6):545-562. [doi: 10.1007/s10389-016-0753-4]

Abbreviations

CRF: clinical report form
DCM: detailed clinical model
DWH: datawarehouse
EDC: electronic data capture
eHealth: electronic health
EHR: electronic health record
ELMO: Electronic Laboratory Medicine ordering with evidence-based Order sets in primary care
GP: general practitioner
LHS: learning health system
LOINC: logical observation identifiers names and codes

XSL•FO
RenderX

SSIN: social security identification number

Edited by G Eysenbach; submitted 29.06.18; peer-reviewed by J Cimino, B Kijsanayotin; comments to author 22.08.18; revised version received 03.09.18; accepted 04.09.18; published 19.11.18.

Please cite as:
Delvaux N, Aertgeerts B, van Bussel JCH, Goderis G, Vaes B, Vermandere M
Health Data for Research Through a Nationwide Privacy-Proof System in Belgium: Design and Implementation
JMIR Med Inform 2018;6(4):e11428
URL: http://medinform.jmir.org/2018/4/e11428/
doi:10.2196/11428
PMID:30455164

©Nicolas Delvaux, Bert Aertgeerts, Johan CH van Bussel, Geert Goderis, Bert Vaes, Mieke Vermandere. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 19.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Proposal

Contextual Anonymization for Secondary Use of Big Data in Biomedical Research: Proposal for an Anonymization Matrix

John Rumbold[1*], PhD, MB, ChB, GDL; Barbara Pierscionek[1*], LLM, PhD

School of Science and Technology, Nottingham Trent University, Nottingham, United Kingdom

[*]all authors contributed equally

Corresponding Author:
John Rumbold, PhD, MB, ChB, GDL
School of Science and Technology
Nottingham Trent University
50 Shakespeare Street
Nottingham, NG1 4FP
United Kingdom
Phone: 44 1158483738
Email: John.Rumbold@NTU.ac.uk

Abstract

Background: The current law on anonymization sets the same standard across all situations, which poses a problem for biomedical research.

Objective: We propose a matrix for setting different standards, which is responsive to context and public expectations.

Methods: The law and ethics applicable to anonymization were reviewed in a scoping study. Social science on public attitudes and research on technical methods of anonymization were applied to formulate a matrix.

Results: The matrix adjusts anonymization standards according to the sensitivity of the data and the safety of the place, people, and projects involved.

Conclusions: The matrix offers a tool with context-specific standards for anonymization in data research.

(JMIR Med Inform 2018;6(4):e47) doi:10.2196/medinform.7096

KEYWORDS

anonymization matrix; big data; data protection; privacy; research ethics

Introduction

The Promise of Big Data Research

The era of big data, which is rendered possible by high-power computing and increasingly cheap data storage, offers possibilities for research that have broad and lasting impact. In the last decade, the cost of memory has dropped from dollars per gigabyte to cents per gigabyte [1]. In 2013, the entire amount of data storage globally was 4.4 zettabytes (10^{21}bytes), but in Utah, the National Security Agency facility's storage capacity alone is now estimated at over 1 yottabyte (10^{24}bytes). Traffic on the internet has now surpassed 1 zettabyte per year [2]. The use of data has the potential to transform many fields with health care as a leading prospect [3,4]. Vast amounts of health care data are already gathered, although not always in an electronic form. The widespread adoption of smartphone apps and wearables will vastly increase the amount of wellness and health data produced. Big data and databank research qualitatively differs from most other forms of health care research. Health data already collected for other purposes is often repurposed as a secondary use. This involves considerable cost savings but introduces the problem of lack of participant consent for research. Such issues are particularly acute with health care and other sensitive data. The potential is enormous, but the benefits are not fully exploited because of issues with consent, even though the research involves minimal risk to participants in most cases [5].

Consent, Privacy, and Inconsistent Standards

Minimal risks, however, do not justify a cavalier approach to public consultation or consent requirements, as the failure of United Kingdom (UK) Care.data project demonstrated [6,7]. Failure to consult or inform the public properly resulted in the program being shelved despite having a firm statutory basis to proceed (although the relevant legislation may be incompatible with the General Data Protection Regulation (GDPR; see Multimedia Appendix 1) [8]). Several commentators have stated

XSL•FO
RenderX

that the "consent or anonymize" model does not work for big data [5,9,10]. These issues have led to inconsistent decision making by governance bodies, which have the potential to hinder research in this rapidly progressing area. This paper justifies an anonymization matrix to guide decision making by research ethics review bodies. It draws on relevant norms in the European Union (EU) but will be applicable in other milieux.

Different standards for data research governance in different jurisdictions cause multinational projects certain issues, which have been addressed in the literature. There is also strong anecdotal evidence for inconsistency in approach among research governance bodies within the same jurisdiction. Reasons for such differences need exploration to ascertain whether the consistency and quality of decision making could be improved. Also pertinent is the consideration of public attitudes to inform decision making by research governance bodies.

Methods

Overview

A scoping study was performed using a recognized 5-step methodology to examine the regulation of data science in North America and the EU [11]. These jurisdictions were chosen because of their cultural connections and ease of access to literature because there were no resources to examine or translate publications in languages other than English. The major relevant statutes are GDPR (EU) and the Health Insurance Portability and Accountability Act (HIPAA; USA; Multimedia Appendix 1). GDPR provides derogations for conduct of research without consent, and these are much narrower in scope for health care and other sensitive data. The EU definition of anonymization is based on the possibility of reidentification using techniques that are "likely reasonably" to be used and without using of additional information (which may be easy to obtain). HIPAA provides a framework for anonymization that is far more prescriptive. The research question for the scoping study was as follows: what research on ethics of data anonymization exists to address public expectations of data management by researchers?

Studies were identified using an electronic search using Google Scholar, Westlaw, PubMed, and citation tracking and by manual search. Titles were selected after examination of abstracts. Data were charted from the small number of relevant studies selected with a narrative review generated from these papers. Results were collated and summarized and are presented in the analysis that forms this paper's main text. This analysis and its subsequent conclusions have informed the construction of the proposed anonymization matrix.

Health Care Data Research: What Are the Issues?

Consent

Health data research presents particular ethical issues. Large numbers of participants are exposed to minimal risks with a large number of projects possible using the same resource. Research tissue banks raise similar matters, and there is a considerable crossover with databank research because tissue banks usually have patient data in conjunction with tissue samples; for example, the UK Biobank project has approximately 500,000 participants and each participant, in addition to providing blood, saliva, and urine samples, completes an extensive lifestyle questionnaire [12]. Their imaging study aims to enroll 100,000 participants [13]. These resources are established for future research, the nature of which cannot be predicted [14,15]. The biobank is the curator of the resource rather than the body that performs research. A large number of participants and potential projects would make obtaining specific consent in each case a massive administrative burden and would inevitably reduce the amount of research performed within a specific cost envelope. Given participants' altruism and minimal risks, if appropriate governance mechanisms were in place, that broad forms of consent are permissible is generally accepted [16-21]. These take several forms:

- Simple broad consent with the possibility of withdrawal at a later date: this suffers the disadvantage that the participant may not be kept aware of further projects to be able to exercise the right to withdraw consent
- Categorical consent: this is narrower—consenting to research in particular areas, which would be compliant with GDPR
- Consent to a form of governance: regulation of the resource can be entrusted to a reflexive governance mechanism that participants trust to make surrogate decisions informed by input of both shareholders and stakeholders [14,22]
- Combinations of these options or some other variation [23,24]

Alternatively, dynamic consent may be required. In this situation, participants are provided with information about each research project to decide whether to provide consent [25]. This mandates specific informed consent for each project, but it has been shown that participants can find this process too demanding [22,26-28].

When research is performed using "found data," the issue of obtaining even broad consent is more problematic [29]. These considerations mean that with appropriate approval, participants' informed consent may not be necessary, contrary to established practice in biomedical research [10,30]. There are broad research exemptions for data science, but derogations to permit research using sensitive data are narrower. GDPR states that research on sensitive data must be "in the public interest" (Recital 53).

There is the potential for several bodies to be involved with decision making on consent to use health care data for research—research ethics committees or institutional review boards, data access committees, data protection authorities, and health service management boards. Some of these bodies have local, national, and supranational arms, each of which may have a different perspective and make different decisions based on the same facts. There are anecdotal reports of divergent opinions on consent to data use between research ethics committees and the Confidentiality Advisory Group of the UK's Health Research Authority (Personal communication from John Fistein). Although the Confidentiality Advisory Group's main remit is to advise the Health Research Authority and Secretary of Health pursuant to the Health Service (Control of Patient Information)

XSL•FO
RenderX

Regulations 2002 and s251 of the Health and Social Care Act 2012, its assessments include an implicit ethical evaluation of whether confidential patient information can be processed. Similar inconsistencies and tendencies toward risk aversion have been described in relation to administrative data [5].

Potential harms that participants in data research might be exposed were examined in a scoping study conducted at the Mason Institute and Farr CIPHER for the Nuffield Council on Bioethics Working Party on Biological and Health Data and by the Wellcome Trust's Expert Advisory Group on Data Access [31]. Limited owing to time and resource constraints, their study focused on the instances of financial damage and emotional distress to individuals. There may be substantial harm to organizations as well, including reputational damage and loss of trust. Many incidents they identified were related to maladministration, and this reinforces the need for secure systems for data science.

Difficulties with consent illustrate that merely gaining consent is not a panacea for all data research issues even when practicable. The standard paradigm for data research is to "consent or anonymize." Therefore, if consent is not practicable for big data projects, the researcher might choose to anonymize data. This is not necessarily straightforward and introduces a host of other issues.

Anonymization

Anonymization is a means of preventing a breach of confidentiality and preserving privacy. Anonymized data are not protected under data protection law. Confidentiality and privacy are related concepts: confidentiality is a *duty* owed, often by a professional, to an individual in particular circumstances; privacy is a *right* that a person enjoys. An individual divulges many sensitive facts to professionals, particularly in law and medicine [32], with the understanding that the professional has a professional, legal, and ethical duty to maintain the information and data in confidence or face hefty sanctions for breaching these duties [33-35]. Duty of confidentiality does not apply where data have been anonymized. A duty of confidentiality is included in the Hippocratic Oath [36] and the Geneva Declaration [37], but there is an additional duty in the Geneva Declaration which is to: "share my medical knowledge for the benefit of the patient and the advancement of healthcare."

This injunction could be interpreted as placing a duty on physicians to share data for purposes of medical research when conducted for the common good. In the UK, the Information Commissioner's Office (ICO) and the Caldicott Review have commented on the problem of not using data [38,39]. Caldicott made this an added seventh principle: "The duty to share information can be as important as the duty to protect patient confidentiality." Although this added principle is in the context of particular duties to an individual, rather than research *per se*, it could be interpreted to include a duty to use data to improve health care.

The distinction between privacy and confidentiality is acknowledged in data protection law in which particular protections apply to those "who in the circumstances owe a duty of confidentiality which is equivalent to that which would arise if that person were a health professional" (UK Data Protection Act 1998 (Multimedia Appendix 1); similar provisions apply in other transpositions of the Data Protection Directive). Data safe havens require researchers' contractual duty to maintain confidentiality and not to attempt reidentification [40,41]. Hefty sanctions should be applied only to those intentionally breaching guidelines; otherwise, a tendency will arise to restrict data sharing unnecessarily [31]. This is one factor behind the tendency of not sharing data when doing so is both legally and ethically acceptable [5].

Anonymization is the procedure that removes data from the remit of data protection law, which pertains only to personal data. Data about or relating to a person are no longer personal if these cannot be linked to the person. Anonymization requires more than just removal of identifiers; the combination of 3 pieces of data could identify 87% of US residents—5-digit zip code, birth date, and sex (note that this would not satisfy the HIPAA Privacy Rules anonymization criteria for 2 of the 3 fields, see Multimedia Appendix 2) [42].

GDPR defines personal data as relating to "an identified or identifiable natural person ('data subject')," as included in Article 4.1. The definition of "an identifiable natural person" covers identification by direct or indirect means and can range from a name to social identity. The nature of personal data is not further defined. Although some facts about an individual are trivial, nonetheless, defining content of personal data that would cover all individuals in all situations and be universally acceptable is difficult. The UK Anonymisation Network (UKAN), run by a consortium of the University of Manchester, University of Southampton, the Open Data Institute, and the Office for National Statistics to establish best practice for anonymization, has classified data, as shown in Table 1.

Defining personal data purely by content is problematic, perhaps because some data tangentially refers to a person, for example, a vehicle registration plate (which would be secondary personal data under the UKAN schema), or because whether the data identifies someone depends on many other factors. This issue is illustrated vividly by the decision in *Breyer v Germany* on whether a dynamic internet provider address is personal data (Multimedia Appendix 1). UKAN states that anonymization depends not only on data but also on the environment within which that data are found [43].

Table 1. Four types of data depending on whether they are about people and whether they are identifiable [43]. Source: Anonymisation Decision-Making Framework.

About individuals	Nonidentifiable data	Identifiable data
Yes	Anonymized data	Primary personal data
No	Apersonal data	Secondary personal data

XSL•FO
RenderX

UKAN has divided anonymization into the following 4 types: formal, guaranteed, statistical, and functional [43]. First, formal anonymization means that direct identifiers have been removed. This does not satisfy the EU legal standard. Second, guaranteed or absolute anonymization, as the name suggests, provides security against any possible reidentification but often at the expense of the removal of large amounts of data. Data protection law does not require this but individual data controllers may deem it necessary. Third, statistical anonymization is tied to the concept of statistical disclosure control. It seeks to reduce the chance of reidentification to below a given predetermined statistical threshold. This threshold is crucial to whether anonymization provides real protection; for example, with differential privacy, the epsilon value selected by Apple has been severely criticized for providing little protection of privacy [44]. Finally, functional anonymization examines the risk of anonymization within a particular context, taking into account motivations of an attacker, consequences of disclosure, and data divergence among other criteria. Data protection legislation does not consider these factors in legal standards for anonymization.

Difficulties With Anonymization

In light of difficulties in defining and ensuring anonymity, definitions of personal data across the globe are becoming broader [45-48]. Proliferation of data collected by various data controllers about data subjects and evolution of techniques to reidentify data subjects has required reassessment of anonymization. Now, unless data are substantially masked, swapped, grouped, or deleted, an individual can often be identified by someone sufficiently determined, with the right technical skills, and with sufficient additional data [42,49]. Although methods have been developed to achieve tailored levels of processing to maximize data's utility, while ensuring anonymization to a given level, none of these alter the fact that making good use of data and achieving rigorous anonymization are currently incompatible. Statutes do not mandate guaranteed anonymization. Instead, they set a standard of difficulty of reidentification that must be achieved. This depends on factors such as motivations and skills of an intruder and information that might be combined with data. None of the legal standards appear to vary according to data's circumstances or sensitivity, although these factors feed into an assessment of good practice [50,51] and could be incorporated into codes of conduct that would be part of the regulatory milieu encouraged by GDPR Article 40 (Comments made at the Privacy Engineering Research and the GDPR Workshop, KU Leuven November 10, 2017 as observed by JR).

Sensitive personal data are defined in data protection legislation, and health care data are one of those categories (s2, Data Protection Act 1998, UK). There are additional ethical and legal protections for health care data, which may include specific protections for particular categories, for example, sexually transmitted infections (eg, National Health Service, Venereal Diseases, Regulations SI 1974/29-UK, now repealed) and genetic data (eg, SI #687/2007—Data Protection Processing of Genetic Data Regulations 2007, Ireland; outside the EU, there is the Federal Act on Human Genetic Analysis 2004, Switzerland). It has been demonstrated that public conception of sensitive data categories may vary from that defined in legislation [52,53].

Anonymization introduces several problems for data researchers, particularly in health care research. It reduces the quantity and quality of usable data to a variable degree. Anonymization makes it impossible to verify data or act on any results that might have consequences for participants, for example, when imaging studies are performed. It will prevent linking of records, either to form a longitudinal health record or to link datasets on relevant issues such as socioeconomic indicators. Pseudonymization makes several of these objectives possible; however, in GDPR, pseudonymization is specifically excluded from being categorized as anonymization.

Public Attitudes Toward Data Science

The basis of an ethical waiver for consent largely rests on the presumption that the public would consent to having their data used in this way, given the potential for creating public good. This necessitates an assessment of public attitudes. Different projects and datasets may require different approaches for preserving participants' privacy, while maximizing the benefit of research performed. Another consideration is the public's attitude toward data research, in particular, factors that affect the public's expectation of how their data will be processed. This is especially important because the social license on which data research with consent or anonymization relies rests on public support.

The public's attitudes toward use of data for research have been studied by the Scottish Health Informatics Programme (SHIP) and the Ipsos MORI Social Research Institute on behalf of several UK organizations [23,54-58]. Use of deliberative engagement methods has proven crucial because public attitudes to data sharing are affected significantly by provision of sufficient information on how data are used. During their deliberative engagement exercise, SHIP found that initially, members of the public expected consent to be asked for each research project. However, with greater understanding of the number of potential research projects with similarity of issues, they considered broad consent to be as acceptable, if not preferable. A similar result was found in a study of the US public [26]. In recent years, the Ipsos MORI Social Research Institute has conducted studies on behalf of the Medical Research Council (the use of personal health information in medical research, 2007) [23]; the Economic and Social Research Council (Dialogue on Data) [58]; the Wellcome Trust (Commercial access to health data, 2016) [56]; the Royal Statistical Society (Public attitudes to the use and sharing of their data, 2014 [55]); and the Government Data Science Partnership (Public dialogue on the ethics of data science, 2016) [57]. Similar to SHIP, it found that attitudes to data sharing varied considerably depending on the purposes and likelihood of public benefit.

Nissenbaum coined the term "contextual integrity" to denote the binding of data sharing practices to particular contexts [59]: The mere fact that data are in the public domain does not constitute license to disseminate them more widely. Solove also dealt with this issue in his taxonomy of privacy [60]. As the example of *Nader v General Motors Corp* demonstrates, intrusive monitoring of activities performed in public can be an

XSL•FO
RenderX

invasion of privacy; just because Nader was in a bank did not permit anyone to know how much money he was withdrawing (nor, indeed, that he was there to withdraw money at all; Multimedia Appendix 1). Therefore, posting material on social media does not automatically make their use for research ethical. Anonymization may still be necessary and appropriate for Facebook and Twitter posts because posters had not intended their material to be disseminated to a wider audience.

With research on attitudes toward sharing location data in particular, Sadeh has also found that privacy is highly contextual [62-64]. Willingness to share location data depends on several factors including time of day, day of the week, social group requesting data, and location. Sadeh found that the purpose for which data would be used was particularly important in decision making. If location data are crucial to the central purpose, its use is much more frequently acceptable than when it is tangential or unrelated to the app's central purpose. Similarly, an individual who may be willing to share sensitive data, such as in health care, might be unwilling to have socioeconomic data linked with those medical records [65]. This points to a demand for improved, granular consent requirements to reflect the need for data from individuals.

Results

This proposal arises from an ethico-legal analysis completed during our work on the Aegle project. It takes into consideration recent EU legislation, but the resulting matrix is applicable to most jurisdictions.

Discussion

A Framework for Information Governance: A Proposed Solution

Governance is an inclusive form of regulation that encompasses governmental laws and regulations. Information governance frameworks require synthesis of data protection laws, guidance from national data protection officers, and an appreciation of expectations of the public they serve. Governance mechanisms can and ought to be more flexible and responsive than governmental laws and regulations. The main justification for the proposed matrix is all the evidence that supports the common sense notion that people are willing to share different amounts and types of data with different people, in different settings, for different purposes, at different times. Therefore, it is reasonable to conclude that using the same anonymization standard for all data protection and freedom of information purposes does not reflect societal attitudes to data or provide a governance framework that satisfies individuals' reasonable expectations. A proportionate form of governance is preferable to protect individuals as fully as practicable while maintaining the capacity to generate useful insights from data [66,67]. The demands of good governance are usually greater than the legal standard, and this is particularly true for research in which standards are left for the scientific community to decide (eg, GDPR allows consent "in keeping with recognized ethical standards for scientific research," Recital 33) [8]. However, there is no suggestion that this assessment can be done without public engagement.

The trust placed in medical practitioners and academic researchers therefore entails the public's possible acceptance of a lower standard of anonymization, given that data users have a professional or contractual duty to respect confidentiality [35,36]. There is a persuasive case for having different standards of anonymization for medical research conducted in safe havens or at least by researchers under a duty (whether professional or contractual) of confidentiality, including a duty of not attempting reidentification, and for data released to the public whether under a freedom of information request or not. The UK Data Protection Act allows processing of medical data under Schedule 3, Para 8.1b by "a person who in the circumstances owes a duty of confidentiality which is equivalent to that which would arise if that person were a health professional." The trustworthiness and motives of those who examine data are highly relevant to which precautions would be prudent. There is no control over what techniques can be used and by whom once data are released to the public and are therefore "in the wild."

Data protection authorities have commented on the dynamic nature of personal data. The UK ICO (2012) noted that predictions about data available now or later cannot be made with certainty (page18) [39]. The EU's Article 29 Working Party reached a similar conclusion, recognizing that the changing nature of data and its usage as well as growth in information that could aid identification can give rise to new data protection issues. This requires an appreciation that anonymization is not a one-off exercise [68]. Data that at one point in time is anonymized may subsequently be identifiable and thus become personal data once more. Based on these considerations and the fact that once data has been released to the public, we conclude that it cannot be recalled and operations performed cannot be limited in any way; there is justification for applying the most stringent standards of anonymization to data for public release.

This distinction was not considered by the UK ICO in their decision FS50565190 against Queen Mary University of London, relating to a trial of treatment for chronic fatigue syndrome. Their decision held that the University could not withhold data anonymized to the Information Commissioner's satisfaction, despite concerns about activists trying to reidentify participants. The ICO wanted a specific explanation as to how reidentification would be achieved [69]. Section 22A of the Freedom of Information Act now exempts research studies from disclosure prior to publication of papers, but this only extends the timeframe for disclosure rather than absolutely exempting them. The University argued that participants had an expectation that their data would be confidential and that, in a small community, addition of information, for example, about hospital appointments, might enable reidentification. Participants had already withdrawn consent because of such fears, and this required expensive, time-consuming reanalysis of the remaining data.

In summary, we argue that the evidence demonstrates that neither consent nor anonymization to current legal standards is a solution to all data research issues. Limitations of anonymization make the application of the same standard across

XSL•FO
RenderX

the board problematic. Recognition of the current framework's inadequacy has led us to propose an anonymization matrix for treatment of sensitive data, particularly health care data. Our hypothesis is that the matrix will improve proportionate information governance and can therefore improve the trustworthiness and utility of data research. This hypothesis requires testing with empirical research, which is beyond the remit of this paper.

Proposal

An Anonymization Matrix

A tool for research ethics committees, institutional review boards, and data access committees for assessing data protection aspects of a project and achieve consistent proportionate information governance is proposed. This P-R matrix (Table 2) includes a number of levels of anonymization adjusted according to the best evidence about public attitudes to trustworthiness, particularly recent research on public attitudes about data's use for research. The matrix also takes into account the unpredictability of health care data's future identifiability, holding that any data for public release should be subject to the highest standards of anonymization in accordance with the precautionary principle. GDPR and ethical standards demand that when research is not in the public interest, the standard paradigm of "consent or anonymize" should apply.

Levels 1-3 of anonymization referenced in the table are defined in Multimedia Appendix 3. They are developments of the HIPAA Privacy Rule (detailed in Multimedia Appendix 2) with the addition of algorithm-based methods that can adjust processing according to required levels of k-anonymity, l-diversity, t-closeness, and differential privacy. Definitions of particular categories and rationales are provided below. We have also incorporated a UK adaptation for the obscuration of postcodes in Multimedia Appendix 4.

Considering that particular contexts may make an individual more vulnerable or the attempts to breach more skilled or more determined is also appropriate. These special circumstances include data on celebrities or other persons about whom a considerable amount of information is already in the public domain and is widely known. Use of metrics to determine the possibility of reidentification is appropriate, although the Level 3 standard combined with anonymization algorithms to provide k-anonymization may not be sufficient to allow for public release. Synthetic data, or a dataset that has been subtly altered from the original, is a good substitute in many situations because it can be demonstrated to provide results very similar to those obtained from data on which it is based [70,71]. The content of the columns and rows in Table 2 is explained further.

Table 2. P-R anonymization matrix.

Context of data	Data use authorized without consent[a]	Health care data use without consent	Very sensitive health care data[b] use without consent	Special circumstances without consent
Research in safe havens[c]	Anonymization not required	Level 1	Level 1	Level 2
Research to which duty of confidentiality applies	Anonymization not required	Level 1	Level 2	Level 3
Research to which no duty of confidentiality applies[d]	Level 1 + algorithmic manipulation[e]	Level 1 + algorithmic manipulation	Level 2 + algorithmic manipulation	Level 3
Information for public release[f]	Level 3 or synthetic data or no release	Level 3 or synthetic data	Level 3 or synthetic data	Level 3 or synthetic data or no release

[a]Where authorization for data processing without consent has been provided by a specific statutory body, a body that provides appropriate safeguards, or the equivalent for research ethics. These bodies have powers to authorize data use without anonymization; however, good practice requires data minimization with justification for inclusion of all identifying data.

[b]Very sensitive data are not exhaustively defined in this paper because they depend heavily on particular sociocultural sensitivities; for example, alcoholic liver disease would be a sensitive diagnosis in some cultures but not necessarily in all. Sexually transmitted infections are usually considered very sensitive. Public consultation is needed on use of health care data in an ongoing process.

[c]Requirements for accreditation include that researchers are under contractual duties of confidentiality, including not to attempt reidentification [40].

[d]It should be noted that the UK government has signaled an intention to create a new criminal offense of reidentification [72]; other jurisdictions, including New Zealand, Australia, and Canada, are also considering this [73,74]. Currently, reidentification would be merely a breach of data protection law.

[e]Algorithmic manipulation means data masking, clustering, or deletion to satisfy demands of k-anonymity and other metrics such as l-diversity, t-closeness, or differential privacy.

[f]As noted above, the UK Information Commissioner's Office could compel release under the Freedom of Information Act 2000 of data only anonymized to their standard (currently, the motivated intruder). This standard is arguably deficient for public release of health data [61], and we propose statutory change to enable an appropriate level of privacy protection to be required.

Rationale for the Anonymization Matrix

Authorization means that data use has been permitted without consent by a statutory body, research ethics committee, or other empowered governance body. "Duty of confidentiality" in this instance means a professional or contractual duty of confidentiality equivalent to those of health care professionals, additional to a duty of not attempting reidentification.

XSL•FO
RenderX

Research in Safe Havens

Several requirements must be met for a data safe haven to be accredited.

1) Authorization by the appropriate body covers the use of data for research without any anonymization, whether in a data safe haven, when the researcher is bound by a duty of confidentiality in the same way as medical professionals.

2) and 3) Anonymization is required to meet legal requirements where authorization has not been granted. Because research is being conducted in safe havens, there is no requirement for a standard higher than Level 1.

4) Where there are special circumstances, it seems reasonable to expect a higher standard of anonymization because reidentification could occur spontaneously without any deliberate attempt by the researcher.

Research Where Duty of Confidentiality Applies

Duty of confidentiality provides protections for participants but not other safeguards provided in accredited data safe havens. Hence, some additional anonymization may be necessary.

1. Where authorization is granted, there is no need for anonymization.
2. Where no authorization has been granted, Level 1 anonymization will satisfy legal and ethical requirements.
3. Where data are particularly sensitive, risks related to disclosure are correspondingly higher. Therefore, we recommend Level 2 anonymization.
4. Where there are special circumstances, Level 3 anonymization reduces risk of inadvertent reidentification.

Research in Which No Duty of Confidentiality Applies

If researchers are not under a duty of confidentiality, safeguards to prevent reidentification should be stronger. Excessive processing of data can be reduced by resorting to algorithmic manipulation.

1. Although there is no legal requirement when authorization has been granted for any anonymization to be performed, we argue that where there is no duty of confidentiality, ethical bodies should require it. Additionally, algorithmic manipulation should be required to ensure that reidentification cannot occur.
2. Here anonymization is legally required. The safeguard of additional algorithmic manipulation should be required by ethical bodies.
3. Processing of more sensitive health care data warrants the higher level of anonymization with algorithmic manipulation.
4. Research on data in special circumstances where researchers are not bound by a duty of confidentiality is worthy of the highest levels of anonymization (where such research is permitted at all).

Public Release

When information is released to the public, anonymization must be as rigorous as possible, owing to future development of new techniques for reidentification and possible release of further information. For this reason, we recommend that at least Level 3 anonymization be used. Synthetic data are preferable when substitution is feasible.

1) and 4) Without consent or with special circumstances, there is a case for not releasing any data to the public. Synthetic data pose no privacy risk.

2) and 3) Release of rigorously anonymized data are acceptable although synthetic data are preferable.

These requirements would need periodic review. Because data for public release cannot be modified to increase protection from reidentification, standards for anonymization must be robust enough to provide protection for at least the medium term. The proposed matrix provides guidance for research ethics review bodies to harmonize their ethical assessments with data protection requirements, while providing the enhanced protection expected for sensitive data.

Techniques of Anonymization

It has been said many times that “ought” implies “can” [75]. Anonymization is not an all or nothing process, but rather a spectrum of processing that provides greater or lesser degrees of difficulty in reidentifying an individual. Although finite risk is associated with nearly all data science research, the public is willing to accept this if mitigated by appropriate data security and safeguards [76,55]. A further solution to the problem of reidentification is to restrict access to researchers who have given assurances that they will not attempt reidentification. Although potentially attractive and reassuring to the public, this currently makes no difference to whether data are classified as indirectly identifiable. However, such assurances would be good evidence of the provision of appropriate safeguards by the body concerned.

Techniques involved in anonymization reduce utility of data to a greater or lesser extent [51,77,78]. Academic literature has much debated risks associated with anonymized data. Although researchers have demonstrated that datasets can be reidentified in defined circumstances, whether these scenarios reflect what is likely in the real world is contentious [42,49,79].

The 2 approaches to anonymization are rule based and risk based. Rule-based anonymization is typified by the first part of the HIPAA Privacy Rule, which mandates obscuring particular data fields. The HIPAA Privacy Rule is easy to apply, but there are problems with it. In some circumstances, it involves unnecessary deletion of data; in others, it fails to provide adequate protection from reidentification [80]. Ruling out unusual data that can uniquely identify an individual is difficult, an example being the mayor of Ottawa [81]. There is also the issue of where sufficient data are available in the public domain about an individual for reidentification to be feasible [61].

Risk-based anonymization involves processing guided by the calculation of the risk of disclosure and assessing which data need to be obscured to achieve this [50]. It would include the statistical expertise-based standard in the HIPAA Privacy Rule and involves such techniques as k-anonymization, l-diversity, t-closeness, and differential privacy [82-85]. These techniques examine data for patterns that would enable reidentification (if, eg, one dataset has particular attributes) and selective data

XSL•FO
RenderX

masking, clustering, or deletion to reduce the possibility of drawing inferences from deidentified data. These techniques reduce deterioration in data, but they do not eliminate it. If too large a value for epsilon is selected for differential privacy, then privacy protection will be minimal [44]. An overstringent rule-based approach to anonymization is problematic, and a proportionate form of governance has distinct advantages [15,86]. If researchers agree to not attempt reidentifying participants and their duty is reinforced by the prospect of sanctions, it provides reassurance and facilitates preservation of intact data. Conversely, data for public release may be subject to any number of techniques and addition of data from a variety of sources, both legal and illegal [61,87].

Advances that have enabled reidentification also enable other inferences from existing data. This is, after all, the basis for linkage research. One of the most famous examples is the Target customer being sent offers on baby-related items when she had not yet told her father of her pregnancy. The supermarket had inferred the fact of her pregnancy from her purchasing habits [88]. The participant cannot have given permission for the production or storage of these new facts when consent has not been specifically given for research purposes. Recently, ICO fined charities for conducting "wealth screening" of potential donors [89].

Conclusions

The literature on privacy and attitudes toward use of data for research purposes provides support for application of different standards of anonymization depending on circumstances. Additionally, the regulatory burden can be reduced by harmonization of criteria applied by research ethics committees and other governance bodies. For research ethics and data access committees, our anonymization matrix provides guidance that exceeds the requirements of current data protection laws. Each row and column of the matrix corresponds to a meaningful ethico-legal distinction. It offers contextual guidance for research ethics bodies to recommend appropriate levels of anonymization when gaining specific consent is not feasible.

We propose that research ethics bodies should not deny permission on grounds of privacy or consent issues for projects that satisfy these anonymization requirements. Satisfying these requirements should make approval, for example, by the Confidentiality Advisory Group easier. Additionally, compliance with standards that exceed legal requirements help secure the social license and thus ensure data bank projects' legitimacy and longevity.

The major potential advantage of such a matrix is the facilitation of international projects. Any ethico-legal framework that satisfies the requirements of multiple jurisdictions without imposing excessive regulatory burden will be a valuable tool for such projects. To demonstrate the matrix's value for improving research ethics committees' decision making on information governance, we propose its use in EU data science projects on a trial basis.

Acknowledgments

This paper is based on work funded by Horizon2020 grant ICT 2014/1 (Aegle).

Authors' Contributions

Both authors contributed to the writing, design, and analysis within this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Statutes and cases.

[PDF File (Adobe PDF File), 189KB - medinform_v6i4e47_app1.pdf]

Multimedia Appendix 2

The HIPAA (Health Insurance Portability and Accountability Act) privacy rule.

[PDF File (Adobe PDF File), 40KB - medinform_v6i4e47_app2.pdf]

Multimedia Appendix 3

Anonymization standards proposed levels 1-3.

[PDF File (Adobe PDF File), 53KB - medinform_v6i4e47_app3.pdf]

Multimedia Appendix 4

UK modification of geographic rule.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 27KB - medinform_v6i4e47_app4.pdf]

References

1. Komorowski M. A history of storage cost (update). Mkomo 2014 Jul 19 [FREE Full text]
2. Cisco. 2016 Jun 01. The Zetabyte era: Trends and Analysis URL: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html#_Toc484556828 [accessed 2018-09-05] [WebCite Cache ID 72Ce6DtCh]
3. Murdoch T, Detsky A. The inevitable application of big data to health care. JAMA 2013 Apr 03;309(13):1351-1352. [doi: 10.1001/jama.2013.393] [Medline: 23549579]
4. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
5. Stevens LA, Laurie G. The Administrative Data Research Centre Scotland: A Scoping Report on the Legal & Ethical Issues Arising from Access & Linkage of Administrative Data. SSRN Journal 2014 Aug 28. [doi: 10.2139/ssrn.2487971]
6. National Health Executive. 2016 Jul 07. NHS England to close care.data programme following Caldicott Review URL: https://tinyurl.com/yb99b4ye [accessed 2018-09-05] [WebCite Cache ID 72Cf3K9CE]
7. Carter P, Laurie G, Dixon-Woods M. The social licence for research: why care.data ran into trouble. Journal of Medical Ethics 2015;41(5):409. [doi: 10.1136/medethics-2014-102374]
8. Rumbold JMM, Pierscionek B. The Effect of the General Data Protection Regulation on Medical Research. J Med Internet Res 2017 Feb 24;19(2):e47 [FREE Full text] [doi: 10.2196/jmir.7108] [Medline: 28235748]
9. Mostert M, Bredenoord A, Biesaart M, van Delden JJM. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. European Journal of Human Genetics 2016 Dec;24(7):956. [doi: 10.1038/ejhg.2015.239]
10. Mittelstadt B, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. In: Mittelstadt B, Floridi L, editors. The Ethics of Biomedical Big Data. Oxford: Springer; 2016.
11. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
12. UK Biobank. 2017. About UK Biobank | UK Biobank Internet URL: http://www.ukbiobank.ac.uk/about-biobank-uk/ [accessed 2018-09-05] [WebCite Cache ID 72CfFvzBu]
13. UK Biobank. 2016. Scanning study launched URL: http://www.ukbiobank.ac.uk/2016/11/scanning-study-launched/ [accessed 2018-09-05] [WebCite Cache ID 72CgBSM4O]
14. Laurie G. Governing the Spaces In-Between: Law Legitimacy in New Health Technologies. In: Flear ML, Farrell AM, Hervey TK, Murphy T , editors. European Law and New Health Technologies. Oxford: Oxford University Press; 2013:193.
15. Laurie G. Reflexive governance in biobanking: on the value of policy led approaches and the need to recognise the limits of law. Human Genetics 2011 Sep;130(3):347-356. [doi: 10.1007/s00439-011-1066-x] [Medline: 21766192]
16. Richter G, Krawczak M, Lieb W, Wolff L, Schreiber S, Buyx A. Broad consent for health care-embedded biobanking: understanding and reasons to donate in a large patient sample. Genet Med 2018 Dec;20(1):76-82. [doi: 10.1038/gim.2017.82] [Medline: 28640237]
17. Hansson M, Dillner J, Bartram C, Carlson J, Helgesson G. Should donors be allowed to give broad consent to future biobank research? The Lancet Oncology 2006 Mar;7(3):266-269. [doi: 10.1016/S1470-2045(06)70618-0]
18. Sheehan M. Can Broad Consent be Informed Consent? Public Health Ethics 2011 Nov;4(3):226-235 [FREE Full text] [doi: 10.1093/phe/phr020] [Medline: 22102849]
19. Hofmann B. Broadening consent--and diluting ethics? J Med Ethics 2009 Feb;35(2):125-129. [doi: 10.1136/jme.2008.024851] [Medline: 19181887]
20. Lipworth W, Morrell B, Irvine R, Kerridge I. An empirical reappraisal of public trust in biobanking research: rethinking restrictive consent requirements. J Law Med 2009 Aug;17(1):119-132. [Medline: 19771992]
21. Erlich Y, Williams J, Glazer D, Yocum K, Farahany N, Olson M, et al. Redefining genomic privacy: trust and empowerment. PLoS Biol 2014 Nov;12(11):e1001983 [FREE Full text] [doi: 10.1371/journal.pbio.1001983] [Medline: 25369215]
22. Hoeyer K, Olofsson B, Mjörndal T, Lynöe N. Informed consent and biobanks: a population-based study of attitudes towards tissue donation for genetic research. Scand J Public Health 2004;32(3):224-229. [doi: 10.1080/14034940310019506] [Medline: 15204184]
23. Ipsos MORI. The Use of Personal Health Information in Medical Research: General Public Consultation. London: Medical Research Council; 2007.
24. Steinsbekk KS, Kåre Myskja B, Solberg B. Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? Eur J Hum Genet 2013 Sep;21(9):897-902 [FREE Full text] [doi: 10.1038/ejhg.2012.282] [Medline: 23299918]
25. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2015 Feb;23(2):141-146 [FREE Full text] [doi: 10.1038/ejhg.2014.71] [Medline: 24801761]

XSL•FO
RenderX

26. Simon C, L'heureux J, Murray J, Winokur P, Weiner G, Newbury E, et al. Active choice but not too active: public perspectives on biobank consent models. Genet Med 2011 Sep;13(9):821-831 [FREE Full text] [doi: 10.1097/GIM.0b013e31821d2f88] [Medline: 21555942]
27. Steinsbekk KS, Kåre Myskja B, Solberg B. Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? Eur J Hum Genet 2013 Sep;21(9):897-902 [FREE Full text] [doi: 10.1038/ejhg.2012.282] [Medline: 23299918]
28. Murphy J, Scott J, Kaufman D, Geller G, LeRoy L, Hudson K. Public perspectives on informed consent for biobanking. Am J Public Health 2009 Dec;99(12):2128-2134. [doi: 10.2105/AJPH.2008.157099] [Medline: 19833988]
29. Harford T. Big data: are we making a big mistake? Financial Times 2014 Apr 28(Mar 28th) [FREE Full text]
30. Nuffield Council on Bioethics Working Party. The collection, linking and use of data in biomedical research and health care: ethical issues. In: Report. London: Nuffield Council on Bioethics; 2015.
31. Laurie G, Jones K, Stevens L, Dobbs C. A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data. In: Report. London: Nuffield Council on Bioethics/Wellcome Trust; 2014.
32. Carman D, Britten N. Confidentiality of medical records: the patient's perspective. Br J Gen Pract 1995 Sep;45(398):485-488 [FREE Full text] [Medline: 7546873]
33. General Medical Council. Ethical Guidance for Doctors. London: General Medical Council; 2009. Confidentiality: good practice in handling patient information URL: https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/confidentiality [accessed 2018-09-05] [WebCite Cache ID 72Cryed2b]
34. Kalra D, Gertz R, Singleton P, Inskip H. Confidentiality of personal health information used for research. BMJ 2006 Jul 22;333(7560):196-198 [FREE Full text] [doi: 10.1136/bmj.333.7560.196] [Medline: 16858053]
35. Sankar P, Mora S, Merz J, Jones N. Patient perspectives of medical confidentiality: a review of the literature. J Gen Intern Med 2003 Aug;18(8):659-669 [FREE Full text] [Medline: 12911650]
36. Higgins GL. The history of confidentiality in medicine: the physician-patient relationship. Can Fam Physician 1989 Apr;35:921-926 [FREE Full text]
37. Parsa-Parsi R. The Revised Declaration of Geneva: A Modern-Day Physician's Pledge. JAMA 2017 Nov 28;318(20):1971-1972. [doi: 10.1001/jama.2017.16230] [Medline: 29049507]
38. Caldicott F, Panel IIGO. Information: To share or not to share? In: The Information Governance Review. London: HMSO; Mar 2013.
39. Information Commissioner's Office. Anonymisation Code. London: Information Commissioner's Office; 2012. Anonymisation: managing data protection risk code of practice Internet URL: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf [accessed 2018-09-05] [WebCite Cache ID 72CsXW244]
40. HSCIC. NHS Digital. London: HSCIC Accredited Safe Haven Accreditation Process Stage 1 URL: https://tinyurl.com/yc62k6pk [accessed 2018-09-05] [WebCite Cache ID 72CsxA3l1]
41. ICES. 2016. Privacy at ICES URL: https://www.ices.on.ca/Data-and-Privacy/Privacy-at-ICES [accessed 2018-09-05] [WebCite Cache ID 72CtmWv1F]
42. Sweeney L. Data Privacy Working Paper. Pittsburgh: Carnegie Mellon University; 2000. Simple Demographics Often Identify People Uniquely URL: https://dataprivacylab.org/projects/identifiability/paper1.pdf [accessed 2018-09-05] [WebCite Cache ID 72Ctun4w0]
43. Elliott M, Mackey E, O'Hara K, Tudor C. UKAN Report. Manchester: UKAN Publications; 2016. The Anonymisation Decision-Making Framework URL: https://tinyurl.com/ycvjh5fo [accessed 2018-09-05] [WebCite Cache ID 72Cu15ZRQ]
44. Tang J, Korolova A, Bai X, Wang X, Wang X. arXiv. 2017. Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12 URL: https://arxiv.org/abs/1709.02753 [accessed 2018-09-05] [WebCite Cache ID 72CuHFlXX]
45. Chappell A. AdExchanger. 2016. Is All Personal Data Equal Under The Law? URL: https://adexchanger.com/data-driven-thinking/personal-data-equal-law/ [accessed 2018-09-05] [WebCite Cache ID 72CuOyOLO]
46. Cowan P. IT News. Treat anonymised data as personal information: Pilgrim URL: https://www.itnews.com.au/news/treat-anonymised-data-as-personal-information-pilgrim-411949 [accessed 2018-09-05] [WebCite Cache ID 72CuoZN4e]
47. FCC. Washington, DC: FCC; 2016. FCC Releases Proposed Rules to Protect Broadband Consumer Privacy URL: https://www.fcc.gov/document/fcc-releases-proposed-rules-protect-broadband-consumer-privacy [accessed 2018-09-06] [WebCite Cache ID 72E6tG9Gn]
48. Rich J. FTC. Washington, DC: FTC Bureau of Consumer Protection; 2016. Keeping Up with the Online Advertising Industry URL: https://www.ftc.gov/news-events/blogs/business-blog/2016/04/keeping-online-advertising-industry [WebCite Cache ID 72XnUwmlC]
49. Ohm P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Rev 2010;57:1701-1777 [FREE Full text]
50. El Emam EK. Risk-based De-identification of Health Data. IEEE Secur Privacy 2010;8(3):64-67 [FREE Full text]
51. Emam E, Khaled. Guide to the de-identification of personal health information. In: Guide to the De-Identification of Personal Health Information. Ontario: Taylor & Francis; 2018.

XSL•FO
RenderX

52. Directorate-General for Communication. Special Eurobarometer 59. Brussels: European Commission; 2014. Attitudes on Data Protection and Electronic Identity in the European Union URL: http://ec.europa.eu/commfrontoffice/publicopinion/index_en.htm [accessed 2018-09-05] [WebCite Cache ID 72CvTOpwP]
53. Davidson S, McLean C, Treanor S, Mori I, Aitken M, Cunningham-Burley S, et al. Davidson S, McLean C, Treanor S, Mori I, Aitken M, Cunningham-Burley S, Laurie G, Pagliari C, Sethi N. Edinburgh: Scottish Government; 2013. Public acceptability of data sharing between the public, private and third sectors for research purposes Internet URL: https://www.gov.scot/resource/0043/00435458.pdf [accessed 2018-09-05] [WebCite Cache ID 72Cvo5QqC]
54. Aitken M. Scottish Health Informatics Programme, Wellcome Trust. SHIP Public Engagement: Summary of Focus Group Findings URL: http://www.scot-ship.ac.uk/sites/default/files/Reports/Focus_Group_Findings_Briefing_Paper.pdf [accessed 2018-09-05] [WebCite Cache ID 72Cvu6xQH]
55. Royal Statistical Society Report. London: Royal Statistical Society; 2014. Public attitudes to the use and sharing of their data URL: https://www.slideshare.net/IpsosMORI/public-attitudes-to-the-use-and-sharing-of-their-data [accessed 2018-09-07] [WebCite Cache ID 72FiWMqPt]
56. Wellcome Trust. London: Wellcome Trust; 2016. The One-Way Mirror: Public attitudes to commercial access to health data URL: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf [accessed 2018-09-07] [WebCite Cache ID 72FifWCk0]
57. Ipsos MORI Social Research Institute MSRI. Report for Government Data Science Partnership. London: Ipsos MORI; 2016. Public dialogue on the ethics of data science in government URL: https://www.ipsos.com/sites/default/files/2017-05/data-science-ethics-in-government.pdf [accessed 2018-09-07] [WebCite Cache ID 72FimuwV1]
58. Ipsos MORI Social Research Institute. Report for ONS/ESRC. London: Ipsos MORI; 2014. Dialogue on Data URL: https://www.ipsos.com/sites/default/files/publication/1970-01/sri-dialogue-on-data-2014-census.pdf [accessed 2018-09-07] [WebCite Cache ID 72FitskHD]
59. Nissenbaum H. Privacy as contextual integrity. Wash L Rev 2004;79(1):119-158 [FREE Full text]
60. Solove D. A taxonomy of privacy. Univ PA Law Rev 2006;154(3):477-564 [FREE Full text]
61. Culnane C, Rubinstein B, Teague V. arXiv. Health data in an open world a report on re-identifying patients in the MBS/PBS dataset and the implications for future releases of Australian government data URL: https://arxiv.org/ftp/arxiv/papers/1712/1712.05627.pdf [accessed 2018-09-07] [WebCite Cache ID 72FjAeo24]
62. Lin J, Sadeh N, Amini S, Lindqvist J, Hong JI, Zhang J. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In: ACM Proc 2012 ACM Conference on Ubiquitous Computing. New York: ACM; 2012 Presented at: UbiComp '12; 2012; Pittsburgh p. 501-510.
63. Benisch M, Kelley P, Sadeh N, Cranor L. Capturing location-privacy preferences: Quantifying accuracy and user-burden tradeoffs. Personal Ubiquitous Computing 2011;15(7):679-694 [FREE Full text]
64. Liu B, Lin J, Sadeh N. Reconciling mobile app privacy and usability on smartphones: could user privacy profiles help? In: Proc 23rd Int Conf World Wide Web. New York: ACM; 2014 Presented at: 23rd Int Conf World Wide Web Internet; 2014; Seoul p. 201-212.
65. Rumbold J. Observations on the first UK Biobank Annual Meeting Jun 13th 2016. In: Personal Observation. 2016 Presented at: First UK Biobank Annual Meeting; Jun 13th 2016; London.
66. Sethi N, Laurie G. Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together. Med Law Int 2013;13(2):168-204. [doi: 10.1177/0968533213508974]
67. Laurie G, Ainsworth J, Cunningham J, Dobbs C, Jones K, Kalra D, et al. On moving targets and magic bullets: Can the UK lead the way with responsible data linkage for health research? Int J Med Inform 2015;84(11):933-940. [doi: 10.1016/j.ijmedinf.2015.08.011]
68. Article 29 Data Protection Working Party. Article 29 Data Protection Working Party Opinions and Recommendations. Brussels: European Commission; 2007. Opinion 4/2007 on the concept of personal data URL: http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf [accessed 2018-09-05] [WebCite Cache ID 72CzvA3ra]
69. Information Commissioner's Office. ICO Decision Notices. London: Information Commissioner's Office; 2015. FS50565190 Queen Mary's University of London URL: https://ico.org.uk/media/action-weve-taken/decision-notices/2015/1560081/fs_50565190.pdf [accessed 2018-09-05] [WebCite Cache ID 72D00oNke]
70. Academy of Medical Sciences. Data in Safe Havens Internet. In: Workshop Report. London: Academy of Medical Sciences; 2014.
71. Patki, Neha WR. The synthetic data vault. In: IEEE Xplore.: IEEE; 2016 Presented at: 2016 IEEE International Conference on Data Science and Advanced Analytics; 2016; Montreal p. 399-410 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7796926
72. Department of Digital Culture, Media and Sport. Statement of Intent. London: Department of Digital Culture, Media and Sport; 2017. A New Data Protection Bill: Our Planned Reforms Internet URL: http://www.documentcloud.org/documents/3921345-2017-08-07-DP-Bill-Statement-of-Intent.html [accessed 2018-09-05] [WebCite Cache ID 72D0TeLeo]

XSL•FO
RenderX

73. New Zealand Data Futures Forum. New Zealand Data Futures Forum. Wellington: New Zealand Data Futures Forum; 2014. Harnessing the economic and social power of data Internet URL: https://www.nzdatafutures.org.nz/sites/default/files/NZDFF_harness-the-power.pdf [accessed 2018-09-07] [WebCite Cache ID 72Fj2FOii]
74. Phillips M, Dove E, Knoppers B. Criminal Prohibition of Wrongful Re-identification: Legal Solution or Minefield for Big Data? J Bioeth Inq 2017 Dec;14(4):527-539 [FREE Full text] [doi: 10.1007/s11673-017-9806-9] [Medline: 28913771]
75. Stern R. Does 'Ought' Imply 'Can'? And Did Kant Think It Does? Utilitas 2004 Feb 11;16(01):42-61. [doi: 10.1017/S0953820803001055]
76. Edwards SJL, Kirchin S, Huxtable R. Research ethics committees and paternalism. Journal of Medical Ethics 2004 Feb 01;30(1):88-91. [doi: 10.1136/jme.2002.000166]
77. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS One 2011;6(12):e28071 [FREE Full text] [doi: 10.1371/journal.pone.0028071] [Medline: 22164229]
78. El Emam K. Methods for the de-identification of electronic health records for genomic research. Genome Med 2011 Apr 27;3(4):25 [FREE Full text] [doi: 10.1186/gm239] [Medline: 21542889]
79. Cavoukian A, El Emam K. Information Privacy Commissioner's Office. Ontario: Information Privacy Commmissioner's Office; 2011. Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy URL: https://www.ipc.on.ca/images/Resources/anonymization.pdf [accessed 2018-09-06] [WebCite Cache ID 72E52RqgK]
80. Nass S, Levit L, Gostin L. Beyond the HIPAA Privacy Rulenhancing Privacy, Improving Health Through Research. In: National Academies Collection. Washington, DC: National Academies Press; 2009.
81. El Emam K, Jabbouri S, Sams S, Drouet Y, Power M. Evaluating common de-identification heuristics for personal health information. J Med Internet Res 2006 Nov 21;8(4):e28 [FREE Full text] [doi: 10.2196/jmir.8.4.e28] [Medline: 17213047]
82. Dwork C, Roth A. The algorithmic foundations of differential privacy. Found Trends Theor Comput Sci 2014;9(3-4):211-407. [doi: 10.1561/0400000042]
83. Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. l-diversity: Privacy beyond k-anonymity. In: ACM Trans Knowl Discov Data. New York: IEEE; 2007 Presented at: 22nd International Conference on Data Engineering; 3-7th April 2006; Atlanta p. 3.
84. Li N, Li T, Venkatasubramanian S. t-closeness: Privacy beyond k-anonymity and l-diversity. In: Proceedings of 2007 IEEE 23rd International Conference on Data Engineering. New York: IEEE; 2007 Presented at: 2007 IEEE 23rd International Conference on Data Engineering; Apr 16-20th; Istanbul p. 106-115.
85. Sweeney L. k-anonymity: a model for protecting privacy. Int J Uncertainty, Fuzziness Knowledge-based Systems 2002;10(5):557-570. [doi: 10.1142/S0218488502001648]
86. Laurie G, Sethi N. Towards Principles-Based approaches to Governance of Health-Related Research Using Personal Data. Eur J Risk Regulation 2013;4(1):43-57 [FREE Full text]
87. Narayanan A, Shmatikov V. Robust De-anonymization of Large Sparse Datasets. In: SP '08 Proceedings of the 2008 IEEE Symposium on Security and Privacy. New York: ACM; 2008 Presented at: 2008 IEEE Symposium on Security and Privacy; May 18-21 2008; Oakland, CA p. 111-125.
88. Hill K. Forbes. USA: Forbes; 2012. How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did URL: https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/ [accessed 2018-09-06] [WebCite Cache ID 72E5C1qFU]
89. BBC News. London: BBC; 2016. Charities fined over "wealth screening" data breaches URL: https://www.bbc.co.uk/news/uk-38230335 [WebCite Cache ID 72Xnai4QS]

Abbreviations

GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
ICO: Information Commissioner's Office
SHIP: Scottish Health Informatics Programme
UKAN: UK Anonymisation Network
EU: European Union

XSL•FO
RenderX

Edited by G Eysenbach; submitted 06.12.16; peer-reviewed by J Bian, JC Hsieh, A James, M Taylor, J Bell; comments to author 18.02.17; revised version received 20.04.17; accepted 21.06.18; published 22.11.18.

Please cite as:
Rumbold J, Pierscionek B
Contextual Anonymization for Secondary Use of Big Data in Biomedical Research: Proposal for an Anonymization Matrix
JMIR Med Inform 2018;6(4):e47
URL: http://medinform.jmir.org/2018/4/e47/
doi:10.2196/medinform.7096
PMID:30467101

©John Rumbold, Barbara Pierscionek. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 22.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinical Named Entity Recognition From Chinese Electronic Health Records via Machine Learning Methods

Yu Zhang[1], BMS; Xuwen Wang[1], PhD; Zhen Hou[1], BE; Jiao Li[1], PhD

Institute of Medical Information and Library, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

Corresponding Author:
Jiao Li, PhD
Institute of Medical Information and Library
Chinese Academy of Medical Sciences
Peking Union Medical College
3rd Yabao Road, Chaoyang District, Beijing, China
Beijing,
China
Phone: 86 18618461596
Email: li.jiao@imicams.ac.cn

Abstract

Background: Electronic health records (EHRs) are important data resources for clinical studies and applications. Physicians or clinicians describe patients' disorders or treatment procedures in EHRs using free text (unstructured) clinical notes. The narrative information plays an important role in patient treatment and clinical research. However, it is challenging to make machines understand the clinical narratives.

Objective: This study aimed to automatically identify Chinese clinical entities from free text in EHRs and make machines semantically understand diagnoses, tests, body parts, symptoms, treatments, and so on.

Methods: The dataset we used for this study is the benchmark dataset with human annotated Chinese EHRs, released by the China Conference on Knowledge Graph and Semantic Computing 2017 clinical named entity recognition challenge task. Overall, 2 machine learning models, the conditional random fields (CRF) method and bidirectional long short-term memory (LSTM)-CRF, were applied to recognize clinical entities from Chinese EHR data. To train the CRF–based model, we selected features such as bag of Chinese characters, part-of-speech tags, character types, and the position of characters. For the bidirectional LSTM-CRF–based model, character embeddings and segmentation information were used as features. In addition, we also employed a dictionary-based approach as the baseline for the purpose of performance evaluation. Precision, recall, and the harmonic average of precision and recall (F1 score) were used to evaluate the performance of the methods.

Results: Experiments on the test set showed that our methods were able to automatically identify types of Chinese clinical entities such as diagnosis, test, symptom, body part, and treatment simultaneously. With regard to overall performance, CRF and bidirectional LSTM-CRF achieved a precision of 0.9203 and 0.9112, recall of 0.8709 and 0.8974, and F1 score of 0.8949 and 0.9043, respectively. The results also indicated that our methods performed well in recognizing each type of clinical entity, in which the "symptom" type achieved the best F1 score of over 0.96. Moreover, as the number of features increased, the F1 score of the CRF model increased from 0.8547 to 0.8949.

Conclusions: In this study, we employed two computational methods to simultaneously identify types of Chinese clinical entities from free text in EHRs. With training, these methods can effectively identify various types of clinical entities (eg, symptom and treatment) with high accuracy. The deep learning model, bidirectional LSTM-CRF, can achieve better performance than the CRF model with little feature engineering. This study contributed to translating human-readable health information into machine-readable information.

(JMIR Med Inform 2018;6(4):e50) doi:10.2196/medinform.9965

KEYWORDS

clinical named entity recognition; machine learning; treatment; diagnosis; physical examination; human body; syndrome; electronic health records; bidirectional LSTM-CRF

XSL•FO
RenderX

Introduction

Background

Electronic health records (EHRs) comprise individuals' health information such as laboratory test results, diagnosis, and medications. This information includes various data types, from structured information such as laboratory test results consisting of test items and the corresponding values, to unstructured data such as clinical narratives in discharge notes [1]. Benefiting from the development of big data techniques, large-scale EHR data mining has become widely used in data-driven medical studies, clinical decision making, and health management. However, plenty of key information on health care is buried in the large amount of unstructured narratives, which makes it difficult to be analyzed computationally. Therefore, clinical named entity recognition (CNER), which is used to identify the boundary of clinical entities such as body parts and diagnoses and then classify them into predefined categories, has been extensively used to extract structured information automatically from English and Chinese EHRs [2-4].

Early named entity recognition (NER) systems often use rule-based approaches that rely on various dictionary resources. More recently, machine learning (ML)–based approaches have been applied to NER, such as maximum entropy (ME), conditional random fields (CRF), support vector machines (SVM), structural support vector machines (SSVM), and multiple deep learning methods [5-10]. Liu et al [11] employed a CRF model based on multiple features including bag-of-characters (BOC), part-of-speech (POS), dictionary, and word-clustering features to identify clinical entities from EHRs. Experiments on 220 clinical tests with different feature combinations showed that a CRF model based on the combination of features including POS features, dictionary features, as well as word clustering features achieved the best performance with an F1 score of 0.8915. Liang et al [12] proposed a novel cascade-type method, which integrated the sentence category classifier from an SVM and the CRF-based clinical entity recognition, to recognize drug names from 324 Chinese admission notes. Their approach achieved an F1 score of 0.935 for the recognition of traditional Chinese medicine drug names and 0.917 for Western medicine drug names. Lei et al [2] systematically investigated the effects of different types of features and different ML models (including CRF, SVM, ME, and SSVM) for CNER on Chinese EHRs. Experiments on their manually annotated corpus of 400 discharge summaries and 400 admission notes showed that both the "word segmentation" feature and the "section information" feature improved the performance of CNER. In addition, among the ML models, SSVM achieved the best performance with an F1 score of 0.9001 and 0.9352 on discharge summaries and admission notes, respectively.

Traditional ML-based approaches such as CRF can achieve good performance on the sequence-labeling tasks but usually rely heavily on hand-engineered features and medical knowledge. However, deep learning methods such as Convolutional Neural Network (CNN) and Recurrent Neural Networks (RNN) can achieve state-of-the-art performance with little feature engineering. Wu et al [8] applied a deep neural network, developed by Ronan Collobert [13], on the CNER task in Chinese clinical text with only word embeddings, achieving an F1 score of 0.9280. Lample et al [14] proposed a bidirectional long short-term memory (LSTM-CRF) model for NER, which achieved an F1 score of 0.9094 on the CoNLL-2003 test set with word embeddings from supervised and unsupervised learning. Misawa et al [15] proposed a "character based" CNN-bidirectional LSTM-CRF model to extract entities from the Japanese Mainichi newspaper corpus. In their model, a CNN model was first used to extract subword information from Japanese characters, and then, the extracted subword information concatenated with the word embedding was fed into a bidirectional LSTM-CRF model to identify entities. Zhu et al [16] developed an end-to-end deep learning model, named GRAM-CNN, for CNER tasks, in which a modified CNN model was first employed to extract local features around a word, and then, a CRF layer was used to model labels jointly based on the output of GRAM-CNN. Their model achieved an F1 score of 87.26% on the Biocreative II dataset. Hu et al [17] built a vote-based hybrid system for the China Conference on Knowledge Graph and Semantic Computing (CCKS) 2017 CNER challenge task, which received the first place with an F1 score of 0.9102. Their hybrid system integrated 4 individual models, including (1) a rule-based model; (2) a CRF model; and (3) 2 bidirectional LSTM models, a conventional bidirectional LSTM model based on word embeddings and a modified bidirectional LSTM model with a fully connected layer added after the LSTM layer to concatenate some hand-crafted features with the LSTM outputs. A total of 4 models were deployed independently for the CNER task with corresponding F1 scores of 0.8682, 0.8969, 0.9017, and 0.8957, respectively. Finally, a vote-based approach was used to combine their results: an entity is selected only when it has been predicted by at least two methods. However, the hybrid system takes considerable time and effort for feature engineering, model constructing, and parameter tuning.

Most of the previous studies on CNER primarily focus on English clinical texts. Various ML models have shown significant performance on CNER on English EHRs. Compared with English CNER, Chinese CNER faces more obstacles and still remains a challenge, which may due to the following reasons: (1) few open access Chinese EHR corpora; (2) a small number of Chinese medical dictionaries and ontology libraries; and (3) complicated properties of the Chinese language, such as the lack of word boundaries, the complex composition forms, and word forms remaining unchanged in all kinds of tense or POS [18,19]. Until the recent 2 years, the number of studies on Chinese CNER has increased rapidly, boosting the performance of the models on Chinese CNER.

Objectives

In this study, we investigate 2 automatic methods, bidirectional LSTM-CRF and the CRF model, in terms of simultaneously identifying 5 types of clinical entities from Chinese EHR data. Experiment results indicate that the 2 ML models showed significant performance on each type of entity, demonstrating their effectiveness in recognizing multiple types of clinical entities for further data-driven medical research. Our

XSL•FO
RenderX

bidirectional LSTM-CRF model can capture not only the past and future input features through the bidirectional LSTM layer but also the sentence-level tag information via the CRF layer. Its performance is comparable with the Top 1 system (F1 score 0.9043 vs 0.9102) in the CCKS 2017 CNER challenge task and better than that of each of the 4 individual models of the Top 1 hybrid system, which needs much effort for feature engineering and model constructing. The bidirectional LSTM-CRF model achieves state-of-the-art performance by utilizing only the character and segmentation information, which significantly alleviates the human work involved in feature engineering to a large extent.

Methods

Datasets

A total of 2 datasets were used in this study, the first one is an annotated corpus, which is used for training and testing, whereas the second one, regarded as the development set, is an unlabeled corpus for learning character embedding. All data are derived from the progress notes and examination results of in-patients' EHRs released by the CCKS 2017 CNER challenge task [20]. The first dataset involves 400 patients' EHR data, and for each patient, it contains 4 data fields, including (1) general items: usually contain the patient's demographics and the reasons for admission; (2) medical history: consists of the patients' past disease history and corresponding treatment, the reasons for current admission, outpatient test results with diagnosis and treatment, and the tests after hospitalization; (3) diagnosis and treatment: mainly include the tests after hospitalization, corresponding diagnosis, and detailed treatment and body condition after treatment (if worse or new symptoms appear, test once again; corresponding diagnosis and treatment will be contained); and (4) discharge note: involves patients' complaints about their body condition, final tests before discharge, and the doctor's summary of the patients' body condition. Moreover, for each field, 5 types of clinical entities—symptom, test, diagnosis, treatment, and body part—were annotated. An example of the original EHR text is shown in Textbox 1 and its manually annotated gold standard provided by the CCKS organizer is shown in Table 1, in which pos_b and pos_e denote the start and end position of the entity in the text. In summary, a total of 10,142 symptoms, 12,689 tests, 1275 diagnoses, 1513 treatments, and 13,740 body parts were annotated in the first dataset. The first dataset was further divided into 2 parts, 300 patients' data as the training set and 100 patients' data as the test set. The distribution of entities among the training and test sets is shown in Table 2.

The second dataset includes 2605 patients' unlabeled EHR data, which was used for learning character embeddings. Character embeddings were trained using individual Chinese characters.

Dictionary-Based Clinical Named Entity Recognition

Traditionally, dictionary-based CNER approaches utilize medical dictionary resources such as the Unified Medical Language system, Medical Subject Headings, and RxNorm. For Chinese clinical entity recognition, we constructed a new dictionary on the basis of the Chinese Unified Medical Language System (CUMLS) [21] and the training corpus. CUMLS is a knowledge organization system with more than 30,000 medical subject headings and 100,000 medical terms, which incorporate more than 10 thesauruses, taxonomies, glossaries, and medical corpora. In this study, we only choose 54 categories of medical terms, which are related to the 5 types of entities in this study and classify them into 5 predefined categories manually (in terms of the clinical entity types) to build our dictionary. Overall, 6 categories of medical terms from CUMLS are classified as "diagnosis," 22 categories as "test," 9 categories as "body part," 4 categories as "symptom," and 13 categories as "treatment." Finally, we construct the dictionary by integrating all the clinical named entities, derived from the annotated labels provided by the CCKS organizer, in the 300 training set with the selected part of the CUMLS. The dictionary we build contains not only the medical terms from medical vocabulary but also the terms from the clinical text, which makes it more suitable for CNER tasks. Maximum forward matching was adopted while extracting clinical entities based on our dictionary.

Machine Learning Methods for Clinical Named Entity Recognition

CNER is generally converted into a sequence-labeling problem or a classification problem. Sequence-labeling problem means, given a sequence of input tokens $A=(a_1,...,a_n)$ and a predefined set of labels L, determine a sequence of labels $B=(b_1,\ldots,b_n)$ with the largest joint probability for the sequence of input tokens [22] ($b_i \in L$ for $1 \leq i \leq n$). Classification problem means for each input token x, determine the label with the highest probability of classification among the predefined set of labels L. As for CNER, the labels incorporate 2 concepts, the type of the clinical entity and the position of the token within the entity. In this study, we utilize the typical "BIO" labels [23] to represent the position of the tokens within the entities. In BIO labels, B means the token is the beginning of an entity, I means the token is inside an entity, and O means the token is outside of an entity. As there are 5 types of entities, we have overall 11 labels including 5 B classes (B-symptom, B-test, B-diagnosis, B-treatment, and B-body) and their corresponding I classes (I-symptom, I-test, I-diagnosis, I-treatment, and I-body) and O. For instance, a body part entity "心房" is made up of 2 Chinese characters "心" and "房" that are annotated with label B-body and I-body, respectively. In the following sections, we introduce the sequence-labeling algorithm CRF as well as the deep learning models (bidirectional LSTM-CRF) for CNER.

XSL•FO
RenderX

Textbox 1. An example of the original electronic health record text.

1、患儿为4岁儿童，起病急，病程短。2、以咳嗽，发热为主症。3、查体：咽部稍充血，双扁桃体稍肿大。双肺呼吸音粗，可闻及中小水泡音，结合胸片故诊断为：支气管肺炎。给予静点头孢哌酮、炎琥宁联合抗感染、雾化吸入布地奈德、沙丁胺醇减轻气道高反应 (1. Patient is a 4-year-old children, acute onset, short duration. 2. Main symptoms are cough and fever. 3. Examination: the throat is slightly congestive, double tonsils are slightly swollen. Lung breath sounds thick, a small and medium-sized bubble sound can be heard. Combined with the chest x-ray, diagnosed as: bronchopneumonia. Given cefoperazone combined with andrographolide for anti-infection, aerosolized inhaled budesonide and salbutamol to reduce airway hyperresponsiveness.)

Table 1. An example of the manually annotated golden standard.

Entity	pos_b[a]	pos_e[b]	Entity type
咳嗽 (cough)	21	22	Symptom
发热 (fever)	24	25	Symptom
查体 (examination)	32	33	Test
咽部 (throat)	35	36	Body part
充血 (congestion)	38	39	Symptom
双扁桃体 (double tonsils)	41	44	Body part
肿大 (swollen)	46	47	Symptom
双肺 (lung)	49	50	Body part
呼吸音 (breath sound)	51	53	Test
胸片 (chest x-ray)	67	68	Test
支气管肺炎(bronchopneumonia)	74	78	Diagnosis
头孢哌酮 (cefoperazone)	84	87	Treatment
炎琥宁 (andrographolide)	89	91	Treatment
布地奈德 (budesonide)	102	105	Treatment
沙丁胺醇 (salbutamol)	107	110	Treatment
气道 (airway)	113	114	Body part

[a]pos_b: start position.

[b]pos_e: end position.

Table 2. Distribution of entities among the training set and the test set.

Dataset	Number of patients	Body part	Diagnosis	Symptom	Test	Treatment	Total
Training set	300	10,719	722	7831	9546	1048	29,866
Test set	100	3021	553	2311	3143	465	9493
All	400	13,740	1275	10,142	12,689	1513	39,359

Conditional Random Fields–Based Clinical Named Entity Recognition

CRF is a probabilistic undirected graphical model, which was first proposed by Lafferty in 2001 [24]. It overcomes the shortcomings of the Hidden Markov Model and also solves the label-bias problem of the Maximum Entropy Markov Model. As it takes into account the joint probability distribution of the output sequence of labels, it has been widely used for sequence labeling tasks such as POS tagging, Chinese word segmentation, NER, and CNER. The CRF model decodes the sequence-labeling problem by undirected Markov chain and the Viterbi algorithm with the training criteria: maximize the likelihood estimation of conditional probability of the output sequence of labels Y given the input sequence X. In this study, X is the random variable over the input Chinese characters sequence and Y is the random variable over the corresponding label sequence.

Let P(Y|X) be a linear chain conditional random field. Under the condition that the value of random variable X is x (eg, “患者左腹压痛; patients with left abdominal pressing pain”), the conditional probability of which the random variable Y is y (eg, “O, O, B-body, I-body, B-symptom, and I-symptom”) is defined as:

$$P(y|x) = 1/Z(x) * \exp\{ \sum_{i,k} \lambda_k t_k (y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l (y_i, x, i) \}$$

$$Z(x) = \sum_y \exp\{ \sum_{i,k} \lambda_k t_k (y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l (y_i, x, i) \},$$

XSL•FO
RenderX

in which Z(x) denotes the normalization factor, y_i (eg, I-symptom) is the label of x_i (eg, "痛; pain" in the "患者左腹压痛; patients with left abdominal pressing pain"), and then y_{i-1} (B-symptom) is the label of x_{i-1} "压 (pressing)". t_k (depends on the current label y_i and former label y_{i-1}) and s_l (depends on the current label y_i) denote the feature functions, and λ_k and μ_l denote their corresponding weights. Once the corresponding weights are learned, the labels of a new input sequence can be predicted according to P(y|x). The prediction process can be achieved in an efficient way using the Viterbi algorithm. In this study, we use the CRF++ package [25], one of the most popular implementations of CRF, for the implementation of CRF model. The features for training the CRF model are described in the Feature Selection section.

Bidirectional Long Short-Term Memory-Conditional Random Fields–Based Clinical Named Entity Recognition

Recently, multiple deep neural architectures have been exploited for NER tasks [26-28], among which RNN models usually achieve the best performance, especially the bidirectional LSTM-CRF model. In theory, RNNs are capable of capturing long-distance dependencies; in practice, they fail due to the gradient vanishing or exploding problems. LSTMs are variants of RNNs designed to cope with these gradient vanishing problems by incorporating a memory cell [29]. Figure 1 shows the structure of an LSTM unit at step t [7]. An LSTM unit contains an input gate i_t, which controls the proportion of input information to the memory cell; a forget gate f_t, which controls the proportion of the previous information to forget; a memory cell c_t, which memorizes the long-distance context information; and an output gate o_t, which controls the proportion of the output information for the next step. The implementation of the LSTM unit is shown as follows:

$$i_t = \sigma(W_{xi}\, x_t + w_{hi}\, h_{t-1} + W_{ci}\, c_{t-1} + b_i)$$
$$f_t = \sigma(W_{xf}\, x_t + W_{hf}\, h_{t-1} + W_{cf}\, c_{t-1} + b_f)$$
$$c_t = f_t \bullet c_{t-1} + i_t \bullet \tanh(W_{xc}\, x_t + W_{hc}\, h_{t-1} + b_c)$$
$$o_t = \sigma(W_{xo}\, x_t + W_{ho}\, h_{t-1} + W_{co}\, c_t + b_o)$$
$$h_t = o_t \bullet \tanh(c_t),$$

where σ denotes the element-wise sigmoid function; • denotes the element-wise product; and b_i, b_f, b_c, and b_o denote the bias vectors. W denotes the weight matrix, x_t denotes the input vector corresponding to the current Chinese character, and h_t denotes the output vector of the LSTM, which represents the context information of the current Chinese characters.

For many sequence-labeling tasks, it is beneficial to have access to both past (left) and future (right) contexts. However, the LSTM's hidden state h_t takes information only from past. Bidirectional LSTM model presents each sequence forward and backward to 2 separate hidden states to capture past and future information, respectively. Then, the 2 hidden states are concatenated to further form the final output. Bidirectional LSTM-CRF model, which takes advantage of both bidirectional LSTM and CRF, can simultaneously utilize the past and future input features through the forward and the backward LSTM layer and the sentence level tag information via the CRF layer. The architecture of the bidirectional LSTM-CRF is shown in Figure 2, which consists of an input layer, 2 LSTM layers, and a CRF layer.

When predicting the tags of Chinese characters, first, given a sentence $S=(c_1,...,c_n)$, each character c_t ($1 \le t \le n$) is represented by vector x_t (the concatenation of the character embeddings and the segmentation information) generated in the input layer. Second, the forward and the backward LSTM layer take the sequence of character representations $X=(x_1,...,x_n)$ as input and generate the representation of the left ($h_l=h_{l1},...,h_{ln}$) and right ($h_r=h_{r1},...,h_{rn}$) context for each character, respectively. Third, the sequence of overall context representations is $h=(h_1,...,h_n)$, where h_t is the concatenation of h_{lt} and h_{rt}. Finally, the sequence of overall context representations is taken as input for the CRF layer to predict the output label sequence $L=(l_1,...,l_n)$.

XSL•FO
RenderX

Figure 1. A long short-term memory unit. it: input gate; ft: forget gate; ct: memory cell; ot: output gate; ht: output vector of the LSTM.

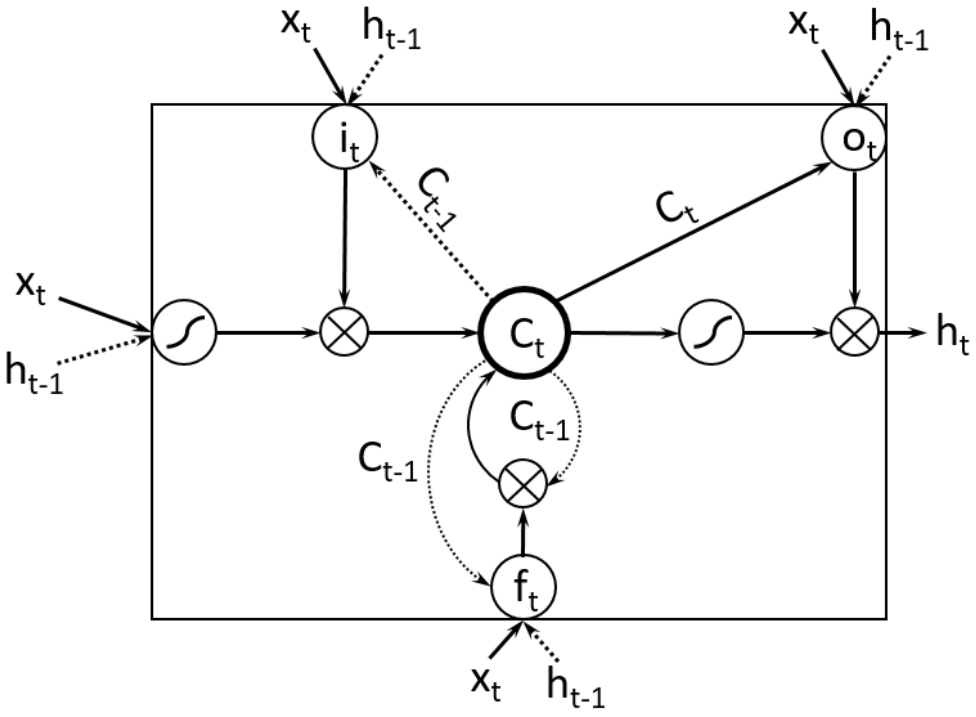
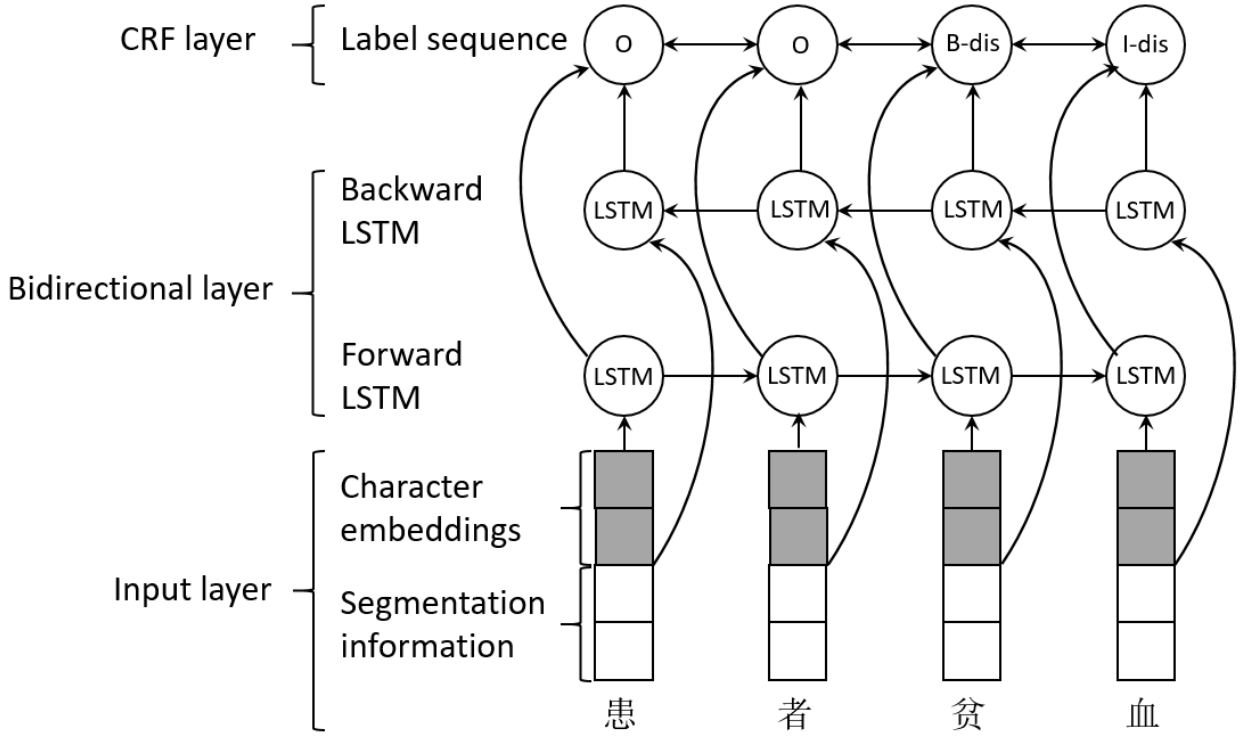


Figure 2. Architecture of the bidirectional long short-term memory-conditional random fields. LSTM: long short-term memory; CRF: conditional random fields; B-dis: B-diagnosis; I-dis: I-diagnosis.



Feature Selection

For training the CRF model, we select 4 types of features, BOC, POS tags, character types (CT), as well as the position of the character in the sentence (POCIS). NLPIR Chinese word segmentation system Institute of Computing Technology, Chinese Lexical Analysis System (ICTCLAS)-2016 [30] is utilized for word segmentation. While using ICTCLAS-2016 for segmentation, POS tags are generated simultaneously. As we use character-level information instead of word-level information, the POS tag of the Chinese character is just the POS tag of the corresponding word, which contains that character. In addition, we manually classify all the characters in the EHR dataset into 5 CT (including W: common character;

XSL•FO
RenderX

D: numbers; L: letters; S: ending punctuation; and P: symmetrical punctuation). To validate the effectiveness of the bidirectional LSTM-CRF model in identifying clinical entities with much less feature engineering than the CRF model, different combinations of features are fed into the CRF model.

For training the bidirectional LSTM-CRF model, we employ the character embeddings and segmentation information as our features. Character embeddings are learned through Google's word2vec [31] on the 2605 patients' unlabeled dataset. The segmentation information is generated by the Jieba segmentation system [32].

Experiment

Experimental Setup

We conducted an experimental study to compare ML-based CNER with the dictionary-based approach. First, we divide the first dataset into 2 parts, the first part, which contains 300 patients' EHR data, for training, and the second part, which involves 100 patients' EHR data, for testing.

In the CRF model, the content window size is set to 5 for extracting character features, including the 2 preceding characters, the current character, and the 2 following characters. Different combinations of features have been tried to train the CRF model, including (1) BOC; (2) BOC+POS tags; (3) BOC+POS tags+CT; and (4) BOC+POS tags+CT+POCIS. In addition, we applied 10-fold cross validation for tuning model parameters. In 10-fold cross validation, the training set was randomly divided into 10 parts; each time, we used 1 part as the test set and the remaining 9 parts as the training set for the experiment. Finally, we used the average F1 score of the 10 experiments to estimate the accuracy of the model and tune the parameters.

As for the deep learning model, we fix the learning rate at 0.0004, the dropout at 0.5, and the character embedding dimension at 100. The number of hidden units in bidirectional LSTM-CRF is set to 100, and the optimizer is set to Adam.

As for dictionary-based CNER, which is regarded as the baseline approach, maximum forward matching based on our dictionary is adopted for extracting clinical entities.

Evaluation Criteria

The evaluation for this CNER challenge task is implemented through the algorithm provided by CCKS 2017 organizers, which reports the Precision (P), Recall (R), and F1 score for all clinical entities using exact matching methods [33]. According to the algorithm, we define $O=(O_1,...,O_m)$ as the output set of the system and $G=(G_1,...,G_n)$ as the manually annotated set (in terms of the golden standard) provided by the task organizer. Then o_i *0* and g_i *G* are strictly equivalent only when:

o_i.mention = g_j.mention

o_i.pos_b = g_j.pos_b

o_i.pos_e = g_j.pos_e

o_i.category = g_j.category

Here, mention represents the content of the entity, pos_b and pos_e separately denote the start and end position of the entity in the EHR text, and category represents the entity type. On the basis of the above equivalence relation, strict evaluation metrics are implemented as follows:

$P=|S \cap G| / |S|$

$R=|S \cap G| / |G|$

$F1 = 2PR / (P+R)$

Results

To validate the effectiveness of the ML models on simultaneously identifying various types of clinical entities from Chinese EHRs, we carried out comparative experiments on the basis of CCKS CNER corpus.

As shown in Table 3, the best overall performance was achieved by the bidirectional LSTM-CRF model with an F1 score of 0.9043, followed by the CRF models with F1 scores from 0.8547 to 0.8949, and finally the dictionary-based model with an F1 score of 0.5924. ML models achieved significantly better performance than the dictionary-based approach. In addition, with the number of features increasing, performance of the CRF model continued to improve, increasing F1 score from 0.8547 to 0.8949. However, even the best CRF model with all 4 types of features was slightly worse than the bidirectional LSTM-CRF model.

Besides the overall performance, Table 4 showed the detailed performance of the ML models as well as the dictionary-based model on each type of clinical entity. The bidirectional LSTM-CRF model achieved the highest recalls in all the 5 types of clinical entities, whereas the CRF model always achieved the highest precisions in each type of entity except for "treatment" type. Among the 5 types of entities, the "symptom" type of entities had the best performance with an F1 score over 0.96 in ML models, followed by the "test" type of entities with an F1 score around 0.94, whereas the "treatment" type of entities always received the worst performance with an F1 score less than 0.75. Furthermore, Figure 3 more intuitively shows the comparison of the detailed performance between the ML-based models and the dictionary-based approach.

XSL•FO
RenderX

Table 3. Overall performance of the bidirectional long short-term memory-conditional random fields model, conditional random fields–based models with different feature combinations, and the dictionary-based model.

Model	Precision	Recall	F1 score
Dictionary-based model	0.5215	0.6855	0.5924
CRF[a] model+BOC[b]	0.8792	0.8316	0.8547
CRF model+BOC+POS[c] tags	0.9065	0.8529	0.8789
CRF model+BOC+POS tags+CT[d]	0.9144	0.8658	0.8895
CRF model+BOC+POS tags+CT+POCIS[e]	0.9203	0.8709	0.8949
Bidirectional LSTM-CRF[f] model	0.9112	0.8974	0.9043

[a]CRF: conditional random fields.

[b]BOC: bag-of-characters.

[c]POS: part-of-speech.

[d]CT: character types.

[e]POCIS: position of the character in the sentence.

[f]LSTM-CRF: long short-term memory-conditional random fields.

Table 4. Detailed performance of the bidirectional long short-term memory-conditional random fields–based, conditional random fields–based, and dictionary-based clinical named entity recognition approaches.

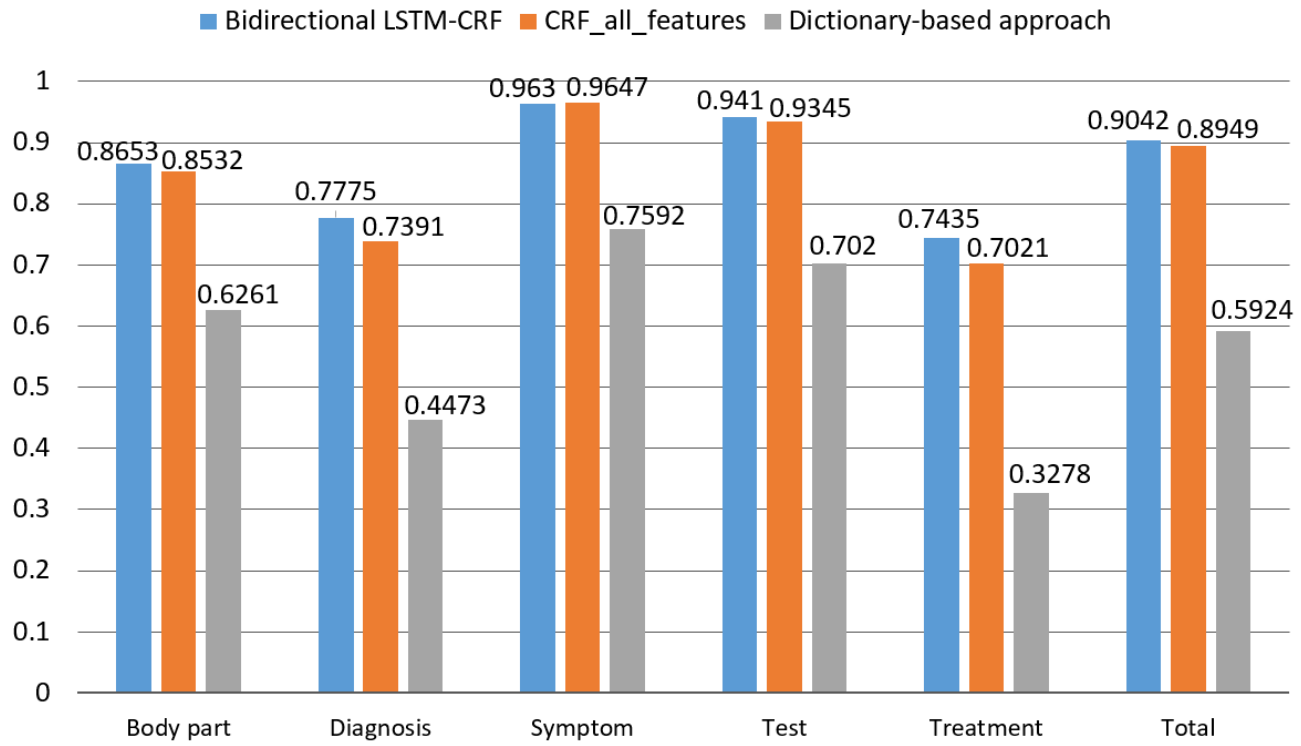
Entity type	Bidirectional LSTM-CRF[a]			CRF[b]_all_features			Dictionary-based approach		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Body part	0.8873	0.8444	0.8653	0.8909	0.8186	0.8532	0.6081	0.6452	0.6261
Diagnosis	0.8086	0.7486	0.7775	0.8148	0.6763	0.7391	0.3545	0.6058	0.4473
Symptom	0.9584	0.9675	0.9630	0.9715	0.9580	0.9647	0.7591	0.7594	0.7592
Test	0.9314	0.9510	0.9411	0.9459	0.9233	0.9345	0.7093	0.6949	0.7020
Treatment	0.7833	0.7075	0.7435	0.7581	0.6538	0.7021	0.2240	0.6108	0.3278
Total	0.9112	0.8974	0.9043	0.9203	0.8709	0.8949	0.5215	0.6855	0.5924

[a]LSTM-CRF: long short-term memory-conditional random fields.

[b]CRF: conditional random fields.

XSL•FO
RenderX

Figure 3. Comparison of F1 scores between dictionary-based approach and machine learning–based approaches among 5 entity types; LSTM-CRF: long short-term memory-conditional random fields; CRF: conditional random fields.



Discussion

Principal Findings

Essentially, recognizing various types of clinical entities allows extraction of the structured information of patients, which can be further exploited for data-driven medical research, clinical decision making, and health management. Compared with previous studies in CNER, ML-based methods can simultaneously extract 5 types of entities. Moreover, the proposed bidirectional LSTM-CRF model achieves a performance that is comparable with the Top 1 system, which is an ensemble model incorporating 4 ML models including a rule-based model, a CRF model, and 2 RNN models, in the CNER challenge only using character embeddings, and the segmentation information, therefore, reduces considerable efforts for feature engineering and model constructing.

Dictionary-Based Clinical Named Entity Recognition Versus Machine Learning–Based Clinical Named Entity Recognition

Experiments on the CCKS 2017 CNER challenge corpus show that ML-based models (bidirectional LSTM-CRF and CRF) achieve remarkably better performance than the dictionary-based method. Different from the maximum forward matching of the dictionary-based CNER, ML methods can sufficiently exploit the context information (eg, bag of Chinese characters and context representation information derived from LSTM), syntactic information (eg, POS tags), and structure information (eg, the position of the Chinese character in the sentence), which makes their performance significantly better. Furthermore, the performance of ML models is comparable with the Top 1 system in the CNER challenge with an overall F1 score of 0.9102, validating the effectiveness of the 2 ML-based methods in simultaneously recognizing multiple types of clinical entities for further data-driven medical studies.

Bidirectional Long Short-Term Memory-Conditional Random Fields Versus Conditional Random Fields

The bidirectional LSTM-CRF model achieves the best overall performance (see Table 3) but only utilizes the character embeddings and the segmentation information. Compared with the traditional CRF model, bidirectional LSTM-CRF not only takes advantage of CRF but also receives the benefits of bidirectional LSTM, which can generate long-distance context representations from the past and future input features. For example, given an input sequence of Chinese characters “生化检查: 谷丙转氨酶23.4 U/L, 谷草转氨酶21.7 U/L...葡萄糖5.78 mmol/L (biochemical tests: alanine aminotransferase 23.4 U/L, aspartate aminotransferase 21.7 U/L...glucose 5.78 mmol/L),” when predicting the labels of “葡萄糖 (glucose),” LSTM can capture the long-distance context information “生化检查 (biochemical tests)” and take it into labels prediction, which may make the prediction of the labels of “葡萄糖 (glucose)” be “B-test, I-test, I-test” rather than “B-treatment, I-treatment, I-treatment.”

Furthermore, by comparing the results of CRF models and the bidirectional LSTM-CRF model in Table 3, we find that, given the same features, bidirectional LSTM-CRF model performs obviously better than the CRF model. Even with more features, the CRF model is still slightly worse than the bidirectional LSTM-CRF model. The bidirectional LSTM-CRF model has a remarkable advantage in taking little effort for feature engineering to get higher efficiency and more robust performance in different types of entity recognition. However,

XSL•FO
RenderX

the CRF model can also perform well in CNER but requires elaborate feature engineering and, thus, lacks efficiency, scalability, and generality. In brief, similar to NER in other domains, deep learning models such as bidirectional LSTM-CRF show great potential on CNER in the medical domain, outperforming the traditional state-of-the-art method CRF, which involves massive feature engineering.

Differences Among the Performance of Five Types of Entities

Despite the impressive overall performance, the ML models do not show superiority over all the 5 types of clinical entities. As shown in Figure 3, among the 5 types of entities, the "symptom" type of entities achieve the best performance, followed by the "test" and "body part," whereas the performance of "diagnosis" and "treatment" is approximately 10% lower than that of the other 3 types. This may be due to 2 reasons: (1) the number of "diagnosis" type entities and "treatment" type entities is almost 10 times less than the other type of entities, as shown in Table 1, and fewer training samples limited the recognition effect and (2) a clinical entity may be annotated as different entity types in different contexts. For example, "头痛" (headache) is annotated as the "symptom" type in the context of "发作性头痛、头晕6年" (paroxysmal headache and dizziness for 6 years) but annotated to the "diagnosis" type in the context of "间断性头痛2周" (intermittent headache for 2 weeks). Incorporating medical domain information into the ML-based models and making a larger training set may help solve the problem.

Error Analysis

An error analysis on our 2 ML-based models shows that plenty of errors often occur when predicting tags on long entities with composite structures. For example, "高血压病腔隙性脑梗死" (hypertension Lacunar Cerebral Infarction), which is annotated as a "diagnosis" type entity in the golden standard, is automatically annotated as 2 entities "高血压" (hypertension) and "腔隙性脑梗死" (lacunar infarction) in our ML models. Especially, we find that, in the EHR text, a "body part" type of entity is often followed by a "symptom" or a "diagnosis" type of entity, which makes it difficult to identify the border between the 2 entities. For instance, in EHR text "股骨骨折 (femoral fracture)," the "body part" type of entity "股骨 (femur)" is followed by a "symptom" type of entity "骨折 (fracture)." Incorporating domain knowledge and medical dictionaries as well as combining the active learning methods with current ML models and increasing the scale of datasets might be the right path.

Furthermore, taking CRF model based on all features (BOC, POS tags, CT, and POCIS) as an example, we conduct an in-depth error analysis on its result to explore the effectiveness and limitations of the ML models on Chinese CNER either from a statistical view or from the clinical view. Table 5 shows the distribution of different types of errors as well as some examples, in which "GT-P" denotes the entities that were not identified by CRF; "P-GT" represents the entities recognized by CRF but are not in the ground truth; and "INTERSECT" denotes that for each entity, there is intersection between the ground truth and the entity predicted by CRF, for example, when extracting entities on EHR text "患者有脂肪肝病史 (the patient has a history of fatty liver)," the entity recognized by CRF is "肝 (liver)," having intersecting part "肝 (liver)" with the ground truth "脂肪肝 (fatty liver)." Overall, there are 1386 errors, 143 (10.32%) errors with type "GT-P," 604 (43.58%) errors with type "P-GT," and 639 (46.10%) errors with type "INTERSECT."

As for "GT-P" type of errors, only 1.51% (143/9493) entities of the test set are missed by the CRF model, which demonstrate its effectiveness in Chinese CNER. After further analysis on type "GT-P" errors from a medical view, we find that some entities missed by CRF model, which may be because the ground truth is not accurate, contain some punctuations that are not related to the entities. For example, the ground truth "肿，(swollen,)" should be "肿 (swollen)" rather than "肿 (swollen)" with punctuation ",". Moreover, some entities such as "对称 (symmetry)" do not belong to each type of clinical entity from the clinical view and should not appear in the ground truth. These entities are not recognized by the CRF model, which is not a problem of the model but a problem of ground truth. With more accurate ground truth, our results can be better. Moreover, some errors such as the "Symptoms and signs" type of new entities "活动障碍" (activity disorder), "听力下降" (hearing loss), and "功能障碍" (dysfunction) were not recognized by CRF, which may be because they never appear in the training set. Without sufficient training examples, it is challenging to effectively identify clinical entities, especially the unknown ones, for supervised ML models. Some studies [34-36] have attempted to apply unsupervised ML methods to recognize entities from clinical text on the basis of lexical resources, syntactic knowledge, and corpus statistics. It is worth making further efforts in Chinese clinical entity recognition using the unsupervised methods when lacking training data.

In addition, through the analysis on "P-GT" type of errors, we find that most of the entities in these types of errors are clinically meaningful, such as "2型糖尿病 (type 2 diabetes)" and "冠心病 (coronary disease)." These entities recognized by the CRF model should be the ground truth rather than errors. The reason behind this may be due to missing annotations while manually building the ground truth. Thus, these type of "errors" should be the advantage of our models, which could maintain high efficiency and accuracy during CNER, rather than errors. Moreover, some entities such as "腔隙性脑梗 (lacunar clog)" are new entities that never appeared in the ground truth. These entities are meaningful to clinicians and should be recognized. This proves that our model has the ability to identify a few new clinical entities from Chinese EHR. However, some entities recognized by our models, such as "比重 (proportion)," do not make any sense.

Finally, the deep analysis of "INTERSECT" type of errors shows that most of the errors are due to the different granularities between our results and the ground truth. For example, the ground truth for clinical text "患者于去年诊断为脑水肿 (the patient was diagnosed with cerebral edema last year)" is "水肿 (edema)" and our result is "脑水肿 (cerebral edema)." This is a limitation of ML models that cannot accurately identify entities at the appropriate granularity. However, plenty of entities appear to be annotated at different granularities in different EHR documents when building the

XSL•FO
RenderX

ground truth. For example, text “脑梗死 (cerebral infarction)” is sometimes annotated as “脑 (brain)” and sometimes annotated as “脑梗死 (cerebral infarction)” and text “右侧丘脑腔隙性脑梗死 (right thalamic lacunar infarction)” is sometimes annotated as “右侧丘脑 (right thalamic)” and “腔隙性脑梗死 (right lacunar infarction),” whereas it is sometimes annotated as “脑梗死 (cerebral infarction).” The ambiguity of the granularities in the ground truth will make the ML models more difficult to extract clinical entities on appropriate granularities. Specific annotation rules on annotation granularities as well as high-quality datasets could be constructed to further improve the performance of ML models on Chinese CNER.

Future Directions

In the future, we will not only develop new ML methods to enhance the accuracy of CNER but will also try to collect and standardize the recognized entities into the standard medical lexicons. Considering that different types of entities have different distributions in different fields of EHR, for instance, “treatment” type of entities often concentrates on the “diagnosis and treatment” field and rarely appears in the “general items” field, separately building ML-based models on each type of field data rather than on all EHR data may be a worthwhile study. As the amount of Chinese EHR data is limited, incorporating the active learning methods with ML models may be a possible future direction. Furthermore, when such structural patient information is used for data-driven medical studies, the time order of the clinical entities as well as their modifications are usually required. Therefore, a future direction is to identify more details of the clinical entities.

Table 5. Distribution of different types of errors in the results of the conditional random fields model based on all the 4 types of features (N=1386).

GT-P[a] (N=143)	P-GT[b] (N=604)	INTERSECT[c] (GT vs P; N=639)
尿蛋白- （urinary protein-）	2型糖尿病 （type 2 diabetes）	右侧丘脑腔隙性脑梗死 versus 右侧丘脑 + 腔隙性脑梗死 （right thalamic lacunar infarction vs right thalamic+lacunar infarction）
低血糖 （hypoglycemia）	冠心病（coronary disease）	胃肠 versus 急性胃肠炎 （stomach and intestine vs acute gastroenteritis）
对称 （symmetry）	胸 （chest）	糖尿病肾病 versus 糖尿病 + 肾病 （diabetic nephropathy vs diabetes+nephropathy）
瞳孔 （pupil）	腔隙性脑梗 （lacunar clog）	右下后 versus 右下后牙 （right lower back vs lower right posterior teeth）
冠心病 （coronary disease）	脂肪肝 （fatty liver）	水肿 versus 脑水肿 （edema vs brain edema）
肿， （swollen,）	比重 （proportion）	氨溴索注射液祛痰 versus 氨溴索注射液 （ambroxol injection to remove phlegm vs ambroxol injection）
胃粘膜 （gastric mucosa）	角膜 （cornea）	皮肤、粘膜 versus 皮肤 + 黏膜 （skin、mucous membrane vs skin+mucous membrane）
寒战 （chill）	脑萎缩 （encephalatrophy）	脂肪肝 versus 肝 （fatty liver vs liver）
无力 （faintness）	峰值 （peak value）	尼群地平药物 versus 尼群地平 （nitrendipine drug vs nitrendipine）
皮肤 （skin）	活动障碍 （activity disorder）	脑 versus 脑梗死 （brain vs cerebral infarction）

[a]GT-P: Entities that were not identified by CRF.

[b]P-GT: Entities recognized by CRF but are not in the ground truth.

[c]INTERSECT: For each entity, there is an intersection between the ground truth and the entity predicted by CRF.

Conclusions

CNER is one of the basic works of data-driven medical research. However, previous studies usually focused on recognizing a single type of clinical entity. In this study, we implemented 2 ML methods, including the bidirectional LSTM-CRF and the CRF models, for simultaneously recognizing 5 types of clinical entities from the Chinese EHR corpus provided by the CNER challenge of CCKS 2017. Compared with the baseline dictionary-based approach, ML methods show remarkably better performance than the former. Moreover, the deep learning model bidirectional LSTM-CRF, outperforming the traditional CRF model in the overall result, achieves state-of-the-art performance on the basis of the character and segmentation information, which alleviates the human work involved in feature engineering to a large extent.

Acknowledgments

The authors would like to thank the CCKS 2017 CNER challenge organizers for providing the training, development, and test corpora. The authors would also like to thank Kesheng Liu for the help with program development. This research is funded by the National Natural Science Foundation of China (Grant No. 81601573), the National Key Research and Development Program of China (Grant No. 2016YFC0901901), the National Population and Health Scientific Data Sharing Program of China, the National Population and Health Scientific Data Sharing Program of China, the Knowledge Centre for Engineering Sciences and Technology (Medical Centre), the Key Laboratory of Medical Information Intelligent Technology Chinese Academy of Medical

XSL•FO
RenderX

Sciences, the Key Laboratory of Knowledge Technology for Medical Integrative Publishing, and the Chinese Academy of Medical Sciences (Grant No. 2016RC330005) and supported by the Fundamental Research Funds for the Central Universities (Grant No. 3332018153).

Conflicts of Interest

None declared.

References

1. Jensen PB, Jensen LJ, 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]
2. Lei J, Tang B, Lu X, Gao K, Jiang M, Xu H. A comprehensive study of named entity recognition in Chinese clinical text. J Am Med Inform Assoc 2014 Sep;21(5):808-814 [FREE Full text] [doi: 10.1136/amiajnl-2013-002381] [Medline: 24347408]
3. Wu Y, Xu J, Jiang M, Zhang Y, Xu H. A study of neural word embeddings for named entity recognition in clinical text. AMIA Annu Symp Proc 2015;2015:1326-1333 [FREE Full text] [Medline: 26958273]
4. 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. [doi: 10.1093/bioinformatics/btx228] [Medline: 28881963]
5. He Y, Kayaalp M. Biological entity recognition with conditional random fields. AMIA Annu Symp Proc 2008 Nov 06:293-297 [FREE Full text] [Medline: 18999272]
6. Skeppstedt M, Kvist M, Nilsson GH, Dalianis H. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: an annotation and machine learning study. J Biomed Inform 2014 Jun;49:148-158 [FREE Full text] [doi: 10.1016/j.jbi.2014.01.012] [Medline: 24508177]
7. Liu Z, Yang M, Wang X, Chen Q, Tang B, Wang Z, et al. Entity recognition from clinical texts via recurrent neural network. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):67 [FREE Full text] [doi: 10.1186/s12911-017-0468-7] [Medline: 28699566]
8. Wu Y, Jiang M, Lei J, Xu H. Named entity recognition in Chinese clinical text using deep neural network. Stud Health Technol Inform 2015;216:624-628 [FREE Full text] [Medline: 26262126]
9. Tang B, Cao H, Wu Y, Jiang M, Xu H. Recognizing clinical entities in hospital discharge summaries using Structural Support Vector Machines with word representation features. BMC Med Inform Decis Mak 2013;13(Suppl 1):S1 [FREE Full text] [doi: 10.1186/1472-6947-13-S1-S1] [Medline: 23566040]
10. Jagannatha AN, Yu H. Structured prediction models for RNN based sequence labeling in clinical text. Proc Conf Empir Methods Nat Lang Process 2016 Nov;2016:856-865 [FREE Full text] [Medline: 28004040]
11. Liu K, Hu Q, Liu J, Xing C. Named Entity Recognition in Chinese Electronic Medical Records Based on CRF. 2017 Presented at: Web Information Systems and Applications Conference; November 11-12; Guang Xi, China p. 105-110.
12. Liang J, Xian X, He X, Xu M, Dai S, Xin J, et al. A novel approach towards medical entity recognition in Chinese clinical text. J Healthc Eng 2017;2017:4898963-4898916 [FREE Full text] [doi: 10.1155/2017/4898963] [Medline: 29065612]
13. Collobert R, Weston J, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (Almost) from scratch. J Mach Learn Res 2011;12(1):2493-2537.
14. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural Architectures for Named Entity Recognition. 2016 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 12-17 June, 2016; San Diego, California p. 260-270.
15. Misawa S, Taniguchi M, Miura Y, Ohkuma T. Character-based Bidirectional LSTM-CRF with words and characters for Japanese Named Entity Recognition. 2017 Presented at: The Workshop on Subword & Character Level MODELS in Nlp; 7 September, 2017; Copenhagen, Denmark p. 97-102.
16. Zhu Q, Li X, Conesa A, Pereira C. GRAM-CNN: a deep learning approach with local context for named entity recognition in biomedical text. Bioinformatics 2018 May 01;34(9):1547-1554 [FREE Full text] [doi: 10.1093/bioinformatics/btx815] [Medline: 29272325]
17. Hu J, Liu Z, Wang X, Chen Q, Tang B. A Hybrid System for Entity Recognition from Chinese Clinical Text. 2017 Presented at: Proceedings of the Evaluation Task at the China Conference on Knowledge Graph and Semantic Computing (CCKS 2017); 26-29 August, 2017; Chengdu, China URL: http://ceur-ws.org/Vol-1976/paper05.pdf
18. Yi HK, Huang JM, Yang SQ. A Chinese named entity recognition system with neural networks. ITM Web of Conferences 2017;12:04002. [doi: 10.1051/itmconf/20171204002]
19. Duan H, Zheng Y. A study on features of the CRFs-based Chinese named entity recognition. IJAIP 2011;3(2):287-294.
20. CCKS 2017 CNER challenge. 2017 Jun. URL: http://www.ccks2017.com/?page_id=51.pdf [accessed 2017-06-15]
21. Li DY, Hu TJ, Li JL, Qian Q, Zhu WY. Construction and application of the Chinese unified medical language system. Journal of Intelligence 2011;30(2):147-151.
22. Doan S, Collier N, Xu H, Pham HD, Tu MP. Recognition of medication information from discharge summaries using ensembles of classifiers. BMC Med Inform Decis Mak 2012 May 07;12:36 [FREE Full text] [doi: 10.1186/1472-6947-12-36] [Medline: 22564405]

XSL•FO
RenderX

23. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, et al. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. J Am Med Inform Assoc 2011 Sep;18(5):601-606 [FREE Full text] [doi: 10.1136/amiajnl-2011-000163] [Medline: 21508414]
24. Lafferty JD, Mccallum A, Pereira FC. Conditional Random Fields: Probabilistic Models for SegmentingLabeling Sequence Data. : Morgan Kaufmann Publishers Inc; 2001 Presented at: Eighteenth International Conference on Machine Learning; June 28 - July 1, 2001; Williamstown, MA p. 282-289.
25. CRF++. 2017 Jun. URL: https://sourceforge.net/projects/crfpp/.pdf [accessed 2017-06-15]
26. Chiu JP. Named entity recognition with bidirectional LSTM-CNNs. Trans Assoc Comput Linguist 2016;4:357-370 [FREE Full text]
27. Gridach M. Character-level neural network for biomedical named entity recognition. J Biomed Inform 2017 Jun;70:85-91. [doi: 10.1016/j.jbi.2017.05.002] [Medline: 28502909]
28. Ma X, Hovy E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. 2016 Presented at: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics; 7-12 August, 2017; Berlin, Germany URL: http://acl2016.org/index.php?article_id=68#long_papers.pdf
29. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [Medline: 9377276]
30. SEWM. ICTCLAS: Institute of Computing Technology, Chinese Lexical Analysis System URL: http://sewm.pku.edu.cn/QA/reference/ICTCLAS/FreeICTCLAS/.pdf [accessed 2017-06-15]
31. word2vec. URL: https://code.google.com/p/word2vec/.pdf [accessed 2017-06-15]
32. Python Software Foundation. URL: https://pypi.python.org/pypi/jieba/.pdf [accessed 2017-06-15]
33. Biendata. Evaluation criteria URL: https://biendata.com/competition/CCKS2017_2/evaluation/.pdf [accessed 2017-06-15]
34. Zhang S, Elhadad N. Unsupervised biomedical named entity recognition: experiments with clinical and biological texts. J Biomed Inform 2013 Dec;46(6):1088-1098 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.004] [Medline: 23954592]
35. Xu J, Gan L, Cheng M, Wu Q. Unsupervised medical entity recognition and linking in Chinese online medical text. J Healthc Eng 2018;2018:2548537 [FREE Full text] [doi: 10.1155/2018/2548537] [Medline: 29849994]
36. Shimoyama M. Novel Unsupervised Named Entity Recognition Used in Text Annotation Tool (OntoMate) At Rat Genome Database. 2017 Presented at: ACM International Conference on Bioinformatics, Computational Biology,Health Informatics; 20-23 August, 2017; Boston, MA, USA p. 590-590.

Abbreviations

BOC: bag-of-characters
CCKS: China Conference on Knowledge Graph and Semantic Computing
CNER: clinical named entity recognition
CNN: Convolutional Neural Network
CRF: conditional random fields
CT: character types
CUMLS: Chinese Unified Medical Language System
EHRs: electronic health records
ICTCLAS: Institute of Computing Technology, Chinese Lexical Analysis System
LSTM: long short-term memory
ME: maximum entropy
ML: machine learning
NER: named entity recognition
POCIS: position of the character in the sentence
POS: part-of-speech
RNN: Recurrent Neural Networks
SSVM: structural support vector machines
SVM: support vector machines

Edited by G Eysenbach; submitted 30.01.18; peer-reviewed by H Xu, J Lei, Y Wang; comments to author 28.06.18; revised version received 30.09.18; accepted 27.10.18; published 17.12.18.

Please cite as:
Zhang Y, Wang X, Hou Z, Li J
Clinical Named Entity Recognition From Chinese Electronic Health Records via Machine Learning Methods
JMIR Med Inform 2018;6(4):e50
URL: http://medinform.jmir.org/2018/4/e50/
doi:10.2196/medinform.9965
PMID:30559093

XSL•FO
RenderX

©Yu Zhang, Xuwen Wang, Zhen Hou, Jiao Li. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 17.12.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of a Multimedia Patient Decision Aid to Supplement the Informed Consent Process of a Peripherally Inserted Central Venous Catheter Procedure: Pre-Post Quasi-Experimental Study

Azizeh K Sowan[1], PhD, RN, MSDA; Arlienita R Beraya[2], MSN, RN, CNL; Adrian Carrola[2], RN, VA-BC; Charles Reed[2], PhD, RN, CNRN

[1]School of Nursing, University of Texas Health at San Antonio, San Antonio, TX, United States

[2]University Health System, San Antonio, TX, United States

Corresponding Author:
Azizeh K Sowan, PhD, RN, MSDA
School of Nursing
University of Texas Health at San Antonio
Suite 2.628
7703 Floyd Curl Drive
San Antonio, TX, 78229
United States
Phone: 1 210 567 5799
Fax: 1 210 567 1719
Email: sowan@uthscsa.edu

Abstract

Background: Informed consent is a complex process to help patients engage in care processes and reach the best treatment decisions. There are many limitations to the conventional consent process that is based on oral discussion of information related to treatment procedures by the health care provider. A conclusive body of research supports the effectiveness of multimedia patient decision aids (PtDAs) in the consent process in terms of patient satisfaction, increased knowledge about the procedure, reduced anxiety level, and higher engagement in the decision making. Little information is available about the effectiveness of multimedia PtDAs in the consent process of invasive therapeutic procedures such as the peripherally inserted central venous catheter (PICC).

Objective: The objective of this study was to examine the effectiveness of a multimedia PtDA in supplementing the consent process of the PICC for patients in 10 acute and intensive care units in terms of knowledge recall, knowledge retention, satisfaction with the consent process, and satisfaction with the PICC multimedia PtDA.

Methods: This pre-post quasi-experimental study included 130 patients for whom a PICC was ordered. Patients in the control group (n=65) received the conventional consent process for the PICC, while those in the intervention group (n=65) received the multimedia PtDA to support the consent process of a PICC. All patients were surveyed for knowledge recall and retention about the procedure and satisfaction with the consent process. Patients in the intervention group were also surveyed for their satisfaction with the multimedia PtDA.

Results: Compared with the control group, patients in the intervention group scored around 2 points higher on knowledge recall (t_{125}=4.9, P<.001) and knowledge retention (t_{126}=4.8, P<.001). All patients in the intervention group were highly satisfied with the multimedia PtDA, with a mean score of >4.5 out of 5 on all items. Items with the highest mean scores were related to the effect of the multimedia PtDA on knowledge retention (mean 4.9 [SD 0.2]), patient readiness to learn (mean 4.8 [SD 0.5]), complete understanding of the procedure-related complications (mean 4.8 [SD 0.4]), and patient role in maintaining the safety of the PICC (mean 4.8 [SD 0.5]). Patients in the two groups were highly satisfied with the consent process. However, 15% (10/65) patients in the control group reported that the following information was omitted from the discussion: patient and provider roles in the safety of the PICC, other treatment options, and common side effects. Furthermore, 2 patients commented that they were not ready to engage in the discussion.

Conclusions: The multimedia PtDA is an effective standardized, structured, self-paced learning tool to supplement the consent process of the PICC and improve patient satisfaction with the process, knowledge recall, and knowledge retention.

XSL•FO
RenderX

(JMIR Med Inform 2018;6(4):e11056) doi:10.2196/11056

KEYWORDS

multimedia; central venous catheters; decision support techniques; informed consent; intensive care units

Introduction

Informed consent is a complex process that aims to discuss with patients complete, clear, and easy-to-understand information about a medical procedure, its treatment indications, associated risks and benefits, and other treatment alternatives using a two-way communication with a teach-back mechanism. Signing the consent form by the patient does not always indicate the process was informed and does not cover the health care team members from liability or protect them against lawsuits. There are numerous challenges related to informed consent, and they may be patient related (ie, special situations such as pain and discomfort, health literacy issues), clinician related (ie, concerns of providing extra information), workflow related (ie, time pressure and workload), organizational culture related (ie, lack of understanding the implications of the informed consent, lack of clear policies and procedures), or resources related (ie, lack of appropriate patient educational tools and decision aids to supplement the consent process).

The consent process is traditionally based on oral discussion of information related to treatment procedures by the health care provider. The effectiveness of this method is questionable given its unstructured and unstandardized nature and the differences in patients' information needs [1,2]. To improve the conventional consent process, multimedia patient decision aids (PtDAs) have been heavily utilized as supplemental educational tools. A conclusive body of research supports the effectiveness of multimedia PtDAs in the consent process for patients undergoing surgeries in terms of higher patient satisfaction with the consent process, increased knowledge about surgeries, reduced anxiety level, and higher patient engagement in the decision-making process [3-10]. On the other hand, little information is available about the effectiveness of multimedia PtDAs in the consent process of invasive therapeutic procedures such as the peripherally inserted central venous catheter (PICC) [11]. Unlike elective surgeries and other medical procedures, when a PICC is ordered, it is often required for patients' condition with little to no other treatment options. The consent process in this case should focus on discussing the procedure with the patient regarding its indications and associated risks, understanding the patient's cultural values that may affect patient acceptance of the procedure, and highlighting the patient role in the safety of the PICC. In this study, we aimed to examine the effectiveness of a multimedia PtDA to supplement the consent process of the PICC for patients in acute care units (ACUs) and intensive care units (ICUs).

PICC is a very common invasive procedure for prolonged administration of medications, parenteral fluids, and blood products. On the other hand, this procedure is a major source for central line-associated bloodstream infections (CLABSIs) [12], central catheter-associated thrombosis [13-15], and other major life-threatening complications, including death. Therefore, patients should play an integral role in maintaining the safety of a PICC and minimizing its complications when competent.

In a previous study, we described the process of developing and "alpha" testing a multimedia PtDA for the PICC following a multidisciplinary, patient-centered, and systematic process [16]. The process was based on the Agency for Healthcare Research and Quality's (AHRQ) Guide for Making Informed Consent an Informed Choice [17], the AHRQ's Health Literacy Universal Precautions Toolkit Guide [18], and the AHRQ's Patient Education Materials Assessment Tool Guide for Audio/Video Materials [19]. The PtDA was evaluated using the PtDAs' quality criteria developed by the International Collaboration for PtDA Standards [20]. The PtDA multimedia program was delivered via the Interactive Patient Care solution GetWell Inpatient, which allows competent patients and family members to review the program as many times as they want from the patient room and when they are ready to do so. The PtDA included information about the purpose of engaging patients and family members in care processes, a disclaimer, and PICC-related content that includes a definition of the PICC, indications, possible insertion sites, other treatment options, risks and complications associated with a PICC, steps of the procedure (before, during, and after a PICC), patient and health care team roles in the care and safety of a PICC during hospitalization, expected time period for having the catheter, safety issues when a patient leaves the hospital with a PICC, and a conclusion. This study describes a "beta" testing of the PICC multimedia PtDA to enhance our understanding of the effectiveness of this tool in practice.

Methods

Design, Setting, and Sample

This pre-post quasi-experimental study was conducted in a university teaching hospital in the Southwest of the United States after obtaining the institutional review board approval. The study included 130 patients from all inpatient units where a PICC was ordered. These include 5 ACUs (2 medical-surgical and 3 surgical), 1 hematology and oncology unit, and 4 ICUs (3 surgical and 1 medical). The preintervention phase included a convenient sample of 65 patients who received the conventional consent process for the PICC, and the post phase included 65 patients who received the multimedia PtDA to support the consent process of a PICC. Only competent patients (ie, those with absence of dementia) were included. Patients were excluded if they had a PICC before this hospitalization, were health care professionals, or had a current diagnosis of depression or anxiety disorders. The sample size was determined on the basis of a desirable improvement in knowledge retention (one of the study outcomes as described below) about the PICC by at least 50%, equal number of patients in the 2 groups, an alpha level of .05, and a beta level of .80. Given these conditions, the enrollment of at least 65 patients in each group should be achieved.

XSL•FO
RenderX

Description of the Consent Process

The PICC consent process starts with the provider who discusses the need for a PICC with the patient and places an order in the electronic medical record (EMR). In the conventional process, a nurse from the vascular access team visits the patient to discuss the procedure details and obtain the consent form. After the discussion and answering the patient's questions, the patient signs a Web-based iMedConsent form for a PICC using an iPad. A certified interpreter is consulted for non-English-speaking patients.

After the creation of the multimedia PtDA for the PICC, the PtDA was integrated into the EMR and delivered to patients using GetWell Inpatient [16]. GetWell Inpatient is a patient education and entertainment system that is integrated into the EMR and used to distribute and track the use of educational videos and multimedia programs in addition to other purposes related to engaging patients in care processes [16]. The use of the PICC multimedia PtDA to supplement the PICC consent process resulted in workflow redesign, as described previously [16]. One of the major changes was related to the new coordination between bedside nurses from the ACUs and ICUs and nurses from the vascular access team. Bedside nurses from the ACUs and ICUs were tasked to make sure that the patient watched the PICC multimedia PtDA within a reasonable timeframe after placing the PICC order in the EMR and before the vascular access team visits the patient for the teach-back and answering questions. The multimedia PtDA was developed in English and Spanish languages [16]. The version watched by the patient was based on his or her preferred language.

As part of standardizing the process during the alpha testing of the product [16], we trained all 6 nurses from the vascular access team and all bedside nurses from the ACUs and ICUs on the new workflow and observed 12 consent processes (2 per nurse from the vascular access team) to ensure adherence to the new workflow. In addition, 2 ICU nurse educators observed the 12 consent processes to ensure that nurses follow a standardized consent process. As mentioned before, helping a patient watch the program was assigned to bedside nurses. Based on the observations' results, all eligible patients watched the PtDA to learn about the procedure before signing the consent form, and family members also watched the program when they were available with patients. The only issue revealed by the observations was lack of clarity about who was supposed to ensure that the patient watched the video (bedside nurses versus vascular access team). Thus, further training was provided, and roles and responsibilities in the new workflow were emphasized.

Measurement and Instrumentation

The main study outcomes were knowledge recall, knowledge retention, and patient satisfaction with the consent process. These are the same outcomes examined in the alpha testing process of the PICC multimedia PtDA [16]. The alpha testing study provided a detailed description of the instruments used to measure these outcomes and instruments' validation processes [16]. In this study, we also assessed patient satisfaction with the multimedia PtDA for the intervention group.

In summary, Patient Knowledge Recall about the PICC Procedure Survey included 19 multiple choice and true or false questions and was developed on the basis of recent PICC clinical practice guidelines [16] and patients' information needs about the procedure indications, benefits, contraindications, insertion site, complications with their probabilities (less common, common, and rare risk factors), and patient and health care team roles in the care and safety of a PICC [16]. The same survey was used to measure knowledge retention.

Patient Satisfaction with the PICC Informed Consent Process Survey was created on the basis of the AHRQ's Guide for Informed Consent; it included 10 items with a 5-point Likert-type scale of agreement [16]. The survey also asked patients about their overall satisfaction with the informed consent process using a 5-point Likert-type scale that ranged from 5 (very satisfied) to 1 (very unsatisfied). For the 2 study groups, the survey measured patient satisfaction with the information provided by the vascular access team. For the intervention group, this discussion took place after watching the multimedia PtDA for the PICC.

Patient Satisfaction with the Multimedia PtDA for the PICC Survey was created on the basis of the AHRQ's Patient Education Materials Assessment Tool Guide for Audio and Video Materials and the Criteria for Effective Patient and Consumer Education Materials [19,21] and included 14 items of a 5-point Likert-type scale of agreement. The survey was followed by a question that measured overall patient satisfaction with the PtDA (a 5-point scale) and 3 other questions that measured the number of times the patient watched the PtDA, whether the patient thinks he or she will watch the program again later, and additional comments. In all surveys, we used the term "video" instead of "multimedia PtDA" to promote patients' understanding of the items because the term "multimedia" is not frequently used by the public.

Study Procedure

Four nurse educators from the ACUs and ICUs administered the surveys. During the study period and at the beginning of every working shift, the vascular access team provided the nurse educators with a list of all PICC orders. The list included the patient name, medical record number, room number, and the unit. After obtaining the consent form, the vascular access team called the nurse educators informing them about the time they obtained the consent form. The nurse educators approached the patients on the list who completed signing the consent form, discussed the study purpose, and emphasized voluntarily participation and patients' rights to withdraw from the study at any point. Owing to a total of 4 surveys in this study to measure the main study outcomes, the nurse educators administered 2 surveys at a time for each patient to improve the response rate and decrease the burden on our patients. For example, Knowledge Recall about the PICC Procedure and Patient Satisfaction with the Multimedia Program surveys were administered 4-8 hours after obtaining the consent form, while Knowledge Retention and Patient Satisfaction with PICC Informed Consent Process surveys were administered 24-36 hours after obtaining the consent form. Overall, 6-7 surveys were collected from each of the 10 inpatient units where a PICC

XSL•FO
RenderX

was ordered to reach the desired sample size for each study phase. In addition, the medical record number was used to connect patients' responses on all surveys. Furthermore, we collected the following patient demographic data from the EMR and patients when the information was not available in the EMR: gender, age, race, education level, preferred language, time of admission, date of admission, and the unit.

The surveys were administered in the English language to all patients, except for 2 patients who indicated Spanish as their preferred language. For these 2 patients, a certified interpreter translated the survey questions.

Results

Patient Characteristics

All patients provided complete responses to all questionnaires, except for 2 patients in the postintervention phase who were excluded from the analysis. We observed no differences in patient characteristics between the control and intervention groups based on chi-square statistic (P>.99; Table 1). Patient age ranged from 24 to 94 years in the control group and 23 to 77 years in the intervention group, and they had similar mean age (mean 51.7 [SD 17.1] vs mean 51.9 [SD 14.3]; P=.90). The majority of patients were white and Hispanic individuals (Table 1). Other races included black (4 in the control group and 4 in the intervention group), American or Alaskan Native (1 in the intervention group), and Asian (1 in the control group). All patients watched the English version of the PICC multimedia PtDA, except for 2 patients who watched the Spanish version (Table 1).

Knowledge Recall and Knowledge Retention

Patients' scores on knowledge recall and knowledge retention ranged from 6 to 18 in the preintervention phase and 9 to 19 in the postintervention phase (out of 19 points). Although knowledge retention was higher than recall at the group level, in general, the 2 study groups scored above average on both surveys (Figure 1). In comparison to the control group, the intervention group scored around 2 points higher on knowledge recall (t_{125}=4.9, P<.001) and knowledge retention (t_{126}=4.8, P<.001; Figure 1).

Table 1. Patient characteristics (N=128).

Characteristic	Control group (n=65)	Intervention group (n=63)	*P* value
Age in years, n (%)			**.41**
<30	11 (17)	5 (8)	
30-50	22 (34)	22 (35)	
51-65	18 (28)	23 (37)	
>65	14 (22)	13 (21)	
Male, n (%)	28 (43)	34 (54)	.20
Race, n (%)			**.92**
Hispanic	30 (46)	25 (40)	
White non-Hispanic	30 (46)	33 (52)	
Others	5 (8)	5 (8)	
Level of education[a], n (%)			**.07**
Illiterate	1 (2)	1 (2)	
Primary education to less than high school	4 (6)	7 (11)	
High school	38 (58)	40 (63)	
College or bachelor	21 (32)	11 (17)	
Graduate	1 (2)	4 (6)	
Language of consent[b], n (%)			**N/A[c]**
Spanish	1 (2)	1 (2)	
English	64 (98)	62 (98)	

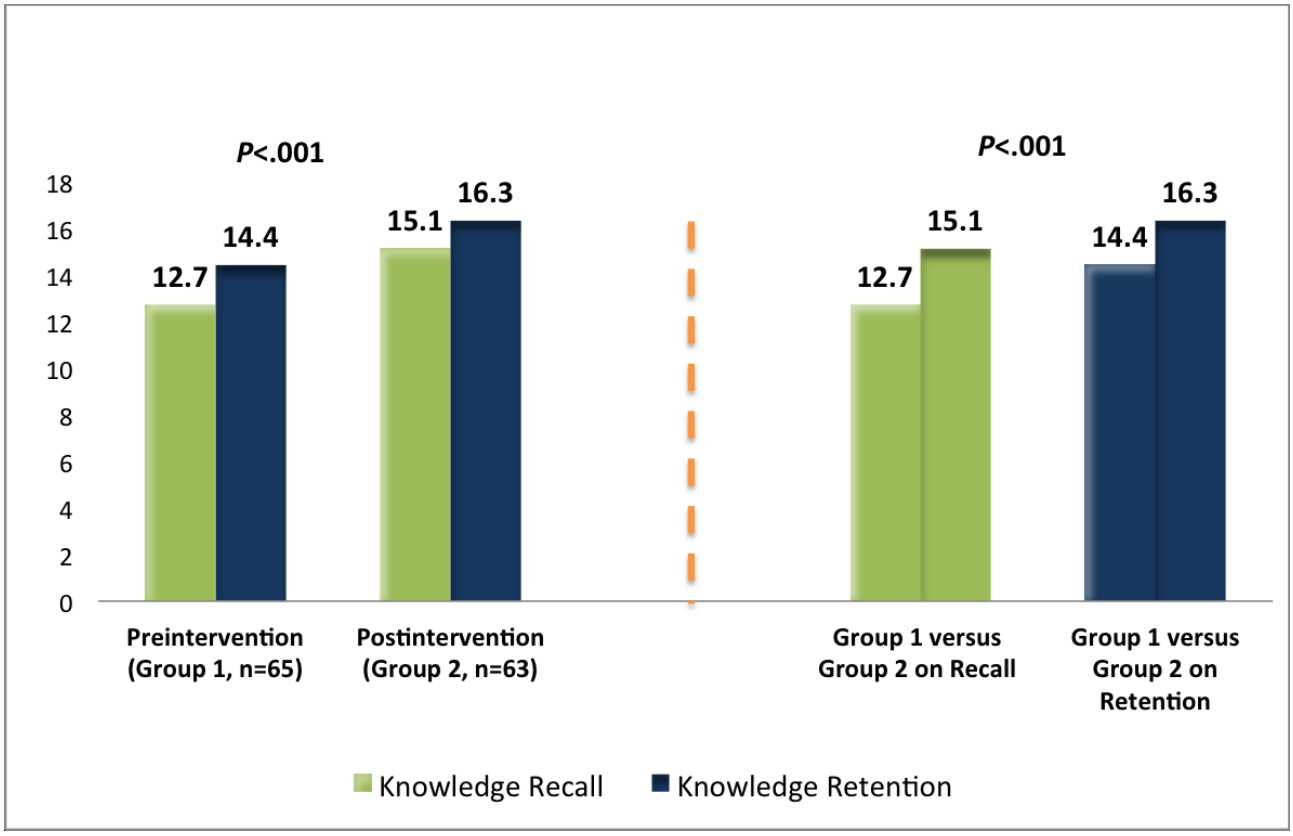
[a]The 3 groups used to perform the chi-square test were high school, college or bachelor, and others. The "others" included all other groups under the level of education because of the small sample size under those categories.

[b]The chi-square test was not performed because of the obvious lack of significance between the categories related to similar frequencies under each category and the small cell size under Spanish.

[c]N/A: not applicable.

XSL•FO
RenderX

Figure 1. Mean difference of correctly answered questions in knowledge recall and retention surveys (19 questions) for the control (n=65 patients) and intervention (n=63 patients) groups.



Patient Satisfaction With the Multimedia Program

Table 2 presents the mean agreement (patients who reported agree or strongly agree) of patient satisfaction with the multimedia PtDA for the PICC procedure for patients in the intervention group. All patients were highly satisfied with the multimedia PtDA with a mean score of >4.5 out of 5 on all items. Items with the highest mean scores were related to the effect of the multimedia PtDA on knowledge retention, patient readiness to learn, and complete understanding of procedure complications and patient role in maintaining the safety of the PICC. Of all patients, 17% (11/63) watched the PtDA program twice, 83% (52/63) watched it once, and 57% (36/63) mentioned that they would like to watch it again. The overall mean patient satisfaction with the multimedia PtDA was 4.6 (SD 0.6).

Patient Satisfaction With the Informed Consent Process

Table 3 presents the mean agreement (patients who reported agree or strongly agree) of patient satisfaction with the informed consent process for the control and intervention groups at the item level. Patients in the control group rated their satisfaction with the discussion between them and the nurse from the vascular access team. Patients in the intervention group rated their satisfaction with the discussion between them and the nurse from the vascular access team after they watched the PtDA program. Patients in the 2 groups were highly satisfied with the process with a reported mean of 4.5-5 out of 5 points on each item. No significant differences at the item level were found between the 2 groups (P>.99).

In this study, 15% (10/65) patients in the control group reported that the following were omitted from the discussion: patient role in the safety of the PICC (5 patients), provider role in the safety of the PICC (4 patients), other treatment options (4 patients), and common side effects (6 patients). Furthermore, 2 patients commented that they were not ready to engage in the discussion.

Overall, patient satisfaction with the process was high for both the groups (mean 4.8 [SD 0.6] versus mean 4.8 [SD 0.5] for the control and intervention groups, respectively, P>.99).

XSL•FO
RenderX

Table 2. Satisfaction with the multimedia decision aid program for the peripherally inserted central venous catheter (PICC) for the intervention group (n=63 patients).

Item	Mean (SD)
1. The video better helped me remember the information about this procedure	4.9 (0.2)
2. The video allows me to listen to information when I am ready to do so	4.8 (0.5)
3. After watching the video, I completely understand the common complications of this procedure and know when to report them	4.8 (0.4)
4. After watching the video, I understand my role as a patient in maintaining the safety of the PICC line	4.8 (0.5)
5. The information in the video was comprehensive to include the following: • Reasons for PICC • Steps of the procedure • Common side effects • Other treatment options • Definition of PICC • Patient role in care and safety of PICC • Provider role in care and safety of PICC	4.7 (0.6)
6. There was almost no disruption during watching the video	4.7 (0.8)
7. Visual aids (eg, showing the PICC line) in the video were helpful	4.7 (0.6)
8. The video was very beneficial to learn about the procedure	4.7 (0.6)
9. The video allows me to listen to information as many times as I need	4.7 (0.7)
10. The information in the video was clear	4.6 (0.7)
11. The information in the video was easy to understand	4.6 (0.8)
12. Speed of presenting the information in the video was reasonable	4.6 (0.8)
13. I highly recommend this video to supplement the consent process for PICC	4.6 (0.8)
14. I feel the video decreased my level of anxiety	4.6 (1.0)

Table 3. Patient satisfaction with the informed consent process (N=128).

Item	Mean (SD)	
	Control group (n=65)	Intervention group (n=63)
1. Provider attitude during the discussion was positive	5 (0)	4.9 (0.1)
2. Speed of information provided was reasonable	4.9 (0.4)	4.9 (0.1)
3. Disruption during the discussion was minimal	4.8 (0.5)	4.8 (0.6)
4. I completely understand the common complications of this procedure and know when to report them	4.8 (0.5)	4.9 (0.4)
5. The information provided by the nurse was clear	4.7 (0.3)	4.9 (0.3)
6. The information provided was easy to understand	4.7 (0.9)	4.8 (0.4)
7. Timing of the discussion was convenient	4.7 (0.6)	4.9 (0.4)
8. I understand my role as a patient in maintaining the safety of the peripherally inserted central catheter (PICC)	4.6 (1)	4.9 (0.4)
9. The nurse answered all of my questions sufficiently	4.6 (0.7)	4.9 (0.3)
10. The information provided was comprehensive to include the following: • Definition of the PICC • Reasons for the PICC • Steps of the procedure • Common side effects • Other treatment options • Patient role in care and safety of the PICC • Provider role in care and safety of the PICC	4.5 (0.8)	N/A[a]

[a]N/A: not applicable.

XSL•FO
RenderX

Discussion

Principal Findings

This study examined the effectiveness of a multimedia PtDA to supplement the consent process of a PICC for patients in the ACUs and ICUs on knowledge recall, knowledge retention, and patient satisfaction with the informed consent process and satisfaction with the PtDA. Patient-centered care and shared decision-making models are based on empowering patients with sufficient, clear, and easy-to-understand information about treatment procedures. Complex care environments are one of the barriers to effective informed consent process even for nonurgent procedures such as a PICC due to time pressure, workload, complexity of medical conditions, special patients' circumstances, and diverse patients' needs. Research studies support the limitations of the conventional informed consent process where providers do not allocate sufficient time to discuss procedures' details or provide a meaningful dialogue [1,2]. Failure to apply these elements in a consent process would result in signing a piece of paper for record keeping rather than informing and engaging patients in care processes. Furthermore, well-designed multimedia PtDAs are effective tools to supplement the consent process of medical procedures in complex care environments.

In this study, patients in the control and intervention groups had a higher level of knowledge retention about the procedure in comparison with knowledge recall; this can be related to the fact that when nurse educators collected the data from patients, although they recorded patients' answers (whether right or wrong), they corrected the patients' misunderstanding about the procedure indications, complications, side effects, and patient and provider roles in safety of the procedure when patients provided wrong answers. Nevertheless, patients in the intervention group significantly achieved a higher level of knowledge recall and knowledge retention, supporting the effectiveness of this tool in supplementing the consent process. In addition, patients in both the groups reported high levels of satisfaction with the informed consent process. Patient satisfaction with the multimedia PtDA for patients in the intervention group was overwhelmingly positive.

Some patients in the control group indicated that the consent process missed discussing different aspects of the procedure such as patient and provider roles in the safety of the procedure, common side effects and complications, and other treatment options. In a previous study, patients reported missing information in the consent process when, in fact, this information was discussed by the health care team [16]. This supports the need for a self-paced resource, such as a multimedia PtDA, accessible to patients when needed to review the information discussed with them.

Our multimedia PtDA for the PICC provided complete information about the procedure. In addition, the PtDA focused on the patient role in the safety of the procedure, an area that is often ignored in consent processes. All surveys also included items that reflect patient role in the safety of the procedure. Although our patients in the 2 groups scored high on these items, one of our patients in the intervention group stopped a nurse from touching his PICC line and asked her to wash her hands based on what he saw in the video.

Hand hygiene is the number one strategy to prevent CLABSIs. CLABSIs are ranked the most common and the most costly hospital-acquired infections (account for US $46,000 per event [22]), and they result in thousands of deaths each year [23]. CLABSIs are "never event" that should never happen if appropriate preventable measures are in place such as hand hygiene and empowering patients to understand and immediately report side effects and complications. In the future, it would be interesting to examine the effect of the multimedia PtDA on CLABSI incidence rate.

The multimedia program was meant to be a self-paced learning tool that takes into consideration patient readiness to engage in care processes. Some of our patients in the control group mentioned that they were not ready for the discussion. Readiness to learn was ranked the second highest item by the intervention group in the patient satisfaction with the multimedia program survey. Many patients in the intervention group watched the video for more than once and indicated their interest to watch it again. The availability of such resource to the patients is necessary to reinforce knowledge and engagement in care processes.

Limitations

This study has the following 2 limitations. First, although the ratios of our Hispanic and white patients were similar to the patient populations we see at the hospital level, almost all of our Hispanic patients who participated in this study indicated English as their preferred language for the discussion and the PtDA. In future research, we need to focus on engaging a larger sample of Hispanic population who would use the Spanish version of the program to test whether the Spanish version has the same positive effect as the English version. Second, the nurse educators who served as data collectors were instructed to correct the wrong responses provided by patients in the knowledge recall and retention questionnaires after recording the original patient responses. Although this reflects ethical and professional practice principles to maintain safety, it could have biased our knowledge retention scores. Nevertheless, the effectiveness of the PtDA was supported by other results such as (1) the significantly higher knowledge recall scores for the intervention group than for the control group; (2) the high agreement scores assigned to the "readiness to learn," "complete understanding of procedure complications," and "complete understanding of the patient role in maintaining the safety of the PICC" items by the intervention group in the patient satisfaction with the multimedia program survey; and (3) the availability of the PtDA as a self-paced learning tool supported by watching the video for more than two times by some patients.

Conclusions

This study reveals that the multimedia PtDA is an effective standardized, structured, and self-paced learning tool to supplement the consent process of the PICC and improve patient satisfaction with the process, knowledge recall, and knowledge retention.

XSL•FO
RenderX

Acknowledgments

This project was funded by the University Health System's Center for Clinical Excellence and the University of Texas Health School of Nursing. The role of the funding sources was limited to providing financial support for the conduct of the research. Sponsors were not involved in study design, collection, analysis or interpretation of data, writing of the manuscript, or the decision to submit the paper for publication.

Conflicts of Interest

None declared.

References

1. McGaughey I. Informed consent and knee arthroscopies: an evaluation of patient understanding and satisfaction. Knee 2004 Jun;11(3):237-242. [doi: 10.1016/S0968-0160(03)00107-8] [Medline: 15194102]
2. Fink AS, Prochazka AV, Henderson WG, Bartenfeld D, Nyirenda C, Webb A, et al. Enhancement of surgical informed consent by addition of repeat back: a multicenter, randomized controlled clinical trial. Ann Surg 2010 Jul;252(1):27-36. [doi: 10.1097/SLA.0b013e3181e3ec61] [Medline: 20562609]
3. Wollinger C, Hirnschall N, Findl O. Computer-based tutorial to enhance the quality and efficiency of the informed-consent process for cataract surgery. J Cataract Refract Surg 2012 Apr;38(4):655-659. [doi: 10.1016/j.jcrs.2011.10.038] [Medline: 22440436]
4. Ihrig A, Herzog W, Huber CG, Hadaschik B, Pahernik S, Hohenfellner M, et al. Multimedia support in preoperative patient education for radical prostatectomy: the physicians' point of view. Patient Educ Couns 2012 May;87(2):239-242. [doi: 10.1016/j.pec.2011.08.014] [Medline: 21945598]
5. Tipotsch-Maca SM, Varsits RM, Ginzel C, Vecsei-Marlovits PV. Effect of a multimedia-assisted informed consent procedure on the information gain, satisfaction, and anxiety of cataract surgery patients. J Cataract Refract Surg 2016 Jan;42(1):110-116. [doi: 10.1016/j.jcrs.2015.08.019] [Medline: 26948785]
6. Batuyong ED, Jowett AJL, Wickramasinghe N, Beischer AD. Using multimedia to enhance the consent process for bunion correction surgery. ANZ J Surg 2014 Apr;84(4):249-254. [Medline: 24812709]
7. Hoppe DJ, Denkers M, Hoppe FM, Wong IH. The use of video before arthroscopic shoulder surgery to enhance patient recall and satisfaction: a randomized-controlled study. J Shoulder Elbow Surg 2014 Jun;23(6):e134-e139. [doi: 10.1016/j.jse.2013.09.008] [Medline: 24295838]
8. Lin Y, Chen C, Lee W, Cheng Y, Lin T, Lin C, et al. Educational video-assisted versus conventional informed consent for trauma-related debridement surgery: a parallel group randomized controlled trial. BMC Med Ethics 2018 Mar 09;19(1):23 [FREE Full text] [doi: 10.1186/s12910-018-0264-7] [Medline: 29523129]
9. Cornoiu A, Beischer AD, Donnan L, Graves S, de Steiger R. Multimedia patient education to assist the informed consent process for knee arthroscopy. ANZ J Surg 2011 Mar;81(3):176-180. [doi: 10.1111/j.1445-2197.2010.05487.x] [Medline: 21342392]
10. Brandel MG, Reid CM, Parmeshwar N, Dobke MK, Gosman AA. Efficacy of a Procedure-Specific Education Module on Informed Consent in Plastic Surgery. Ann Plast Surg 2017 May;78(5 Suppl 4):S225-S228. [doi: 10.1097/SAP.0000000000000970] [Medline: 28118228]
11. Bowers N, Eisenberg E, Montbriand J, Jaskolka J, Roche-Nagle G. Using a multimedia presentation to improve patient understanding and satisfaction with informed consent for minimally invasive vascular procedures. Surgeon 2017 Feb;15(1):7-11. [doi: 10.1016/j.surge.2015.09.001] [Medline: 26464072]
12. Herc E, Patel P, Washer LL, Conlon A, Flanders SA, Chopra V. A Model to Predict Central-Line-Associated Bloodstream Infection Among Patients With Peripherally Inserted Central Catheters: The MPC Score. Infect Control Hosp Epidemiol 2017 Dec;38(10):1155-1166. [doi: 10.1017/ice.2017.167] [Medline: 28807074]
13. Chopra V, Kaatz S, Conlon A, Paje D, Grant PJ, Rogers MAM, et al. The Michigan Risk Score to predict peripherally inserted central catheter-associated thrombosis. J Thromb Haemost 2017 Dec;15(10):1951-1962. [doi: 10.1111/jth.13794] [Medline: 28796444]
14. Grau D, Clarivet B, Lotthé A, Bommart S, Parer S. Complications with peripherally inserted central catheters (PICCs) used in hospitalized patients and outpatients: a prospective cohort study. Antimicrob Resist Infect Control 2017;6:18 [FREE Full text] [doi: 10.1186/s13756-016-0161-0] [Medline: 28149507]
15. Liu Y, Gao Y, Wei L, Chen W, Ma X, Song L. Peripherally inserted central catheter thrombosis incidence and risk factors in cancer patients: a double-center prospective investigation. Ther Clin Risk Manag 2015;11:153-160 [FREE Full text] [doi: 10.2147/TCRM.S73379] [Medline: 25673995]
16. Sowan A, BerayaA, Carrola A, Reed C, Matthews S, Moodley T. Developing, implementing, and evaluating a multimedia decision aid program to reform the informed consent process of a peripherally inserted central venous catheter procedure. JMIR Formative Research. (forthcoming/in press). https://preprints.jmir.org/preprint/10709 2018. [doi: 10.2196/10709]

XSL•FO
RenderX

17. Shoemaker S, Brach C. Prepared under Contract No. Implementation Guide for AHRQ's Making Informed Consent an Informed Choice Training Modules URL: https://www.ahrq.gov/professionals/systems/hospital/informedchoice/informed-implementation-guide.html [WebCite Cache ID 73u62MYJc]
18. Brega A, Barnard J, Mabachi N. (Prepared by Colorado Health Outcomes Program, University of Colorado Anschutz Medical Campus under Contract No. AHRQ health literacy universal precautions toolkit, Second Edition URL: https://www.ahrq.gov/sites/default/files/publications/files/healthlittoolkit2_3.pdf [accessed 2018-04-15] [WebCite Cache ID 6yPFoZddF]
19. Shoemaker S, Wolf M, Brach C. (Prepared by Abt Associates, Inc. The Patient Education Materials Assessment Tool (PEMAT) and User's Guide URL: https://www.ahrq.gov/sites/default/files/publications/files/pemat_guide.pdf [accessed 2018-04-16] [WebCite Cache ID 6yPG4D2N9]
20. International Patient Decision Aid Standards (IPDAS) Collaboration. IPDAS 2005: Criteria for judging the quality of patient decision aids URL: http://ipdas.ohri.ca/ipdas_checklist.pdf [accessed 2018-04-03] [WebCite Cache ID 6yPGgIBuU]
21. Elsevier Clinical Solution (2015). How to develop and use effective patient/consumer education URL: https://www.elsevier.com/__data/assets/pdf_file/0006/139902/Elsevier_WP_PatientEducationWhitepaper5_29_15WEB.pdf [accessed 2018-04-17] [WebCite Cache ID 6zR3ziOJZ]
22. Zimlichman E, Henderson D, Tamir O, Franz C, Song P, Yamin CK, et al. Health care-associated infections: a meta-analysis of costs and financial impact on the US health care system. JAMA Intern Med 2013;173(22):2039-2046. [doi: 10.1001/jamainternmed.2013.9763] [Medline: 23999949]
23. Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 2006 Dec 28;355(26):2725-2732. [doi: 10.1056/NEJMoa061115] [Medline: 17192537]

Abbreviations

ACU: acute care unit
AHRQ: Agency for Healthcare Research and Quality
CLABSI: central line-associated bloodstream infection
EMR: electronic medical record
ICU: intensive care unit
PICC: peripherally inserted central venous catheter
PtDA: patient decision aid

Edited by G Eysenbach; submitted 15.05.18; peer-reviewed by K Glover, F Abujarad; comments to author 09.08.18; revised version received 16.08.18; accepted 16.08.18; published 28.11.18.

Please cite as:
Sowan AK, Beraya AR, Carrola A, Reed C
Effect of a Multimedia Patient Decision Aid to Supplement the Informed Consent Process of a Peripherally Inserted Central Venous Catheter Procedure: Pre-Post Quasi-Experimental Study
JMIR Med Inform 2018;6(4):e11056
URL: http://medinform.jmir.org/2018/4/e11056/
doi:10.2196/11056
PMID:30487117

©Azizeh K Sowan, Arlienita R Beraya, Adrian Carrola, Charles Reed. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 28.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Validation and Testing of Fast Healthcare Interoperability Resources Standards Compliance: Data Analysis

Jason Walonoski[1], MS; Robert Scanlon[1], MS; Conor Dowling[1]; Mario Hyland[2]; Richard Ettema[2], BS; Steven Posnack[3], MS, MHS

[1]The MITRE Corporation, Bedford, MA, United States

[2]AEGIS.net, Inc, Rockville, MD, United States

[3]The Office of the National Coordinator for Health Information Technology, US Department of Health and Human Services, Washington, DC, United States

Corresponding Author:
Jason Walonoski, MS
The MITRE Corporation
202 Burlington Road
Bedford, MA, 01730
United States
Phone: 1 781 271 2021
Email: jwalonoski@mitre.org

Abstract

Background: There is wide recognition that the lack of health data interoperability has significant impacts. Traditionally, health data standards are complex and test-driven methods played important roles in achieving interoperability. The Health Level Seven International (HL7) standard Fast Healthcare Interoperability Resources (FHIR) may be a technical solution that aligns with policy, but systems need to be validated and tested.

Objective: Our objective is to explore the question of whether or not the regular use of validation and testing tools improves server compliance with the HL7 FHIR specification.

Methods: We used two independent validation and testing tools, Crucible and Touchstone, and analyzed the usage and result data to determine their impact on server compliance with the HL7 FHIR specification.

Results: The use of validation and testing tools such as Crucible and Touchstone are strongly correlated with increased compliance and "practice makes perfect." Frequent and thorough testing has clear implications for health data interoperability. Additional data analysis reveals trends over time with respect to vendors, use cases, and FHIR versions.

Conclusions: Validation and testing tools can aid in the transition to an interoperable health care infrastructure. Developers that use testing and validation tools tend to produce more compliant FHIR implementations. When it comes to health data interoperability, "practice makes perfect."

(JMIR Med Inform 2018;6(4):e10870) doi:10.2196/10870

KEYWORDS

electronic health records; health data interoperability; test-driven development; practice makes perfect

Introduction

Lack of Health Interoperability

Despite the relatively rapid nationwide adoption of electronic health records (EHRs), the industry's ability to successfully exchange computable health data has not kept pace. A recent study found that less than 35% of providers report data exchange with other providers within the same organization or affiliated hospitals. The exchange of data across organizations is even more limited, with less than 14% of providers reporting they exchange data with providers in other organizations or unaffiliated hospitals [1]. Limited health data interoperability has significant impacts. As part of providing care for their patients, the typical primary care physician (PCP) coordinates care with 229 other physicians across 117 organizations [2]. Limited interoperability makes the already complicated problem of care coordination even more challenging. Currently, 40% of PCPs report that when they refer a patient to a specialist, they do not efficiently receive the outcomes of the visit, including cases where the patient's plan of care or active medications have

XSL•FO
RenderX

changed [3]. The lack of interoperability leads to gaps in critical information at the point of care. This puts undue burden on patients who currently must fill those gaps in data and when those gaps remain unfilled, they can lead to significant safety issues. An inpatient study found that 18% of medical errors leading to adverse drug events could be traced back to missing data in the patient's medical record [4]. Finally, limited health care data interoperability has an immense cost. For example, West Health Institute, an independent, nonprofit medical research organization, estimates that the lack of medical device interoperability alone leads to over US $30 billion in wasteful spending each year. Beyond medical devices, the broader problem of limited health care data interoperability further contributes to an estimated US $700 billion in wasteful spending annually in health care [5].

Improving Interoperability

As indicated by the JASON report [6], the implications and benefits from a truly open digital health care architecture are wide ranging, from enabling individual patients to obtain, share, and authorize who can view their data, to population health analytics and research. Currently, data and exchange standards in health care do not adequately ensure out-of-the-box interoperability, chiefly due to the complexity and lack of identical interpretations of the published standards by health IT software developers. Rigorous testing and validation will help move the US health care system in the direction of open, accessible, patient-centric care.

To date, the health care community has produced and tolerated data standards that are complex, difficult to understand, and technically challenging to consistently implement and test. While well meaning, such standardization efforts have advanced interoperability only so far and, at the same time, stifled innovation due to high custom development and maintenance costs. In one such situation, MITRE has previously demonstrated in the domain of clinical quality measurement that a test-driven approach can successfully establish a framework for interoperability using national health care standards [7]. Similarly, AEGIS.net has successfully supported health information networks that focus on nationwide scale and standards adoption with its Developers Integration Lab cloud-based Test-Driven-Development (TDD) Test Platform [8].

Health Level Seven International Fast Healthcare Interoperability Resources

The Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) standard [9] offers a better and more innovation-friendly path forward. The FHIR standard is designed to be easily and quickly implemented while also, over time, supporting a broad set of health care use cases. It is also a rapidly evolving standard that consists of data formats for health care resources and an application programming interface (API) for the exchange of this information between client applications and servers. This API uses Representational State Transfer web services, a modern technology pattern that powers most of today's Internet.

Although the Office of the National Coordinator for Health Information Technology's (ONC) 2015 Edition Health IT Certification Criteria includes an API certification criterion—45 United States Code of Federal Regulations 170.315(g)(8) and (g)(9)—that is well suited for FHIR implementation, no government regulations require health IT developers to conform to any published version of the FHIR standard [10]. To facilitate this without regulation, a group of commercial health IT developers established the Argonaut Project. Argonauts have committed to promote health interoperability by using FHIR within the industry by defining implementation guides and profiles for read-only data access and document retrieval [11].

Policy Context

In 2015, the ONC published the document, Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap (the Roadmap) [12]. At a high level, the Roadmap expressed that interoperability needed to focus on three overall themes: (1) a supportive payment and regulatory environment, (2) policy and technical components, and (3) outcomes that could be measured and could impact individuals and providers. Section G of the Roadmap laid out the need for an industry-wide testing and certification infrastructure, stating that a "diverse and complementary set of testing and certification programs will need to be in place to achieve nationwide interoperability." Further, with respect to testing, the Roadmap indicated that "the health IT ecosystem will need to invest in more efficient ways to test health IT that is implemented and used among a diverse set of stakeholders."

The 21st Century Cures Act (the Cures Act), Public Law 114-255 [13], was signed into law in late 2016. The Cures Act includes policies that impact everything from medical devices to precision medicine. In the context of health IT, it includes the most substantial update to the ONC's authority since the Health Information Technology for Economic and Clinical Health Act was passed in 2009. Specifically, the Cures Act includes a statutory definition for *interoperability*; it establishes a new federal advisory committee, the Health IT Advisory Committee, it requires the National Coordinator to amend the ONC Health IT Certification Program to adopt *conditions of certification* that are applicable to health IT developers, and it defines *information blocking* and the penalties associated with doing so.

Importantly, and relevant to this paper, the Cures Act includes two provisions within the conditions of certification related to APIs. First, it charges ONC to require that health IT developers publish APIs that can enable health information to be accessed, exchanged, and used "without special effort." Second, it charges ONC with requiring that health IT developers successfully test the real-world use of their certified technology for interoperability in the type of setting in which the technology is marketed. Taken together, these two statutory requirements signal a growing need for the industry to coalesce and invest in API-testing capacity.

Objectives

To meet the requirements expressed within the Cures Act, health IT developers need substantive tools to validate and test system

XSL•FO
RenderX

conformity to the FHIR specification. Furthermore, the consistent implementation of FHIR will help enable an open and innovation-friendly ecosystem that can make data exchange more efficient and reduce interface costs. Both Crucible and Touchstone projects represent production-ready testing platforms for the FHIR specification, which can immediately be leveraged by industry to support their needs for FHIR-based testing [14,15].

Other available testing tools include Sprinkler, an open-source project developed by Firely, that tested FHIR servers with a web-based application [16]. Sprinkler has been retired and has not been publicly updated in four years. Another FHIR test tool available to the community is the Sync for Science Test Suite, which provides similar FHIR testing capabilities, but is geared specifically for the Sync for Science program [17].

The objective of this research was to examine whether or not the use of validation and test tools, specifically Crucible and Touchstone, had any impact on vendor compliance with the FHIR specification and, by extension, interoperability.

Methods

Overview

Two independent projects—MITRE's Crucible project and AEGIS.net's Touchstone project—provide the capability to rigorously test servers against the FHIR specification. Such testing assures health IT developers and app developers that the standards have been consistently implemented and deployed. This kind of testing is essential to enable interoperable health IT solutions that can be used to deliver safer and more efficient health care.

MITRE Crucible Project

Crucible is a set of open-source testing tools for HL7 International FHIR developed by MITRE through an internally funded research program. It is provided as a free and public service to the FHIR development community to promote correct FHIR implementations. Its capabilities include the testing of servers for conformance to the FHIR standard, scoring patient records for completeness, and generating synthetic patient data suitable for testing [14].

The Crucible tool has been used by FHIR developers in the health care information technology industry since 2015. Developers can test their FHIR implementations through the Crucible website [14] by entering a URL at which their server can be accessed via the FHIR API.

There are three ways that Crucible can be used to test server compliance:

1. Server compliance tests may be manually run through the public instance of Crucible.
2. Server compliance tests of known servers are automated to run every 3 days through the public instance of Crucible.
3. Server compliance tests may be run on private instances of Crucible behind a private firewall.

In our analysis, this paper examines test results from manual and automated tests run through the public instance. Only manually run tests are considered as an indicator of system usage. Tests run on private instances of Crucible are not included, as that data is not available to the researchers.

AEGIS.net, Inc Touchstone Project

Touchstone is an open-access platform which combines nearly 20 years of automated lab-based testing initiatives, most recently the cloud-based Test-as-a-Service Developers Integration Lab developed by AEGIS.net through internal research and development. By leveraging the experience gained and lessons learned supporting ONC onboarding participant organizations to the early stages of Nationwide Health Information Network and later hosting the Sequoia Project formal testing program for eHealth Exchange, AEGIS.net has advanced this test platform to address FHIR [15].

Touchstone has successfully been used by developers and quality assurance experts in health care information technology since 2015. Users can privately test their FHIR implementations by navigating to the Touchstone Project site [15] and create an account to support testing FHIR Implementations. Publication of any Touchstone test results only occurs with a developer's prior approval.

In order to test for conformance and interoperability, Touchstone combines the following features in an open-access platform:

1. Testing both client applications and server implementations, while supporting peer-to-peer, multi-actor scenarios (ie, care coordination and workflow) in a unified testing approach.
2. Testing is based entirely on the FHIR Test Script Resource, allowing for crowdsourcing future test case development.
3. Multi-version FHIR support, which facilitates testing backwards compatibility and future-proofing systems and products to ensure a continuously interoperable ecosystem.

To gauge FHIR implementation conformance, this paper examines test results from manual and API-automated tests run through the public cloud instance of Touchstone. Only vendor-initiated tests against the cloud version of Touchstone are considered as an indicator of system usage. Tests run on private instances of Touchstone are not included as that data is not available to the researchers.

Testing Period

The FHIR specification was originally proposed as a new health care data and exchange standard in August 2011. The first official release as a Draft Standard for Trial Use (DSTU) was published on September 30, 2014. Subsequent official releases of the FHIR specification have occurred on a 1.5-2-year balloting cycle. The FHIR specification has rapidly evolved over a short number of years; until initial stabilization of the specification occurred with the release of the DSTU, the introduction of publicly available testing tools was not feasible. To that point, the Crucible and Touchstone platforms only became available starting in 2015 when the test execution results data used in this statistical analysis began to be collected. For this study, data from Crucible ranged from December 1, 2015, to May 31, 2017, and data from Touchstone ranged from September 27, 2015, to September 3, 2017.

XSL•FO
RenderX

Data Collection

Data was collected for this study through the usage of the Crucible and Touchstone projects. During the study period, software developers executed tests using both projects either autonomously or as part of a FHIR Connectathon. Both projects automatically collected usage data on the tests that they execute. This included the following: which FHIR server was under test, the version of FHIR being tested, which tests were being executed and how those tests map to the FHIR specification, the results of each test (eg, pass, fail, skip), as well as step-by-step interactions between the testing system and the target FHIR server (eg, every HTTP request including headers and body and every HTTP response including headers and body), and detailed introspection and checks of those results.

Results

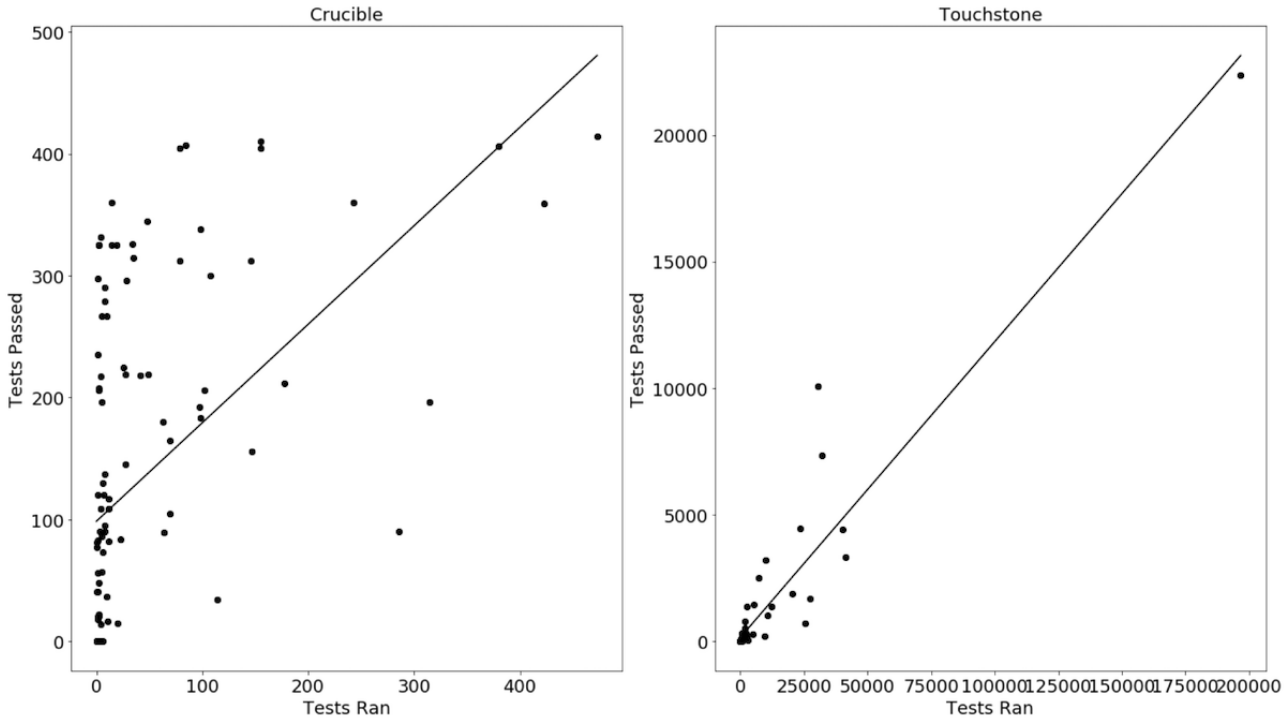
We wanted to know whether or not there was a relationship between testing and compliance. Therefore, we explored whether a statistically significant correlation could be found between the frequency with which vendors execute tests and their conformance with the FHIR specification. For this regression, servers were grouped together by vendor and as many vendors tested FHIR implementations using multiple servers. The number of manual tests executed was used as a measure of an organization's usage level. The number of distinct test suites supported (ie, tests successfully passed) across all the servers was used to measure vendor performance. This metric is a good approximation of the number of features a vendor has implemented successfully and completely.

The number of tests executed were log-normalized to reflect decreasing marginal returns. This is because the most complex test suites tend to be implemented by developers last and require more implementation hours and testing. Regressing log tests executed against the number of supported suites gives a statistically significant (P<.005, n=115) positive correlation between Crucible usage and vendor performance. In other words, using linear regression to predict vendor performance (ie, number of test suites passed), it was found that the number of tests executed (beta=.80, P<.005) was a significant predictor. The model fit was R-squared=.262.

A similar analysis for Touchstone shows a statistically significant (P<.005, n=70) positive correlation between Touchstone usage and vendor performance. In other words, using linear regression to predict vendor performance (ie, number of unique tests passed), it was found that the number of tests executed (beta=0.11, P<.005) were significant predictors. The model fit was R-squared=.883.

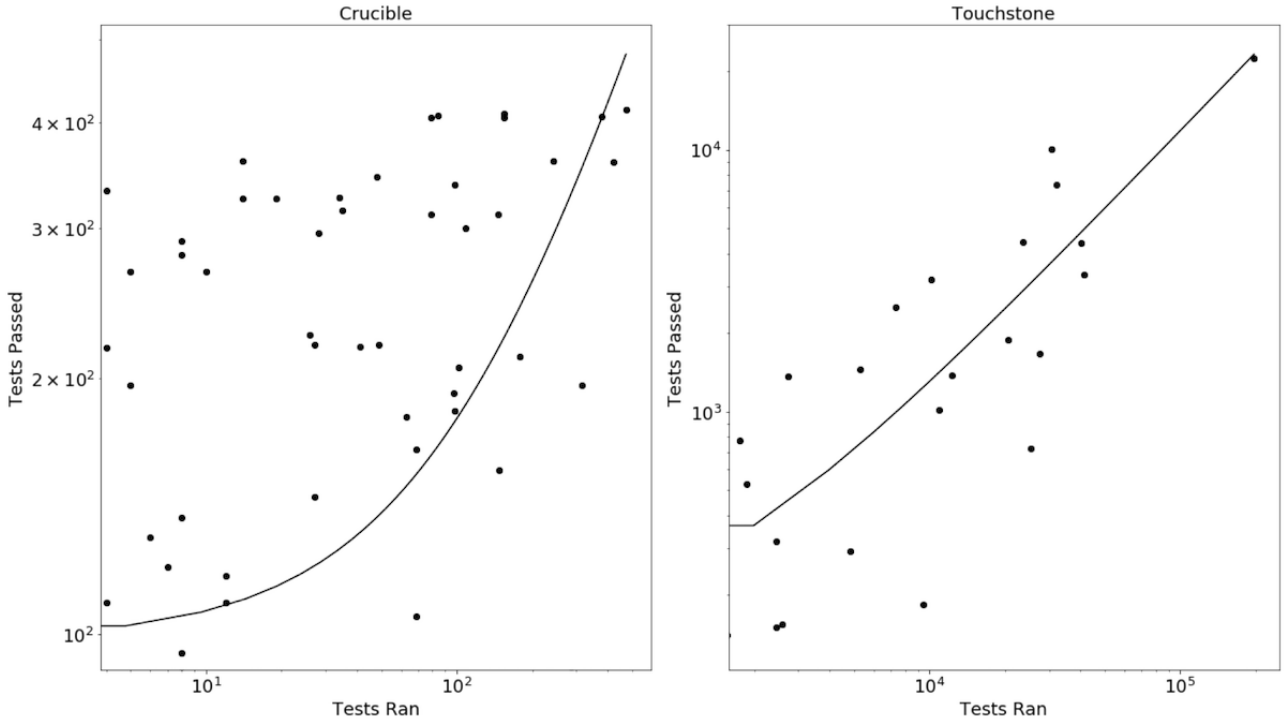
These simple regressions—plotted in Figure 1 in a linear scale and Figure 2 with a log scale—indicate that committed FHIR developers are gaining value from Crucible and Touchstone through repeated use of testing services and incremental improvements of their implementations. In other words, the classic adage "practice makes perfect" unsurprisingly proves true. Test-driven development (*practice*) leads to improved specification adherence (*perfection*).

Figure 1. Predicting suites passed by tests executed.



XSL•FO
RenderX

Figure 2. Predicting suites passed by tests executed (log scale).



Discussion

Principal Findings

The results of our data analysis indicate that as the frequency of testing or number of tests increases, the performance of a server against those tests increases. This should not be surprising as software developers will address issues and fix defects in order to pass the tests, so long as they are discovering these issues and defects by repeated testing. Assuming the tests accurately and adequately cover the depth and breadth of the FHIR specification, then FHIR servers developed and tested using these tests in a test-driven manner should more accurately adhere to the FHIR specification. If compatibility with FHIR equates to health data interoperability, then it seems that fair and neutral testing is critical to achieving that goal. Of course, health data interoperability is vastly more complex than FHIR alone; other factors include, but are not limited to, clinical terminologies, security and trust frameworks, clinical workflow compatibility, and financial incentives. But the correct implementation of software that adheres to the FHIR specification is a good first step to exchanging data.

Tests Over Time

As shown in Figure 3, Crucible has seen use since its launch, with a period of high usage during February 2016, corresponding with increased Argonaut testing and spikes in usage during HL7 Connectathons. Over its lifetime, Crucible has averaged just over 42 test executions per week with testing volume trending upward over time. Touchstone has also seen significant growth in use since its inception as FHIR has grown in popularity and importance.

Tracking Vendors Over Time

Using regular automated testing on known FHIR servers, Crucible can track the progress of a server over time. Because vendors often use temporary server URLs for testing purposes, to track the weekly progress of an individual vendor, we can aggregate the results of all known servers for that vendor and use the best results to track their progress implementing FHIR. Crucible's tracking of one anonymized vendor's Standard for Trial Use version 3 (STU3) servers is shown below in Figure 4. They show a slow improvement before a drastic increase in performance in February 2017.

Similarly, looking specifically at the top active anonymized users of Touchstone—Vendor A (188 uses), who started testing with Touchstone in February 2017, Vendor B (378 uses), Vendor C (321 uses), and Vendor D (207 uses)—there is evidence of both high use of Touchstone and improvement in their FHIR implementations. These implementations used Touchstone consistently during the study period, with their results progressively improving (ie, passing less than 20 tests initially to passing over 1000 tests). Touchstone's TDD testing capabilities allowed these developers to implement their FHIR servers faster by finding errors and confirming the correctness of their implementations, including managing version upgrades.

It is important to note that Vendor A accomplished in 6 weeks what many organizations accomplish in 12-24 weeks, by leveraging TDD—and testing on a daily basis—and integrating continuous testing into their development lifecycle.

XSL•FO
RenderX

Figure 3. Test runs per week.

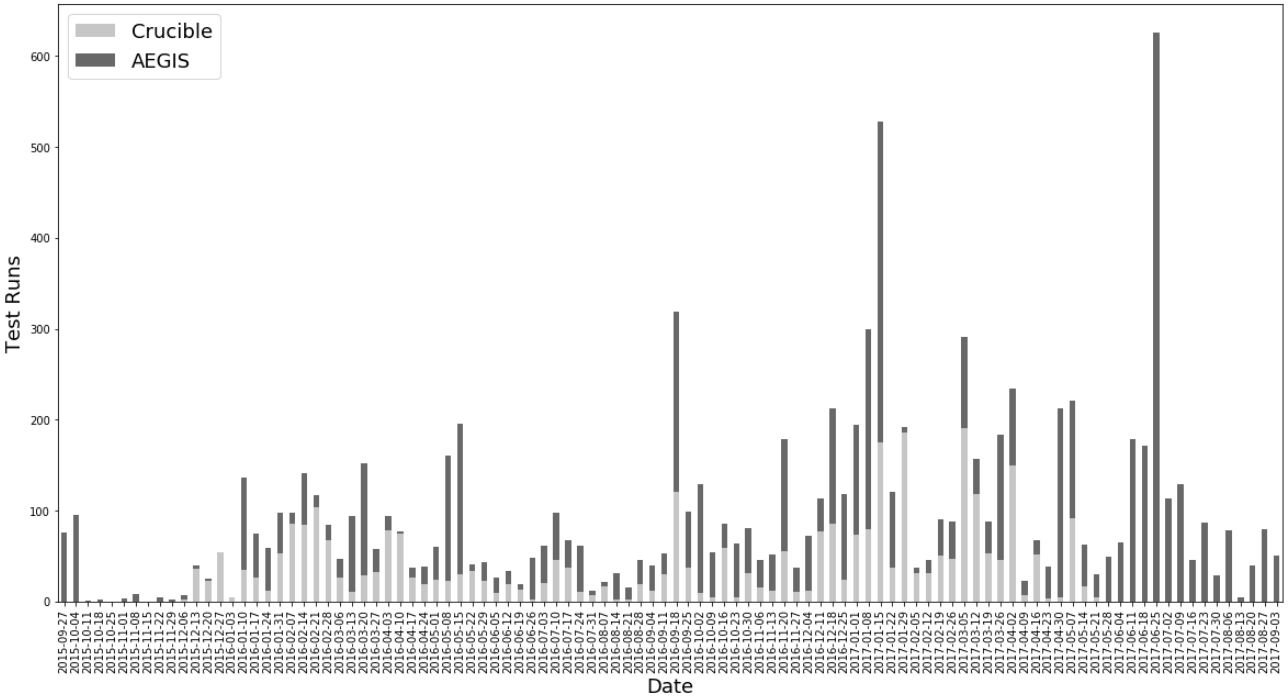
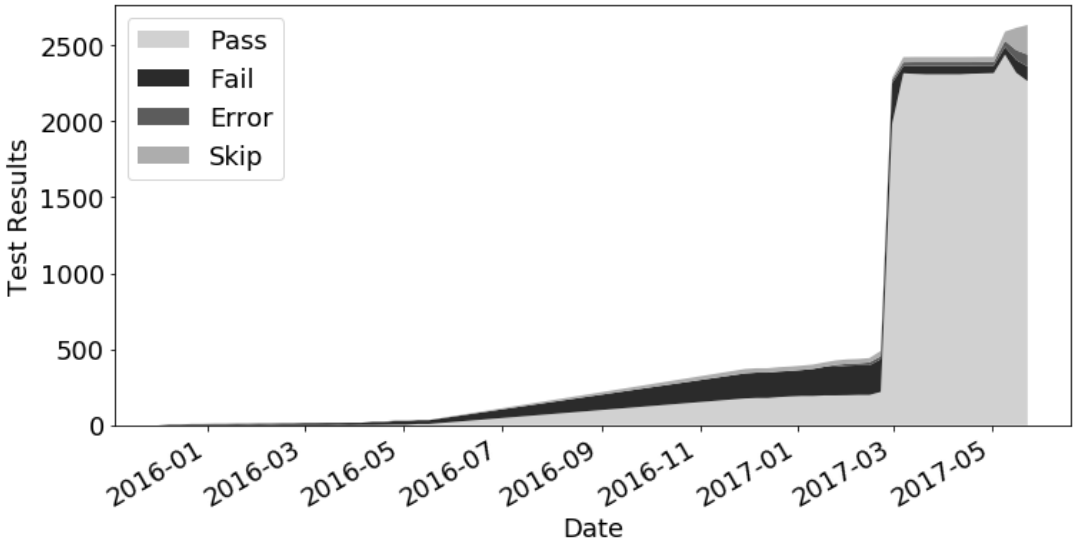


Figure 4. Vendor A STU3 (Standard for Trial Use version 3) servers.



Tests by Use Case

During the study period there were 3253 identified user-initiated test executions on Crucible, of which 1970 included only a single test suite. Four of the top 10 test suites executed were Argonaut suites. The other commonly executed suites include the most general tests: reading, searching, history retrieval, and formatting, as well as the transaction and batch test. The FHIR patient-resource test was the most-used resource test since it is one of the most important and central resources in the FHIR specification.

Within Touchstone, there were 529,847 tests run during the study period. A total of 99,848 (18.8%) of the tests executed were specifically testing the FHIR Patient Resource, while 55,163 (10.4%) tested the terminology functionality. Touchstone also includes tests for HL7 Connectathon tracks, which comprised 125,720 (23.7%) of the tests run by volume, although many of these tests were most likely run outside of Connectathons.

The top tests executed on Crucible and Touchstone are listed in Table 1.

Tests by Version

FHIR is an evolving standard that has seen three major releases in the last 4 years and a dozen minor releases in the same time frame [9]. The topic of FHIR versioning has similarly evolved

XSL•FO
RenderX

as various vendors begin to build and deploy production services. Only recently, after the study period, did HL7 add versioning information to the FHIR specification, at a maturity level of *not applicable* (N/A) and status of *informative*, meaning it is merely information and not rules to be followed [18,19]. Currently, and during the study period, both Touchstone and Crucible examined the FHIR Server Capability Statement to determine the declared version of FHIR supported.

Crucible supports testing the last two major versions of FHIR, while Touchstone supports testing all point releases since FHIR 1.0. Figures 5 and 6 show the community shift from testing one version of FHIR to the next. Examining tests tagged with specific FHIR versions, we can see growth in testing usage throughout the lifetime of Touchstone. The dip in testing volume of FHIR 3.0.0 is indicative of the swift transition to FHIR 3.0.1.

Community Engagement

Beyond providing the tools themselves, the Crucible and Touchstone teams have maintained considerable involvement with the FHIR development community by attending Connectathons sponsored by HL7, assisting the Argonaut group by providing tailor-made tests for their use cases, and in the case of Crucible, reaching out to the open-source community for involvement in the development of the software.

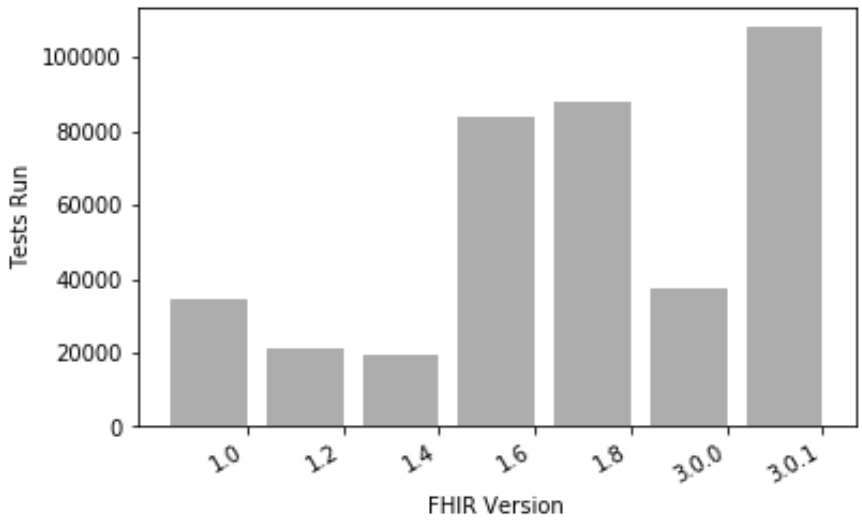
Connectathons

The Crucible development team has attended each HL7-sponsored FHIR Connectathon since Connectathon 8 in January 2015 through Connectathon 17 in January 2018. The AEGIS.net team has attended each HL7-sponsored FHIR Connectathon since Connectathon 4 in September 2013 and introduced Touchstone at Connectathon 10 in October 2015. Both Crucible and Touchstone develop and support a suite of tests for each Connectathon, specific to that event's tracks. The Touchstone team regularly runs a "Developers Introduction to FHIR" session parallel to each Connectathon introducing FHIR and TDD.

Table 1. Top tests executed by users.

Rank	Crucible		Touchstone	
	Test ID	Number of executions	Test ID	Number of executions
1	Argonaut Sprint 1	858	Patient Resource Test	132,328
2	Read Test	664	ValueSet Resource Test	59,162
3	Argonaut Sprint 3	549	Practitioner Resource Test	26,282
4	Argonaut Sprint 4	539	Organization Resource Test	22,124
5	History001	476	Location Resource Test	12,243
6	Search001	475	Observation Resource Test	11,592
7	Format001	460	Device Resource Test	10,698
8	Argonaut Sprint 5	451	AllergyIntolerance Resource Test	10,621
9	Transaction and Batch Test	447	Appointment Resource Test	10,232
10	Patient Resource Test	445	Condition Resource Test	9588

Figure 5. Touchstone usage by Fast Healthcare Interoperability Resources (FHIR) version.



XSL•FO
RenderX

Figure 6. Touchstone Fast Healthcare Interoperability Resources (FHIR) by version over time. DSTU2: Draft Standard for Trial Use version 2; STU3: Standard for Trial Use version 3.

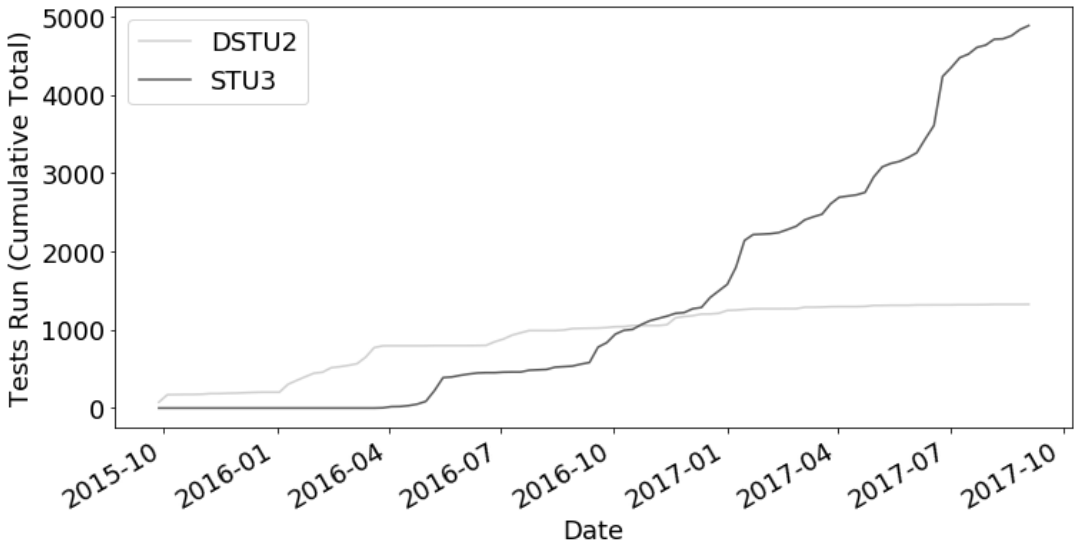


Table 2. Anonymized performance of Argonaut members at the completion of each Argonaut Sprint.

Vendor	Sprint			Resprint			Sprint				Argonaut Connectathon tests
	1	2	3	1	2	3	4	5	6	7	
E	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
F	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
G	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail
I	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail
M	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
N	Pass	Fail	Pass	Fail	Fail	Fail	Pass	Pass	Pass	Pass	Pass
O	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail
P	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail
Q	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
R	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
S	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail
T	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
U	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
V	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail

Argonaut

The Argonaut Project is a private sector initiative with the mission of advancing industry adoption of modern open interoperability standards. Its stated purpose is to “develop a first-generation FHIR-based API and Core Data Services specification to enable expanded information sharing for electronic health records and other health information technology using existing Internet standards and architectural patterns and styles” [11].

With Touchstone’s and Crucible’s missions to advance the adoption of the FHIR API, both teams collaborated with Argonaut vendors to develop a series of test suites to help them test their FHIR implementations. Crucible’s test suite results show almost all Argonaut members failed these test suites initially. However, as shown in Table 2, many of the members now support most of the test suites. This could be attributed to the high volume of testing that was performed on the Argonaut suites during the initial Argonauts implementation sprints.

Caveats

Electronic health records have structured and unstructured data. FHIR supports both of these data types: structured data using Resources and unstructured data using Binary and DocumentReference [20,21]. Neither Crucible nor Touchstone test for clinical correctness; they focus purely on technical correctness. Achieving health data interoperability may also

XSL•FO
RenderX

require a higher level of interoperability (eg, semantic) beyond technical exchange.

Conclusions

Crucible and Touchstone have proven to be valuable tools for the FHIR developer community. These tools can aid in the transition to an interoperable health care infrastructure by providing open reference implementations for FHIR testing and support future Cures Act requirements. Our research shows that developers that use testing and validation tools tend to produce more compliant FHIR implementations. The test data collected by MITRE and AEGIS.net during the study period shows that when it comes to health data interoperability, "practice makes perfect." This gives us hope that a future with ubiquitous health care information interoperability is possible.

Acknowledgments

MITRE research reported in this publication was supported by The MITRE Innovation Program (Approved for Public Release; Distribution Unlimited, Case Numbers 16-0597 and 17-3214-6) and the ONC, US Department of Health and Human Services, Washington, DC, USA.

Conflicts of Interest

JW, RS, and CD are employed by The MITRE Corporation, which funded the development of Crucible software. MH and RE are employed by AEGIS.net, which owns and develops the Touchstone software.

Authors' Contributions

Crucible development was led by JW and RS. Touchstone development was led by RE and MH. Data analysis was completed by CD, MH, and JW. Policy input and review was provided by SP. All authors contributed to the writing and final approval of this manuscript.

References

1. Furukawa M, King J, Patel V, Hsiao C, Adler-Milstein J, Jha A. Health Affairs. 2014. Despite substantial progress in EHR adoption, health information exchange and patient engagement remain low in office settings URL: https://www.healthaffairs.org/doi/pdf/10.1377/hlthaff.2014.0445 [accessed 2018-04-24] [WebCite Cache ID 6yvAfFpTx]
2. Pham H, O'Malley AS, Bach P, Saiontz-Martinez C, Schrag D. Primary care physicians' links to other physicians through Medicare patients: The scope of care coordination. Ann Intern Med 2009 Feb 17;150(4):236-242 [FREE Full text] [Medline: 19221375]
3. Davis K, Stremikis K, Schoen C, Squires D. Mirror, Mirror on the Wall: How the US Health Care System Compares Internationally. 2014 Update. Washington, DC: Commonwealth Fund; 2014 Jun. URL: http://www.commonwealthfund.org/~/media/files/publications/fund-report/2014/jun/1755_davis_mirror_mirror_2014.pdf [accessed 2018-04-24] [WebCite Cache ID 6yvHla1mt]
4. Middleton B. ResearchGate. Wellesley, MA: Center for Information Technology Leadership; 2004. The value of healthcare information exchange and interoperability URL: https://www.researchgate.net/publication/228874733_The_value_of_healthcare_information_exchange_and_interoperability [accessed 2018-10-15] [WebCite Cache ID 73CHcFhaF]
5. The Value of Medical Device Interoperability: Improving Patient Care With More Than $30 Billion in Annual Health Care Savings. San Diego, CA: West Health Institute; 2013 Mar. URL: http://www.westhealth.org/wp-content/uploads/2015/02/The-Value-of-Medical-Device-Interoperability.pdf [accessed 2018-04-24] [WebCite Cache ID 6yvImmiVA]
6. JASON, The MITRE Corporation. A Robust Health Data Infrastructure. Rockville, MD: Agency for Healthcare Research and Quality; 2014 Apr. URL: https://www.healthit.gov/sites/default/files/ptp13-700hhs_white.pdf [accessed 2018-04-05] [WebCite Cache ID 6yvIre3Kc]
7. HealthIT.gov. Cypress URL: https://www.healthit.gov/cypress/ [WebCite Cache ID 6yvIv8DGl]
8. AEGIS. AEGIS Developers Integration Lab URL: http://www.aegis.net/dil.html [WebCite Cache ID 6yvJBZi8s]
9. Health Level Seven International. FHIR URL: http://hl7.org/fhir [WebCite Cache ID 6yvJI1RY5]
10. Federal Register: The Daily Journal of the United States Government. Washington, DC: Department of Health and Human Services; 2015 Oct 16. 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.gpo.gov/fdsys/pkg/FR-2015-10-16/pdf/2015-25597.pdf [accessed 2018-10-11] [WebCite Cache ID 7360baQEE]
11. Health Level Seven International. Argonaut Project URL: http://argonautwiki.hl7.org/index.php?title=Main_Page [WebCite Cache ID 6yvJfZmZo]
12. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap. Washington, DC: The Office of the National Coordinator for Health Information Technology; 2015. URL: https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf [accessed 2018-04-24] [WebCite Cache ID 6yvK561QK]

XSL•FO
RenderX

13. United States 114th Congress. The 21st Century Cures Act, Public Law 114-255. Washington, DC: Government Publishing Office; 2016 Dec 13. URL: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf [accessed 2018-04-24] [WebCite Cache ID 6yvK91Fhm]
14. Project Crucible. URL: https://projectcrucible.org/ [WebCite Cache ID 6yvKBeAdB]
15. Touchstone. URL: http://www.touchstone.com/ [WebCite Cache ID 7361UfMja]
16. Sprinkler. URL: https://github.com/FirelyTeam/sprinkler [WebCite Cache ID 72kNgAF3G]
17. S4S Test Suite. URL: https://tests.demo.syncfor.science [WebCite Cache ID 72kNnhDh6]
18. Health Level Seven International. Managing Multiple FHIR Versions URL: http://hl7.org/fhir/2018Sep/versioning.html [WebCite Cache ID 72kOpWaOV]
19. Health Level Seven International. RESTful API: FHIR Version Parameter URL: http://hl7.org/fhir/2018Sep/http.html#version-parameter [WebCite Cache ID 72kOutmps]
20. Health Level Seven International. Resource Binary - Content URL: http://hl7.org/fhir/2018Sep/binary.html [WebCite Cache ID 72kQ1n0yn]
21. Health Level Seven International. Resource DocumentReference - Content URL: http://hl7.org/fhir/2018Sep/documentreference.html [WebCite Cache ID 72kQ8tXYa]

Abbreviations

API: application programming interface
DSTU: Draft Standard for Trial Use
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level Seven International
ONC: Office of the National Coordinator for Health Information Technology
PCP: primary care physician
STU3: Standard for Trial Use version 3
TDD: Test-Driven-Development

Edited by G Eysenbach; submitted 24.04.18; peer-reviewed by G Grieve, Y Chu; comments to author 25.08.18; revised version received 07.09.18; accepted 29.09.18; published 23.10.18.

Please cite as:
Walonoski J, Scanlon R, Dowling C, Hyland M, Ettema R, Posnack S
Validation and Testing of Fast Healthcare Interoperability Resources Standards Compliance: Data Analysis
JMIR Med Inform 2018;6(4):e10870
URL: http://medinform.jmir.org/2018/4/e10870/
doi:10.2196/10870
PMID:30355549

©Jason Walonoski, Robert Scanlon, Conor Dowling, Mario Hyland, Richard Ettema, Steven Posnack. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 23.10.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX