JMIR Medical Informatics

Impact Factor (2022): 3.2
Volume 6 (2018), Issue 2 ISSN 2291-9694 Editor in Chief: Christian Lovis, MD, MPH, FACMI

Contents

Original Papers

Reviews

Original Paper

Privacy-Preserving Patient Similarity Learning in a Federated Environment: Development and Analysis

Junghye Lee[1,2,3], PhD; Jimeng Sun[4], PhD; Fei Wang[5], PhD; Shuang Wang[2], PhD; Chi-Hyuck Jun[3], PhD; Xiaoqian Jiang[2], PhD

[1]School of Management Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic Of Korea

[2]Department of Biomedical Informatics, University of California San Diego, San Diego, CA, United States

[3]Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Republic Of Korea

[4]College of Computing, Georgia Institute of Technology, Atlanta, GA, United States

[5]Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medical College, Cornell University, New York City, NY, United States

Corresponding Author:
Junghye Lee, PhD
School of Management Engineering
Ulsan National Institute of Science and Technology
UNIST Industrial complex campus
10, Techno Saneop-ro 55beon-gil, Nam-gu
Ulsan, 44776
Republic Of Korea
Phone: 82 52 217 3129
Email: jul289@ucsd.edu

Abstract

Background: There is an urgent need for the development of global analytic frameworks that can perform analyses in a privacy-preserving federated environment across multiple institutions without privacy leakage. A few studies on the topic of federated medical analysis have been conducted recently with the focus on several algorithms. However, none of them have solved similar patient matching, which is useful for applications such as cohort construction for cross-institution observational studies, disease surveillance, and clinical trials recruitment.

Objective: The aim of this study was to present a privacy-preserving platform in a federated setting for patient similarity learning across institutions. Without sharing patient-level information, our model can find similar patients from one hospital to another.

Methods: We proposed a federated patient hashing framework and developed a novel algorithm to learn context-specific hash codes to represent patients across institutions. The similarities between patients can be efficiently computed using the resulting hash codes of corresponding patients. To avoid security attack from reverse engineering on the model, we applied homomorphic encryption to patient similarity search in a federated setting.

Results: We used sequential medical events extracted from the Multiparameter Intelligent Monitoring in Intensive Care-III database to evaluate the proposed algorithm in predicting the incidence of five diseases independently. Our algorithm achieved averaged area under the curves of 0.9154 and 0.8012 with balanced and imbalanced data, respectively, in κ-nearest neighbor with κ=3. We also confirmed privacy preservation in similarity search by using homomorphic encryption.

Conclusions: The proposed algorithm can help search similar patients across institutions effectively to support federated data analysis in a privacy-preserving manner.

(JMIR Med Inform 2018;6(2):e20) doi:10.2196/medinform.7744

KEYWORDS

privacy; federated environment; similarity learning; hashing; homomorphic encryption

Introduction

Data-Driven Decision Making in Medical Fields

Electronic health records (EHRs) are becoming ubiquitous across almost all medical institutions. They provide insight into diagnoses [1-6], as well as prognoses [7-10] and can assist in the development of cost-effective treatment and management programs [8,11-14]. All kinds of data across institutions are being collected in EHRs, including diagnosis, medication, lab results, procedures, and clinical notes. In the recently announced precision medicine initiative, many more other types of data including omics data such as genomic and proteomic data and behavior data such as activity sensor data are being generated and collected by doctors and patients. As such rich and heterogeneous health data become available, the entire medical research and practice are shifting from the knowledge or guideline-driven approaches to the data or evidence-driven paradigm, where effective and efficient algorithms become the key for clinical research and practice.

Limitations of Single-Institutional Studies

Previously, many biomedical studies were conducted within a single institution having limited EHR data because of the lack of federated data analysis framework and the institutional privacy concerns on data sharing. However, such an approach has many limitations. For example, it has been demonstrated that genome-wide association studies on EHR data often failed to discover known biomarkers from a single institution because of limited sample size [15,16]. To enable cross-institutional studies, many collaborative networks have been proposed, such as mini-sentinel [17], Observational Health Data Sciences and Informatics [18], National Patient-Centered Clinical Research Network [19], and i2b2 Shared Health Research Informatics Network [20]. These frameworks enable certain analyses (such as database queries with very specific inclusion or exclusion criteria) to be conducted efficiently in a federated manner. However, more sophisticated analyses such as predictive models [21] and context-specific patient similarity search [22] are still a challenge for most existing frameworks, as cross-institutional EHR data exchange is required to build such models, which is usually infeasible because of the institutional privacy and security concerns. There is an urgent need for the development of novel frameworks that can perform analyses in a privacy-preserving federated environment across multiple institutions. In this way, global analytic models can be built collectively without sharing raw EHR data. A few studies on the topic of federated clinical analysis [23-26] have been conducted recently with the focus on different algorithms. However, none of them have solved the problem of similar patient matching, which is important for many biomedical studies. Therefore, we plan to develop a privacy-preserving analytic platform that focuses on a suite of algorithmic challenges on patient similarity learning.

Patient Similarity Learning

Patient similarity learning aims to develop computational algorithms for defining and locating clinically similar patients to a query patient under a specific clinical context [7,27-30]. The patient similarity search is very challenging because the raw EHR data is sparse, high-dimensional, and noisy, which makes finding an exact match among patients using EHR data almost impossible. Besides, patient similarity learning is often context-specific. For example, patient similarity measure for heart disease management can be very different from cancer management. The fundamental challenge is how we can perform effective context-specific patient similarity learning in a federated setting, which enables many different applications:

- Cohort construction: cross-institution observational studies are challenging but necessary as many studies require a large and specific patient cohort that does not exist within a single institution. To conduct such a study, an efficient similarity search needs to be conducted across institutions to identify the focused patient cohort.
- Disease surveillance: The Centers for Disease Control and Prevention monitors thousands of hospitals for potential epidemics. When a suspicious case is reported, there is a need to find similar cases across geographies.
- Clinical trial recruitment: pharmaceutical companies often need to spend significant amount of time and resources to identify targeted patients through many different clinical institutions. Ideally, they would like to be able to perform patient similarity search across all clinical institutions to identify where those relevant patients are. Then they can quickly focus on recruiting patients from the right clinical institutions.

Patient similarity learning involves two computational phases: (1) patient representation learning is to learn the context-specific representation of patients based on their EHR data. For example, patients may be given different representations in heart disease management versus cancer management and (2) patient similarity search is to find similar patients based on their corresponding representations. In a federated environment where multiple institutions exist, patient similarity learning has many unique challenges: (1) how to design an efficient but flexible patient representation that enables fast similarity search? (2) how to learn patient representation from heterogeneous data sources? and (3) how to preserve privacy while still allowing the computation of the patient representation and the search of similar patients across institutions?

Research Objective

The main objective of this paper was to develop a privacy-preserving analytic platform for patient similarity learning in a distributed manner. We propose to learn context-specific binary hash codes to represent patients across institutions. The similarities between patients can be efficiently computed as the hamming distance using the resulting hash codes of corresponding patients; the hamming distance is defined to be the number of places where two binary codes differ. As patient data are heterogeneous from multiple sources such as diagnosis, medication, and lab results, we propose a multi-hash approach that learns a hash function for each data source. Then, the patient similarity is calculated by hash codes from data sources. To avoid the potential security risk because of the attack from malicious users, we also adopt homomorphic encryption [31] to support secure patient similarity search in a

federated setting. Finally, the proposed algorithm is applied and validated on real data.

Methods

Feature Construction

For K feature domains, we assume a vector-based representation for patients in every feature domain ($1 \leq k \leq K$). There are different ways to construct the feature vectors: (1) for nominal features with standard dictionaries, such as diagnosis and procedure codes, we can use either binary value for presence, or code frequency within the observation period (where the features are extracted from); (2) for continuous features such as age or lab test values, we can use them as they are or we can first quantize them and treat each quantized region as a nominal feature. For example, the values of a specific lab test can be quantized as critical low, low, normal, high, and critical high; and (3) for time-evolving features, if we want to consider the temporal trends in the feature construction process, we can first construct a temporal pattern dictionary with either data-driven method or expertise knowledge, and then treat each pattern as a nominal feature. For example, if there are four types of features including two demographics, 20 prescriptions, 15 lab tests, and 10 diagnoses, we can construct a vector-based representation for patient A as shown in Figure 1. We represent gender as a binary value and age as it is. For diagnosis, prescription, and lab test, we add a one-hot representation of each event (ie, $\{0,1\}^{|C|}$ with the number of codes $|C|$).

Hashing

In general, hashing is an approach of transforming the data item to a low-dimensional representation, or equivalently a short code consisting of a sequence of bits (Figure 2).

Hashing technologies can be applied in many applications such as Bloom filter [32] and cryptography [33]. Similarity-based hashing [34] is one specific type of hashing that aims to preserve the data similarities in their original space with hash codes. On the basis of the availability of supervision information, a similarity-based hashing method can be categorized as unsupervised [35-40], semi-supervised [41-43], or supervised hashing [44,45]. Unsupervised methods learn hash functions purely based on data distributions. Supervised methods exploit the labeled pairwise relationship between entities to capture the high-level data semantics. Semi-supervised methods lie in between them, that is, they explore both data distribution characteristics and labeled pairwise data relationships to learn the hash functions. Most of these existing methods assume a single vector-based representation for every data object.

However, one challenge in our scenario is that the patient features are highly heterogeneous, that is, the features for characterizing the patients are of different types. In this case, it may not be effective to represent each patient as a single vector (simple concatenation will not work as different features are of different types and have different value range). There are some existing multi-modal hashing methods [46-52] that aim to derive a unified single-hash table for encoding the data objects with heterogeneous features. The problem with single-hash (or uni-hash) table is that it is difficult to discover the latent similarity components [53] derived from different feature types, which is crucial in our scenario. For example, it is important to know how similar two patients are, but also why (eg, patients A and B are similar to each other mainly because of their similar demographics and patients B and C are similar because of their similar diagnosis history and lab test values).

Figure 1. Example of feature construction. Prescription, lab test, and diagnosis are denoted by p, l, and d, respectively.

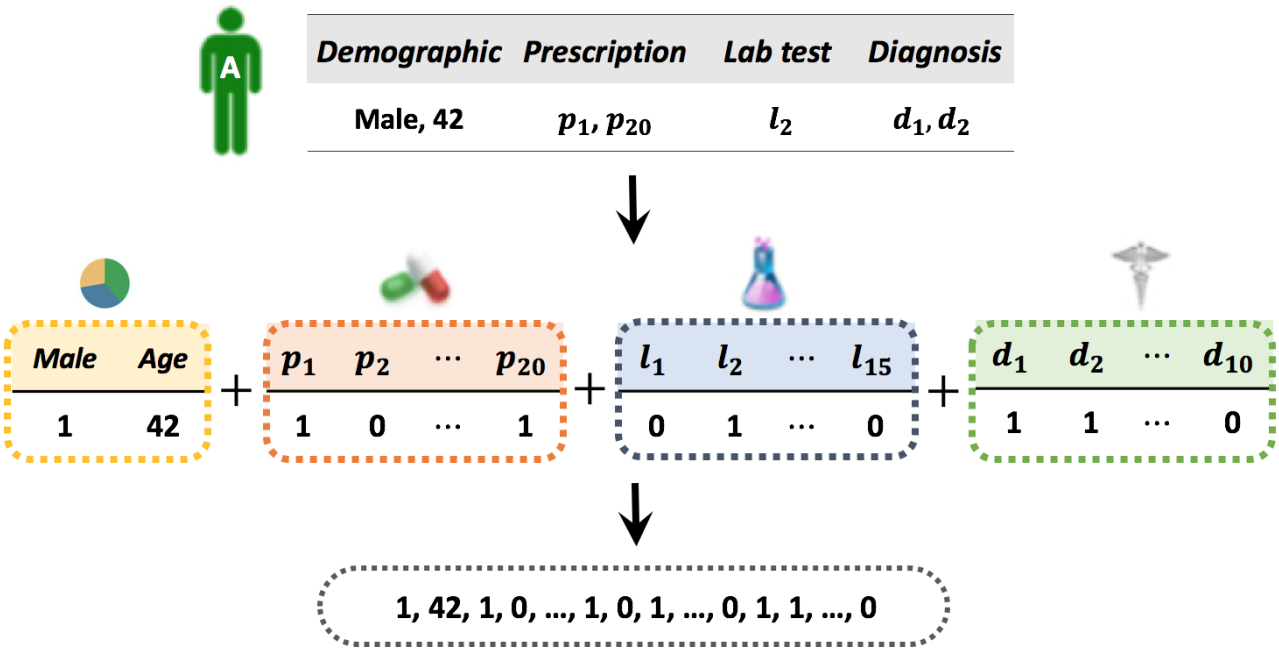
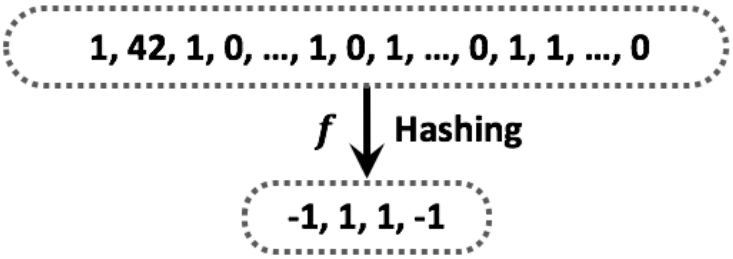


Figure 2. Example of hashing.



Federated Patient Hashing Framework

Symbols used in this paper are listed in Textbox 1.

Figure 3 illustrates the overall federated patient matching framework. Suppose there are M sites with the i-th site S^i which owns a patient population P^i. We use p^i_j to represent the j-th patient in P^i. Then, our problem is, given a query patient, how to retrieve similar patients from those M sites without explicitly accessing the patient feature vectors. Our plan is to resolve this problem using similarity based hashing, which transforms the patient's raw features into a binary vector representing patient characteristics (patient representation learning). The pairwise patient similarities will be evaluated as the pairwise distance based on those signatures (patient similarity search). In this paper, we will focus on feature-based hashing, that is, those binary patient signatures are obtained by proper transformation from patient features. Therefore, to perform hashing, we need to first construct feature-based representation for patients.

Without the loss of generality, we assume there are K different feature types to characterize every p^i_j, and we use p^i_{jk} (k=1,2,…, K) to represent the k-th type of feature vector of p^i_j. The goal is to derive an effective computational framework for patient matching in a federated environment, and the key idea is to learn a good hash function that can transform the patient features into binary hash codes. A uni-hash table approach shown in Figures 2 and 3 is to learn only one hash function for the feature vector f: $R^d \rightarrow \{-1,+1\}^b$, where d is the dimensionality of the whole feature vector, and b is the number of bits of the hash codes learned d by f. In this paper, we propose a multi-hash approach for patient hashing that aims to learn a hash function f_k: $R^d_k \rightarrow \{-1,+1\}^b_k$ for every patient feature type k (k=1,2,…, K); d_k is the dimensionality of the k-th feature type, and b_k is the number of bits of the learned hash codes for the k-th feature type. Each f_k (k=1,2,…, K) is shared across all the M sites. We use the sign function to construct the hash codes, that is, sign $(Q^i_k) \in \{-1,+1\}^{b_k \times N_i}$, where Q^i_k is transformed numerical data from original data of i-th site for k-th type of feature $P^i_k \in R^{d_k \times N_i}$ by a hash function f_k that incorporates function coefficients for the k-th feature type $W_k \in {}^{d_k \times b_k}$; N_i is the population size of i-th site. How these components are formulated is described in the next paragraph in detail. We use H^i_k=sign(Q^i_k) to denote the hash codes of k-th feature type for the patients at i-th site. Figure 4 shows the process of patient similarity calculation with a multi-hash approach.

The u-th column of H^i_k, $h^i_{uk} \in \{-1,+1\}^b_k$ is the hash codes of p^i_{uk}. Then, the similarity between p^i_{uk} and p^i_{uk} can be evaluated as the inner product of h^i_{uk} and h^i_{uk} as shown in equation 1:

(1) $s^i_{kuv} = 1/b_k (h^i_{uk})^T (h^i_{vk})$

Thus, the overall similarity can be computed as the average of K similarities, as shown in equation 2, which is bounded on the interval of (−1,1).

(2) $s^i_{uv} = 1/K \sum_k (h^i_{uk})^T (h^i_{vk})$

Here, we suggest a general framework for learning $\{W_k\}^K_{k=1}$, which is the most important component. The framework basically constructs an objective function in terms of $\{W_k\}^K_{k=1}$ such as shown in equation 3, where λ_S, λ_U, and λ_W are regularizers of S ($\{W_k\}^K_{k=1}$), U ($\{W_k\}^K_{k=1}$), and $\Omega(\{W_k\}^K_{k=1})$, respectively, and then minimizes (or maximizes) it:

(3) $J(\{W_k\}^K_{k=1}) = \Psi(\{W_k\}^K_{k=1}) + \lambda_S S(\{W_k\}^K_{k=1}) + \lambda_U U(\{W_k\}^K_{k=1}) + \lambda_W \Omega(\{W_k\}^K_{k=1})$

$\Psi(\{W_k\}^K_{k=1})$ is a reconfiguration error term between the low-dimensional representation of the original data and hash codes, which is the main term of the objective function, and generates the hash codes from the original data, as shown in equation 4, where $\|\cdot\|_F$ is a Frobenius norm [54]. On the basis of this term, the hash function in our framework is formed as $f_k(P^i_k) = W^T_k P^i_k$, and this transformation results in H^i_k=sign(Q^i_k).

(4) $\Psi(\{W_k\}^K_{k=1}) = \sum_i \sum_k \|W^T_k P^i_k - H^i_k\|^2_F$

The objective function can incorporate regularizers, as well as the main term to obtain better solutions of $\{W_k\}^K_{k=1}$ by (1) introducing additional information to improve either unsupervised or supervised learning if desired, (2) solving an ill-posed problem, and 3) preventing overfitting. Possible regularizers are listed as follows:

$S(\{W_k\}^K_{k=1})$ is a supervised loss term that measures the quantization loss during the hashing process when supervision information is available for the patients. Here, the supervision information could be the labels of the patients, such as the disease the patients have. For example, if both p^i_u and p^i_v have the same disease, then their relationship r^i_{uv}=1, otherwise r^i_{uv}=-1. Then, we can set $S(\{W_k\}^K_{k=1})$ as shown in equation 5:

(5) $S(\{W_k\}^K_{k=1}) = \sum_i \sum_k \sum_{u,v} - s^i_{kuv} r^i_{uv}$

Textbox 1. List of symbols.

M: the number of local sites

K: the number of feature types (domains)

S^i: i-th local site

P^i: patient population in S^i

N^i: patient population size of S^i

P^i_k: patient population for k-th type of feature in S^i

p^i_j: j-th column of P^i, j-th patient in P^i

p^i_{jk}: j-th column of P^i_k, k-th type of feature vector for p^i_j

f_k: k-th hash function

d_k: dimensionality of the k-th feature type

b_k: the number of bits of the learned hash codes for the k-th feature type

W_k: function coefficients of the hash function for the k-th feature type

w_{ik}: i-th column of W_k

Q^i_k: numerical data transformed from P^i_k

sign(Q^i_k): signed Q^i_k

H^i_k: hash codes for P^i_k(=sign(Q^i_k))

h^i_{jk}: j-th column of H^i_k, the hash codes of p^i_{jk}

$\Psi(\{W_k\}^K_{k=1})$: reconfiguration error term for $\{W_k\}^K_{k=1}$

$S(\{W_k\}^K_{k=1})$: supervised loss term for $\{W_k\}^K_{k=1}$

$U(\{W_k\}^K_{k=1})$: unsupervised loss term for $\{W_k\}^K_{k=1}$

$\Omega(\{W_k\}^K_{k=1})$: term related to $\{W_k\}^K_{k=1}$ itself

$L(x, y)$: loss function between x and y

λ_S: regularizer of $S(\{W_k\}^K_{k=1})$

λ_U: regularizer of $U(\{W_k\}^K_{k=1})$

λ_W: regularizer of $\Omega(\{W_k\}^K_{k=1})$

λ: regularizer of a supervised loss term

η: regularizer of a Frobenius norm for Q

σ^i_{kuv}: similarity between p^i_{uk} and p^i_{vk}

R^i: pairwise relationship of R^i for labeled information

S^i_k: pairwise similarity of P^i_k

r^i_{uv}: relationship between p^i_{uk} and p^i_{vk} for labeled information

s^i_{kuv}: similarity between p^i_{uk} and p^i_{vk}

$S_L(Q^i_k)$: approximated sign function for Q^i_k

Figure 3. The whole process of patient matching in a federated environment. The user sends a patient matching request to the service center, which is delegated to patient data resources from several clinical sites. Due to the privacy concerns, the center does not have access to the raw patient data. All patients within different sites need to be first hashed, and the center only has the patient's signatures after hashing. The hash functions are shared across different sites.

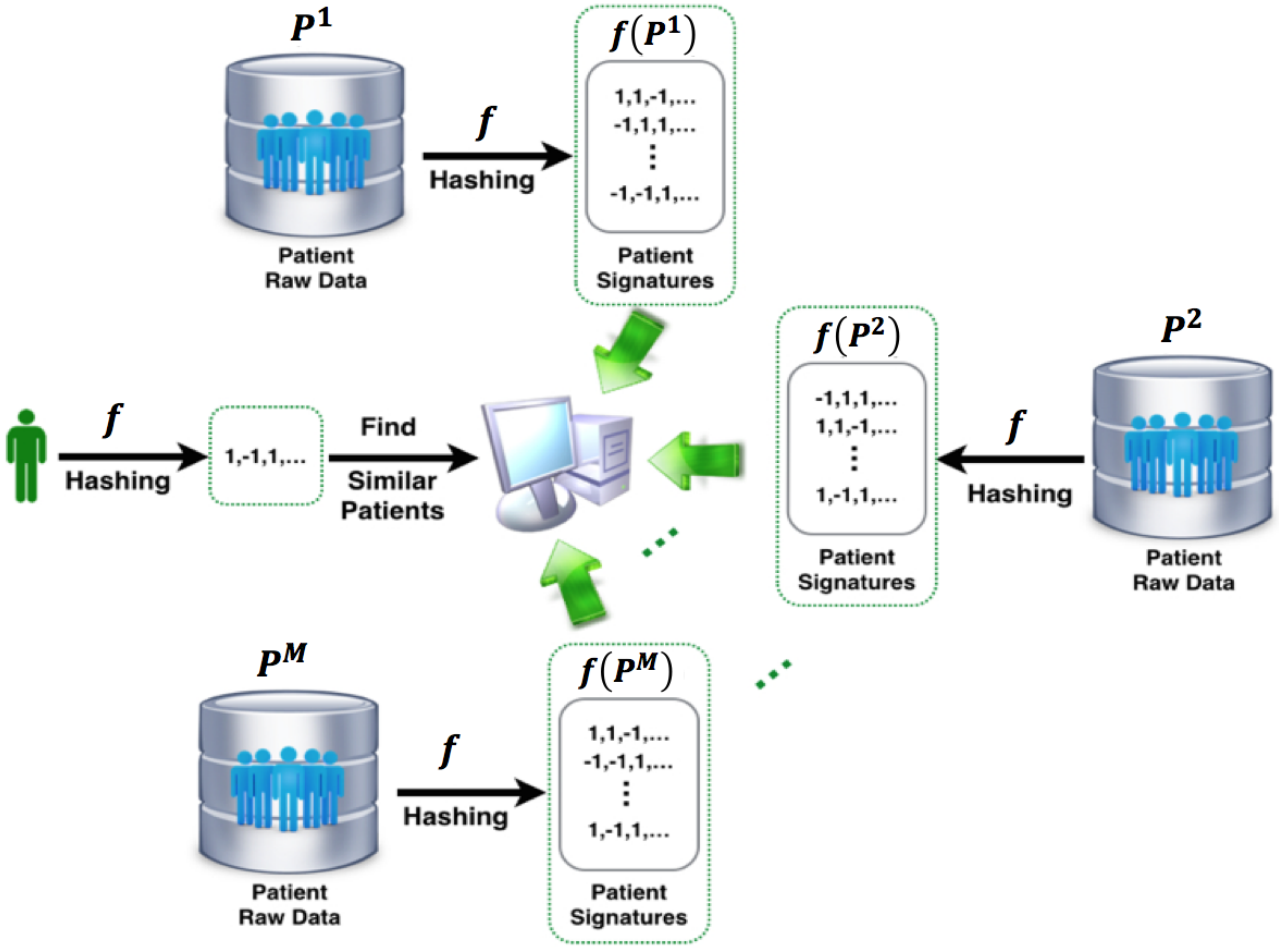
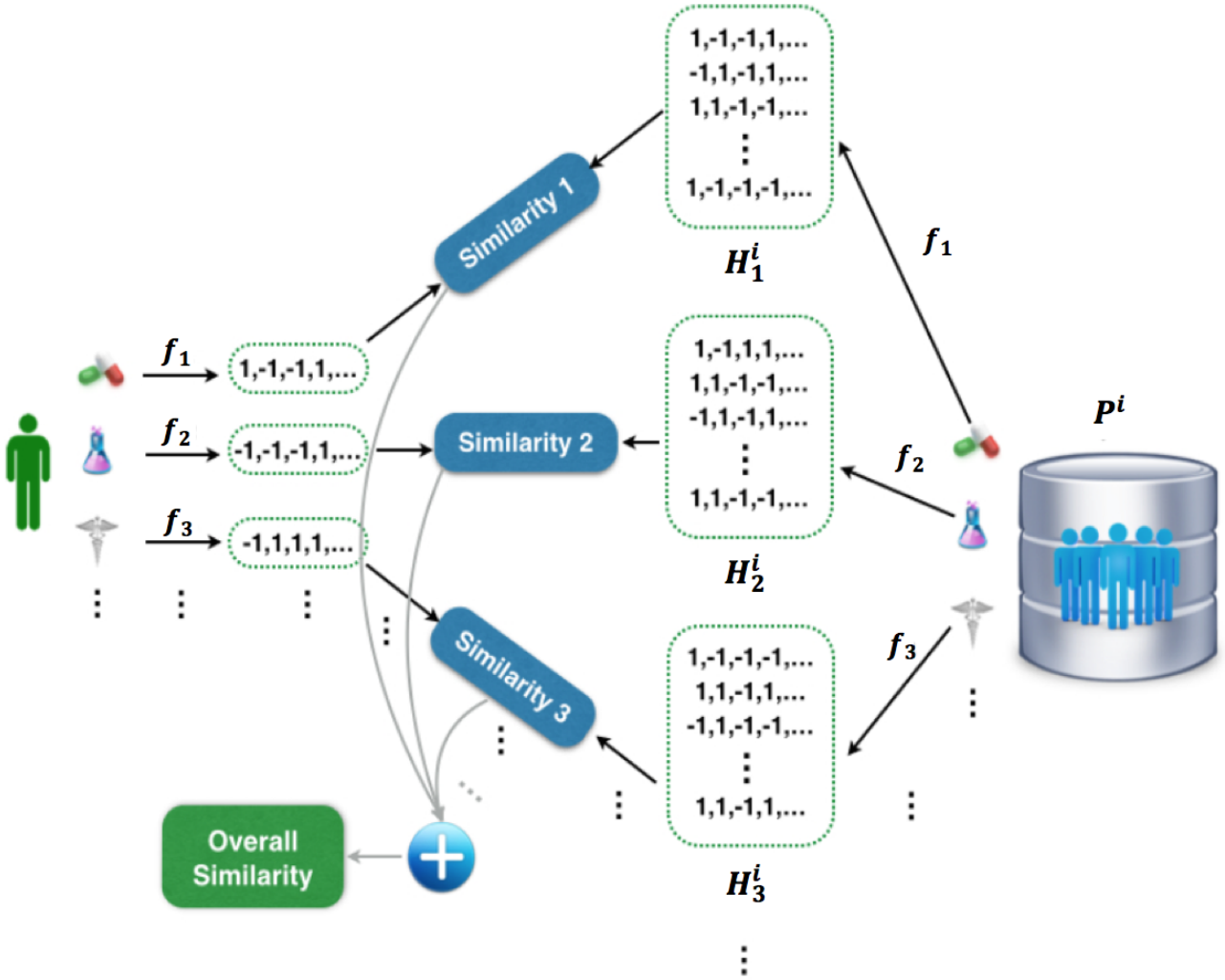


Figure 4. The process of calculating patient similarity with a multi-hash approach.



The possible choices of supervised loss term could be any loss function $L(x,y)$, and examples include $L(x,y)=-xy$ and well-known binary loss functions such as (1) logistic loss, $L(x,y)=\log(1+\exp(-xy))$ and (2) hinge loss, $L(x,y)=\max(0,1-xy)$.

Note that $U(\{W_k\}^K_{k=1})$ is an unsupervised term that exploits the intrinsic data distribution and enforces the resultant hash codes to comply with the distribution. For example, we can request similar patients to have similar hash codes on each feature type. This can be achieved by minimizing the below regularizer, as shown in equation 6, where σ^i_{kuv} is a similarity between p^i_{uk} and p^i_{vk} based on, for example, a Gaussian function for continuous valued features or a cosine function after Term Frequency-Inverse Document Frequency normalization on bag-of-code (eg, diagnosis code or procedure code):

(6) $U(\{W_k\}^K_{k=1}) = \sum_i\sum_k\sum_{u,v}\sigma^i_{kuv}\|h^i_{uk} - h^i_{uk}\|^2_F$

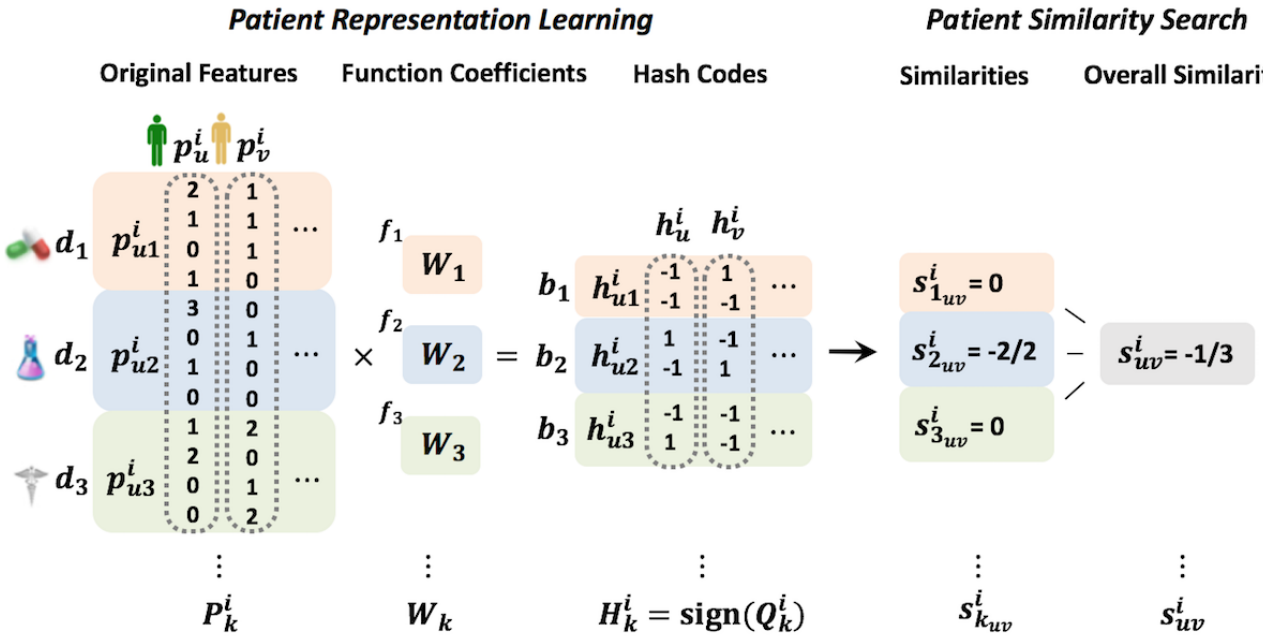
$\Omega(\{W_k\}^K_{k=1})$ is a term related to $\{W_k\}^K_{k=1}$ themselves, which is independent of the patient features. Examples of $\Omega(\{W_k\}^K_{k=1})$ include (1) Frobenius norm regularizer $\sum^K_{k=1}\|W_k\|^2_F$, which can be used for improving the numerical stability of the solution process and (2) orthogonality regularizer $\sum^K_{k=1}\sum_{i\neq j}\|w^T_{ik}w_{jk}\|^2$, where w_{ik} is the i-th column of W_k, which can encourage the diversity of the learned hash codes and thus improve their representation effectiveness.

Figure 5 shows a running example of the proposed hashing methodology. Such optimization problems can be solved with Block Coordinate Descent technologies [55], with $\{W_k\}^K_{k=1}$ as variable blocks that alternatively update W_k ($1\leq k\leq K$) one by one. Moreover, as different sites are continuously receiving new patients (or new patient features), we will need to continuously update the hash functions as well. Fortunately, as can be observed from equations 4, 5, and 6, those terms are fully decomposable with respect to different sites. Therefore, we can update the hash functions in an asynchronous manner, that is, we can update the current $\{W_k\}^K_{k=1}$ as soon as new patient data is received on site i.

Privacy-Preserving Patient Representation Learning in a Federated Setting

Without loss of generality, let us instantiate the objective function with the regularizer λ of empirical error on the labeled data for a family of hash codes; this choice might be the most basic approach to similar patient learning based on the fact that supervised learning is more commonly used than unsupervised learning because data generated in the medical field usually have label information.

Figure 5. Example of transformation of patient vectors into hash codes and computation of similarity between hash codes.



When solving the initiated objective function, two possible problems because of the sign function for Q arise. First, Q may not be a unique solution, and thus, the objective function is difficult to converge without considering any regularizer about Q. We add a Frobenius norm regularizer η to solve this problem. In addition, the objective function $f(W,Q)$ is nondifferentiable in terms of Q. We can approximate the sign function with the surrogate function. Then, we have the final objective function, as shown in equation 7, where $R^i \in R^{N_i \times N_i}$ is the pairwise relationship of P^i for labeled information:

(7) $f(W, Q) = \min \sum_i \sum_k \|W_k^T P^i_k - S_L(Q^i_k)\|^2_F + \lambda \sum_i \sum_k \mathrm{tr}(-S_L(Q^i_k) R^i S_L(Q^i_k)^T) + \eta \sum_i \sum_k \|Q^i_k\|^2_F$

If both p^i_u and p^i_v have the same disease, then their relationship r^i_{uv}=1, otherwise r^i_{uv}=-1, and $S_L(\cdot)$ is the surrogate function, as shown in equation 8, where is the hadamard (elementwise) product:

(8) $S_L(Q^i_k) = (Q^i_k \; Q^i_k + \xi)^{-1/2} Q^i_k$

The detailed process to derive the final objective function is given in Multimedia Appendix 1 (Note ⊗ is the Kronecker product [56]). The objective function for $\{W_k\}^K_{k=1}$ and $\{Q^i_k\}^{K,M}_{k,i=1}$ can be solved one by one iteratively as variable blocks [55] by using the Newton-Raphson method [57] until the estimates converge. To be specific, this approach first allows us to update W_k for each of k (k=1,2,…, K) with other W_l for all l ($1 \le l \ne k \le K$) and Q being fixed:

(9) $W^{new}_k = W_k - (\partial^2 f / \partial W^2_k)^{-1} \partial f / \partial W_k$

Then, similarly, we update Q^i_k for each combination of (i, k) ($1 \le i \le M, 1 \le k \le K$) with other combinations of (j, l) ($1 \le j \ne i \le M, 1 \le l \ne k \le K$) and W being fixed:

(10) $Q^{i,new}_k = Q^i_k - (\partial^2 f / \partial Q^{i\,2}_k)^{-1} \partial f / \partial Q^i_k$

The derivation process for the first and second derivatives of W and Q is described in Multimedia Appendix 1. As derivatives are linearly decomposable by sites i, the objective function defined in equation 7 can be computed in a distributed manner. This means the optimization only requires locally computed statistics to be delivered to estimate the $\{W_k\}^K_{k=1}$ iteratively until convergence.

The time complexity at each iteration depends on feature type k and site i. When updating W_k for each of k ($1 \le k \le K$) with other W_l for all l ($1 \le l \ne k \le K$) and Q being fixed, the time complexity is $O(d_k^3)$ because each site has to inverse the $d_k \times d_k$ Hessian matrix. When updating Q^i_k for each combination of (i, k) ($1 \le i \le M$, $1 \le k \le K$) with all other combinations of (j, l) ($1 \le j \ne i \le M$, $1 \le l \ne k \le K$) and W being fixed, the time complexity is $O(b_k^3 N_i^3)$ because S^i has to inverse the $b_k N_i \times b_k N_i$ Hessian matrix. Therefore, parameters that have a significant effect on time complexity include original and projection dimensions by feature type and population size by site. Other parameters such as the number of sites M and the number of feature types K along with the number of iterations are excluded in the big O notation because they are just constants. That is unless the number of site or the number of feature type goes to infinity, it only has a small impact on the complexity.

Privacy-Preserving Patient Similarity Search in a Federated Setting

To find similar patients across sites, hash codes for each site H^i (ie, $\{H^i_k\}^K_{k=1}$ have to be exchanged across institutions originally. However, when all other sites expect for i-th site receive H^i for similarity search, the patient-level information of i-th site can

be leaked by equation 4; other sites and a server can be united for reverse engineering to extract P^ibecause they have both $\{W_k\}^K_{k=1}$ and H^i, as well as their information in equation 4. Figure 6 illustrates the situation mentioned.

Therefore, we suggest the way to search similarity among different sites by avoiding revealing H^i_k but able to compute similarities based on H^i_k. We introduce homomorphic encryption specifically that is a form of encryption where a specific algebraic operation performed on the plaintext is equivalent to another algebraic operation performed on the cipher-text, and an encrypted result, when decrypted, matches the result of the same operation performed on the plaintext. Unlike traditional encryption schemes that do not allow any computations to be performed on the cipher-text without first decrypting it, homomorphic encryption allows computations to be performed without decrypting the data. The results of the computations remain encrypted and can only be read and interpreted by someone with access to the decryption key. Therefore, it is appropriate to use homomorphic encryption in our case that other sites and a server can attack maliciously. It enables cross-site comparison of health care statistics with protecting privacy for each site. The procedure of homomorphic encryption in this paper is summarized as follows: first, *i*-th site encrypts hash codes for its query data and delivers encrypted codes to *j*-th site. Next, *j*-th site performs the computation between delivered encrypted codes of *i*-th site and encrypted codes of *j*-th site without a decryption key and sends the computed value to *i*-th site. Finally, *i*-th site decrypts the value to get the hamming distance of hash codes between query data and data of *j*-th site. Each site is restricted to only answer the hamming distance to avoid the risk of privacy leakage. This process is depicted in Figure 7.

We note that homomorphic encryption provides an extra layer of privacy protection especially during patient similarity search.

Security

There are several participants in our framework.

- Data custodians (DCs) represent institutions or hospitals who have access to patient data and would like to collaborate in learning about similar patients.
- Crypto service provider (CSP) generates public and private keys. The public key is provided to the data custodians to safeguard the intermediary statistics.
- Cloud server (CS) computes over summary statistics from individual data custodians to obtain a global patient similarity model.

Our goal is that a DC does not learn patient-level information from other DCs during the process. We also want to ensure CS cannot infer patient-level information from the data. We assume a CSP is trustworthy and provides encryption keys (public and private). In the threat model, we assume the CS to be semi-honest, that is, it is honest to follow the protocol but curious about patient's private information while executing the protocol. We make the following basic assumptions: (1) DC and CS do not collude, (2) CS and CSP also do not collude, and (3) DC always receives correct keys from the CSP. To evaluate the security of our system, it is assumed that the security of the system is compromised if patient-level data or intermediary statistics that can infer patient-level data are leaked. CSP is only involved in generating public and private keys and transferring those keys to DCs, and no access to unintended fine-grained local information is involved in this process.

The leakage is related to computation of $\{W_k\}^K_{k=1}$, and possible scenarios according to the participants are as follows:

- Leakage to CSP in each computation: CSP does not participate in computation at all. Therefore, there is no leakage.
- Leakage to DC in each computation: each DC cannot indirectly learn patient data from other DCs only with $\{W_k\}^K_{k=1}$ and its local information $\{W_k\}^K_{k=1}$ and $\{Q^i_k\}^K_{k=1}$. If all DCs except for one collude, it is infeasible for the other DCs to reconstruct P^i_k of that one DC because the first and second derivatives of W_k have a nonlinear relationship for P^i_k. Specifically, it is not possible to specify a certain matrix only given information of covariance matrix because of insufficient equations. They also do not have information (first and second derivatives) about Q^i_k.
- Leakage to CS in each computation: CS cannot infer patient data from $\{W_k\}^K_{k=1}$. Even though CS receives local information for the first and second derivatives of $\{W_k\}^K_{k=1}$, it is infeasible for CS to recover $\{P^i_k\}^K_{k=1}$ for the same reason as the collusion among DCs. In finding similar patients, hash codes for each site $\{H^i_k\}^K_{k=1}$ have to be exchanged across institutions originally, but the use of homomorphic encryption prevents direct exchange of hash codes $\{H^i_k\}^K_{k=1}$ between DCs, and thus, there is no leakage.

Figure 6. Example of potential privacy leakage in patient similarity search across sites.

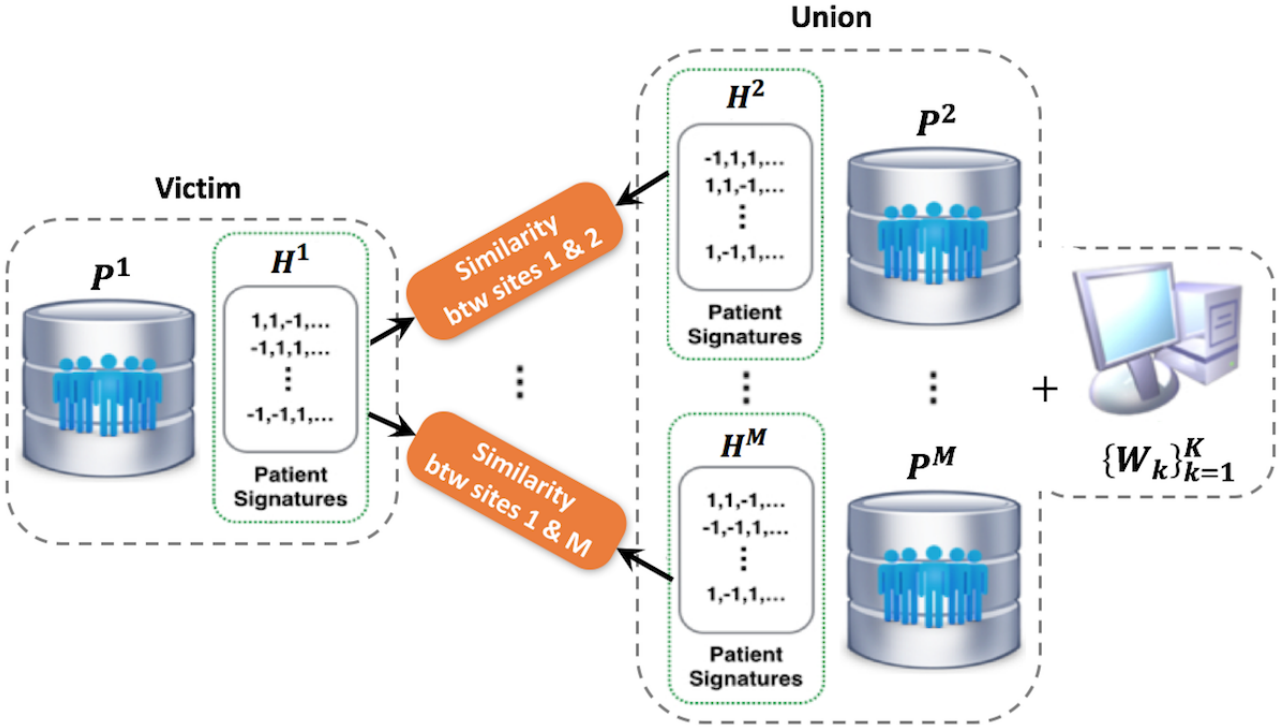
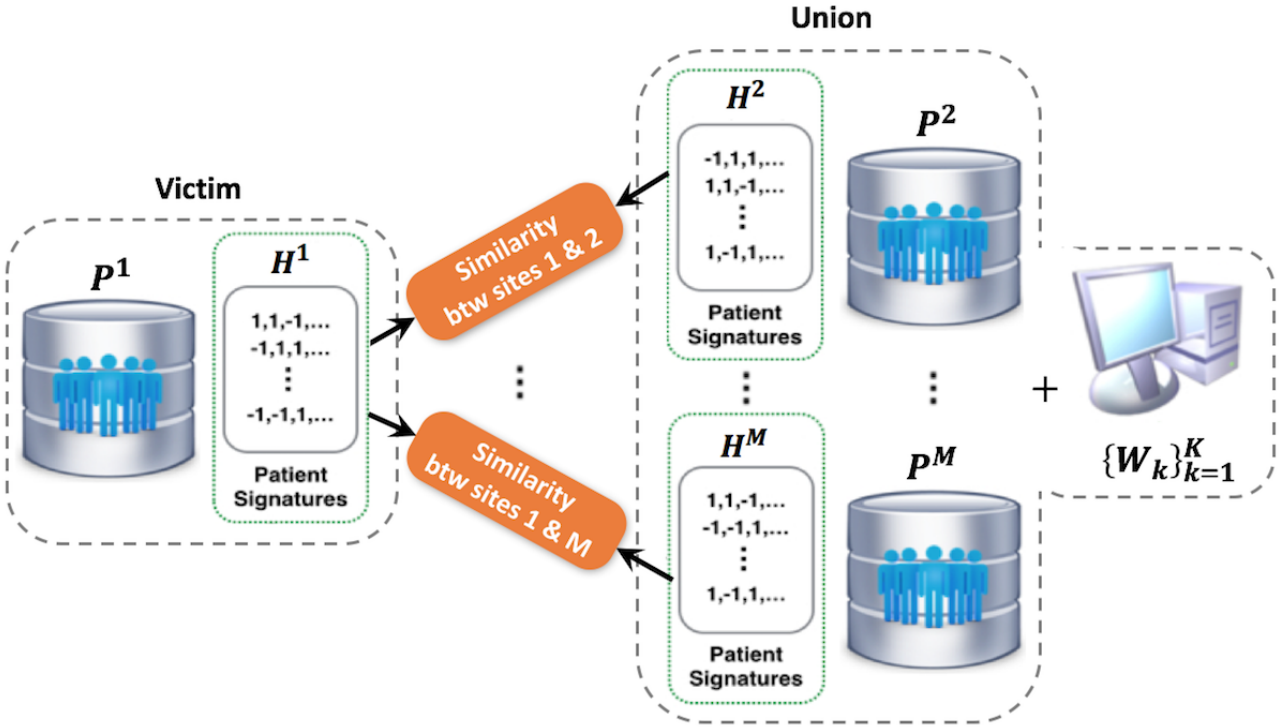


Figure 7. Privacy-preserving patient similarity search by homomorphic encryption; green key: encryption (public) key, blue key: decryption (private) key.



Results

Experimental Setting

We conducted experiments to validate our proposed method on real data. For comparison, we assumed two different systems against our system according to connection among *M* sites: open and closed system. In the open system, *M* sites can exchange their patient's information without any restrictions; in the closed system, each site can only utilize patient's information in each site. Ours is in the middle of two systems. For better understanding of these systems, let us assume that there are three sites A, B, and C with the same number of patients *N*. In this situation, an open system means that every site can access the complete information of the entire patient cohort ($3 \times N$), including information from other sites as well, and thus, three sites work like one site without any concerns on privacy. On the other hand, closed system indicates that each site can only access its patient-level information (*N*) exclusively. Open system and closed system are derived based on an idealistic situation and a realistic situation, respectively, and our system is in between these two systems, which cannot access patient's information from other sites but can utilize it through $\{W_k\}^K_{k=1}$. Then, we predicted the incidence of a certain disease and compared the standard κ-nearest neighbor (κ-NN) classification results based on hamming distance of multi-hash codes from our system with those based on hamming distance of multi-hash codes from open and closed systems, as well as uni-hash codes from open and closed systems. We also provided baseline results based on four similarity distances of raw data without using hashing for open and closed systems: Euclidean, cityblock, cosine, and correlation. We utilized five-fold cross validation (CV) that randomly splits patients into five folds with the equal size; we used four folds for training, and one fold for testing. As an evaluation measure, we used the area under the curve (AUC) where the true positive rate (TPR; ie, the number of true positives divided by the sum of true positives and false negatives) is plotted against the false positive rate (ie, the number of false positives divided by the sum of false positives and true positives) at various thresholds. AUC as a summarized single value for the curve has desirable properties that are independent to the threshold and invariant to a priori class probability distributions. An area of 1 represents a perfect model, and an area of 0.5 represents a worthless model. As we repeated CV ten times, we obtained ten vectors consisting of probabilities based on κ nearest neighbors' voting. The program was implemented by MATLAB 2015b (MathWorks).

Temporal Sequence Construction

A sequence is composed of lab tests, prescriptions, diagnoses, conditions, and symptoms that were given to a patient in multiple hospital admissions. We only extracted common lab tests, prescriptions, diagnoses, conditions, and symptoms (prefixed with "*l*_," "*p*_," "*d*_," "*c*_," and "*s*_," respectively). We used the International Classification of Diseases, 9th revision (ICD-9) level 3 codes instead of level 4 or 5 to avoid extreme sparsity of diagnoses. We assumed space in time between all events to be same. Then, we constructed data for incidence of a target disease as follows: for patients in which a target disease occurs, we sliced the very admission that includes the diagnosis event of a target disease out of the sequence, and used only events before that admission as a feature sequence. For other patients, we used all events. We utilized temporal information of a sequence to make a time-decayed vector representation; when we add a one-hot representation for each event, it is multiplied by the time decaying function (ie, $\exp(-\gamma t)$ with the decay constant γ) that enables to weaken the effect of old event but to strengthen the effect of recent event. A graphical illustration of this sequence and its vector representation is presented in Figure 8.

Multiparameter Intelligent Monitoring in Intensive Care-III Database

We used Multiparameter Intelligent Monitoring in Intensive Care-III (MIMIC-III) database that contains health-related data associated with 46,520 patients and 58,976 admissions to the intensive care unit of Beth Israel Deaconess Medical Center from 2001 to 2012. The database consists of detailed information about patients, including demographics such as gender, age, and race; admissions; lab test results; prescription records; procedures; and discharge ICD diagnoses. On the basis of this database, we randomly selected several common diseases (ie, diseases with relatively large number of positives) as a target disease to verify that our method can perform well in general not only for a specific disease. Then, we extracted temporal sequences and constructed following six feature vectors (*K*=6) for patients in *i*-th site: demographic information $P^i_1 \in R^{d_1 \times N_i}$, lab results $P^i_2 \in R^{d_2 \times N_i}$, diagnoses $P^i_3 \in R^{d_3 \times N_i}$, prescriptions $P^i_4 \in R^{d_4 \times N_i}$, conditions $P^i_5 \in R^{d_5 \times N_i}$, and symptoms $P^i_6 \in R^{d_6 \times N_i}$. Time decay constant γ was set to 0.01. We note that the feature vector of diagnoses in each dataset does not include its outcome of interest. Information of original datasets is described in Table 1.

To test three-site scenario, we made datasets balanced and horizontally partitioned the dataset into three, assuming data are evenly partitioned among sites (*M*=3), $P^1_k \in R^{d_k \times 125}$, $P^2_k \in R^{d_k \times 125}$, and $P^3_k \in R^{d_k \times 125}$ for every k=1,…6; federated system is needed when each institution has a limited sample size that is not enough for an analysis. In addition, from the complexity analysis, time to implement the algorithm exponentially increases in proportion to the number of patients. On the basis of these, we randomly selected and placed 125 patients in each site. Then, we predicted the incidence of five diseases independently. We set parameters for regularizers λ=0.5 and $\eta=10^{-3}$ in common. In addition, for multi-hash approach, we reduced the original dimensions for each feature to ten (ie, b_k=10 for k=2,…,6) except for the demographic feature that was reduced to two (ie, b_1=2), and for uni-hash approach we reduced the total dimensionality to the sum of projection dimensions in multi-hash approach (ie, b=52). We note that the results would be robust to the projection dimensionality unless we have too many or too few of it. Table 2 shows the results of κ-NN with κ=3 based on hamming distance for the following configurations: our system, open and closed systems with multi-hash, as well as open and closed systems with uni-hash.

Figure 8. Example of constructing temporal sequence with target disease in red and its vector representation.

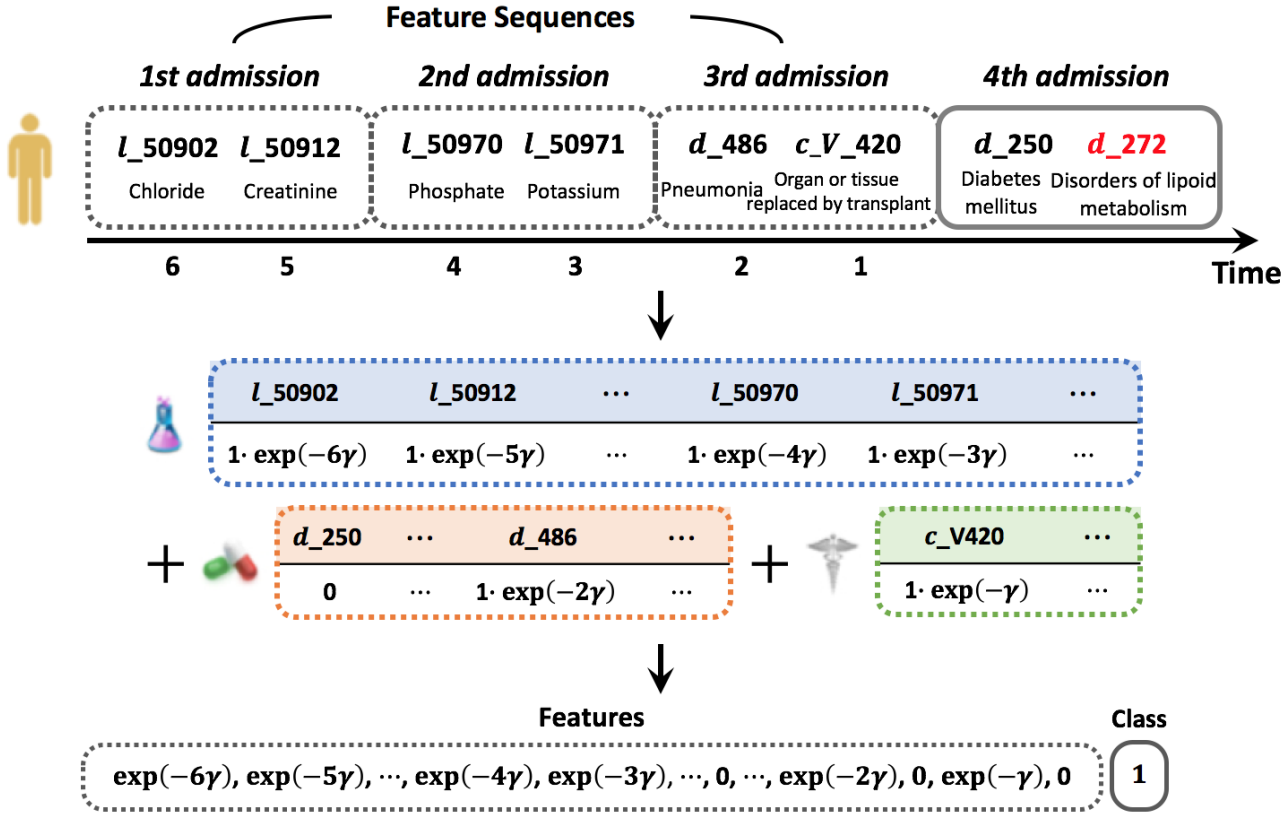


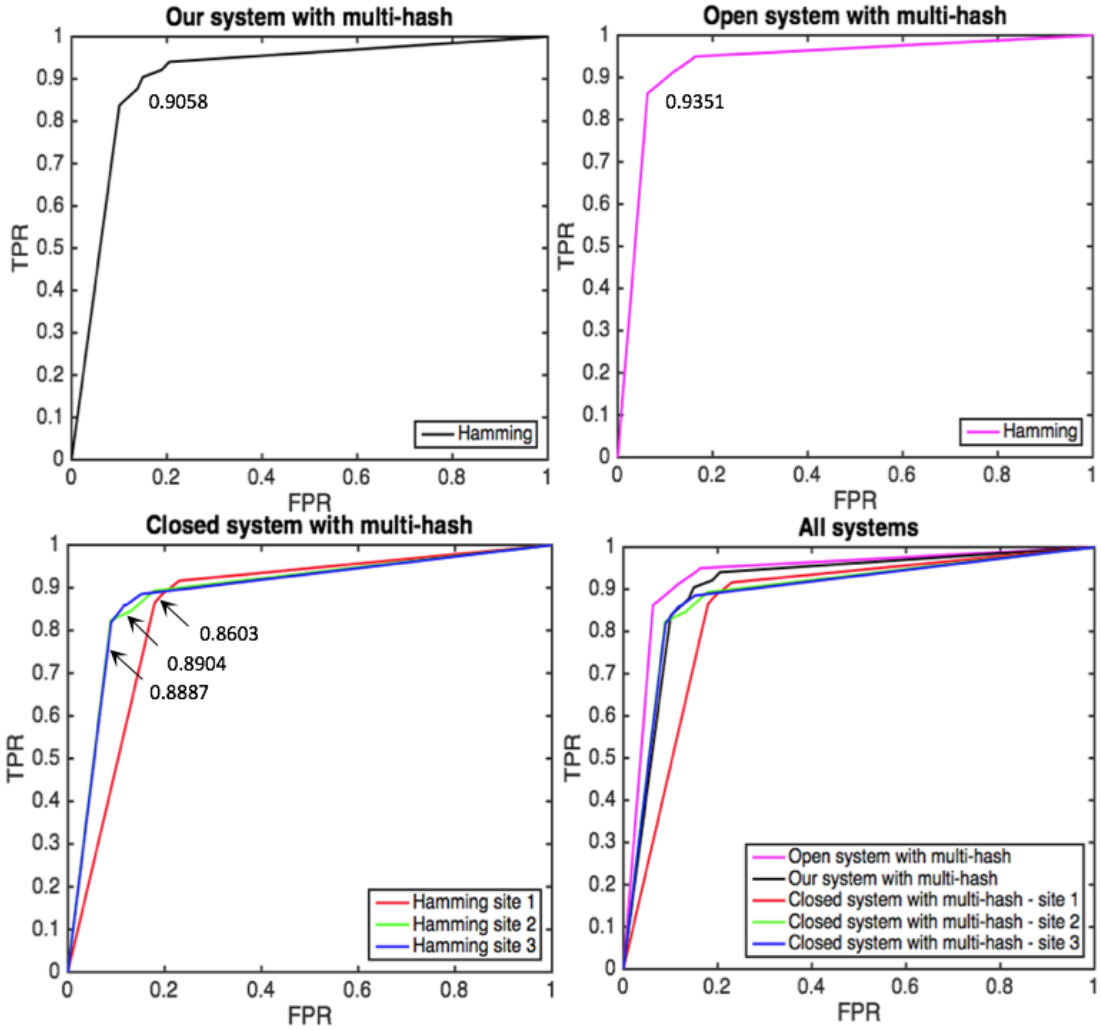
Table 1. Description of five datasets from Multiparameter Intelligent Monitoring in Intensive Care-III (MIMIC-III) database.

Disease	Data size (negative or positive)	Dimension (d_k, k=1,…,6)
Disorders of lipoid metabolism	4546/2990	(12,204,814,1338,262,170)
Hypertensive chronic kidney disease	5652/1884	(12,204,822,1338,266,169)
Cardiac dysrhythmias	3878/3658	(12,204,817,1338,263,169)
Heart failure	4167/3369	(12,204,819,1338,265,169)
Acute renal failure	4182/3354	(12,204,809,1338,268,170)

Table 2. Averaged area under the curve (AUC) with SD of κ-NN (κ=3) based on hamming distance from our, open and closed systems with multi-hash approach and from open and closed systems with uni-hash approach and based on cosine distance from open and closed systems.

Disease	Multi-hash			Uni-hash		Baseline	
	Our system, Averaged AUC (SD)	Open system, Averaged AUC (SD)	Closed system, Averaged AUC (SD)	Open system, Averaged AUC (SD)	Closed system, Averaged AUC (SD)	Open system, Averaged AUC (SD)	Closed system, Averaged AUC (SD)
Disorders of lipoid metabolism	0.9330 (0.0086)	0.9343 (0.0125)	0.9002 (0.0285)	0.9159 (0.0255)	0.8486 (0.0271)	0.8079 (0.0222)	0.7945 (0.0308)
Hypertensive chronic kidney disease	0.9078 (0.0346)	0.9283 (0.0432)	0.8538 (0.0421)	0.9270 (0.0350)	0.8501 (0.0305)	0.7823 (0.0261)	0.7762 (0.0262)
Cardiac dysrhythmias	0.9135 (0.0287)	0.9368 (0.0492)	0.8833 (0.0397)	0.9072 (0.0414)	0.8236 (0.0328)	0.7695 (0.0151)	0.7340 (0.0343)
Heart failure	0.9058 (0.0282)	0.9351 (0.0326)	0.8798 (0.0414)	0.9089 (0.0376)	0.8471 (0.0248)	0.7986 (0.0292)	0.7733 (0.0421)
Acute renal failure	0.9169 (0.0397)	0.9477 (0.0374)	0.8637 (0.0320)	0.8821 (0.0403)	0.7929 (0.0378)	0.7434 (0.0380)	0.7289 (0.0341)

Figure 9. Averaged area under the curve (AUC) of κ-NN (κ=3) for heart failure based on hamming distance from our, open and closed systems with multi-hash approach.



Additionally, Table 2 presents a baseline result based on cosine distance obtained from open and closed systems, which has the highest AUC among baseline results. We note that the results for closed systems are the average of three sites.

Figure 9 shows the comparison for heart failure of our open and closed systelabelms with multi-hash approach as an example. The prediction performance of our system is moderate between those of open and closed systems. It is encouraging that our system approaches open system without sharing local data. Figure 10 also shows the comparison result for heart failure of our system with multi-hash approach and open and closed system with uni-hash approach. We can see the superior performance of our system over closed system as before. However, in this case, our system is comparable with open system and even outperformed it for three diseases; this may come from multi-hash approach is more effective than uni-hash approach to construct context-specific hash codes. Figure 11 shows the results of our system with different κ. The detailed results with different κ are presented in Multimedia Appendix 2. AUC generally increases as κ increases.

However, in real life, different sites have a different specialty and have a different distribution in patient data. To see how our platform works in random and skewed distribution, we differentiated the ratio of samples having negative and positive classes by site. We assumed that three sites, respectively, have 10%, 30%, and 50% of positive class for five diseases. Note that all other settings including the number of sites and patients for each site, projection dimensions, and parameters were set the same as before to test only the change originated from the class imbalance and for experimental convenience; we omitted the uni-hash approach, which is expected to have the similar trend about multi-hash approach to that shown in Table 2. Table 3 shows the averaged AUC results from κ-NN with κ=3 based on hamming distance for our system, open and closed systems with multi-hash, and based on cosine distance for open and closed systems with raw data. For more elaborate comparison, F1, sensitivity (ie, TPR), and specificity (ie, the number of true negatives divided by the sum of true negatives and false positives) [58] were also measured along with AUC (Multimedia Appendix 3); F1 is the harmonic mean of precision and recall where it reaches its best value at 1 and worst at 0. It can be interpreted as weighted average of the precision (ie, the number of true positives divided by the sum of true positives and false positives) and recall (ie, TPR, sensitivity) with their equal contribution.

Figure 10. Averaged area under the curve (AUC) of κ-NN (κ=3) for heart failure based on hamming distance from our system with multi-hash approach and open and closed systems with uni-hash approach.

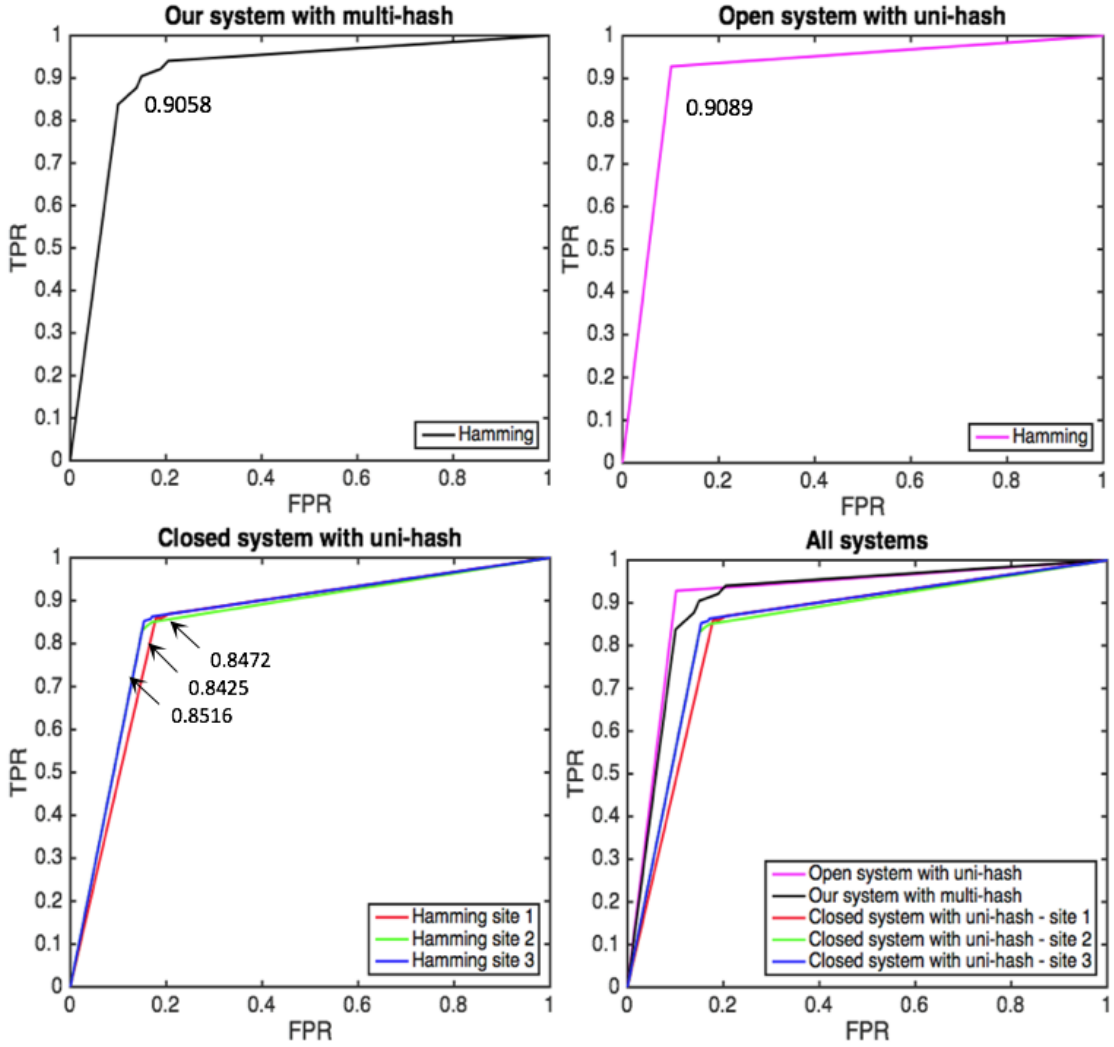


Figure 11. Averaged area under the curve (AUC) of κ-NN with different κ (κ=1,3,9) for five diseases from our system.

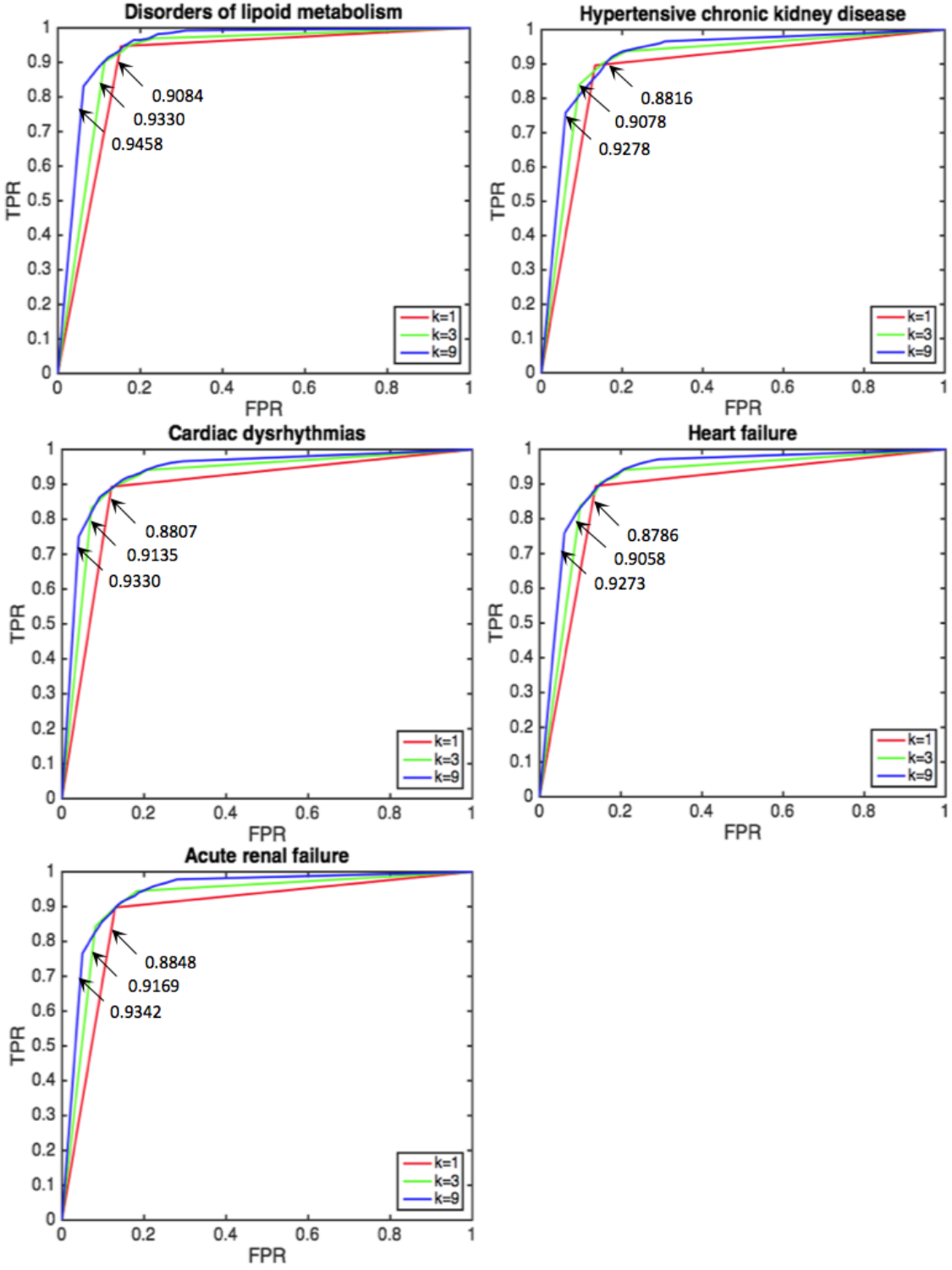


Table 3. Averaged area under the curve (AUC) with SD of κ-NN (κ=3) based on hamming distance from our, open, and closed systems with multi-hash approach and based on cosine distance from open and closed systems.

Disease	Multi-hash			Baseline	
	Our system, AUC (SD)	Open system, AUC (SD)	Closed system, AUC (SD)	Open system, AUC (SD)	Closed system, AUC (SD)
Disorders of lipoid metabolism	0.8056 (0.0386)	0.8309 (0.0412)	0.7629 (0.0295)	0.7525 (0.0212)	0.7104 (0.0187)
Hypertensive chronic kidney disease	0.7637 (0.0367)	0.7924 (0.0209)	0.7275 (0.0266)	0.7296 (0.0215)	0.7141 (0.0207)
Cardiac dysrhythmias	0.7840 (0.0301)	0.7937 (0.0228)	0.7659 (0.0223)	0.7638 (0.0198)	0.7385 (0.0188)
Heart failure	0.8287 (0.0283)	0.8832 (0.0278)	0.7459 (0.0331)	0.7735 (0.0206)	0.6778 (0.0213)
Acute renal failure	0.8239 (0.0326)	0.8704 (0.0335)	0.7558 (0.0263)	0.7304 (0.0218)	0.7415 (0.0225)

Table 4. Averaged execution time of each basic cryptographic operation for five diseases.

Operation	Time (seconds)				
	Disorders of lipoid metabolism	Hypertensive chronic kidney disease	Cardiac dysrhythmias	Heart failure	Acute renal failure
Homomorphic encryption	1.9	2.2	2.2	2.3	2.2
Initialization	5.2	6.3	5.8	6.5	6.0
Comparison	994.2	1243.9	1067.1	1131.7	1066.5
Homomorphic decryption	0.4	0.4	0.4	0.4	0.4

Most of the results can be interpreted in the same context as Table 2, but it should be noted that the degree of performance degradation in our system (~13%) is greater than that at baseline (~5%). Given these results from open and closed systems, as well as our system with multi-hash approach, accuracy might be lost because of the instability caused by updating weights $\{W_k\}^K_{k=1}$ with information from skewed distributions. However, it is encouraging that sensitivity is obtained stably in multi-hash approach rather than baseline. Sensitivity is an important measure in medical analysis because it is much more dangerous to diagnose that the disease has not occurred even though it has already developed than the opposite case. The fact that F1 is significantly larger is consistent with this. Therefore, considering all the results, we believe that our system is a useful alternative.

Next, we performed secure data aggregation and data comparison among different sites in a federated setting by which each site is able to retrieve its hamming distance under certain criteria in a privacy-preserving manner. In our experiments with balanced data, each row has 52 bits (hash code), and a 128-bit encryption key is used for homomorphic encryption. We measured the execution time of some key cryptographic operations in a workstation with an Intel 2.5 GHz CPU, where all the results are averaged over five-fold CV of total time for six cases (three test sets by two training sets). The execution time of each basic cryptographic operation has been profiled and shown in Table 4.

We confirmed that the calculated similarities across sites are the same when exchanging raw $\{H^i_k\}^K_{k=1}$ directly with each other (ie, without homomorphic encryption) or exchanging encrypted $\{H^i_k\}^K_{k=1}$ (ie, with homomorphic encryption) with each other. Therefore, the results after homomorphic encryption were obtained exactly the same as the results in Tables 2 and 3 and Figures 9 to 11 without any privacy leakage.

Discussion

Principal Findings

There are several limitations in the proposed framework. When learning hash functions, the assumption is that each site has common feature events that should be needed. However, different sites, for example, hospitals, may have different event types, and additionally, the notation system for each event type cannot be standardized except for diagnoses, symptoms, and conditions that are based on ICD-9. Even though we have the limitation of common feature events, we believe that our methodology can be still useful for cooperating hospitals eager to find similar patients across sites at the point of care. We are planning to develop a new and more practical approach to relax this assumption.

Basically, our system works better when all the participants have similar distributions. However, we have confirmed through the imbalance class experiment that our system still works well with different distributions, as well at the cost of some performance degradation. We will address more generalized imbalance data problem in future work.

Next, even if we have computational benefits by adopting a multi-hash approach compared with a uni-hash approach, and the computational complexity is not prohibitive in practice, a

technical challenge still remains in scalable hash function learning when the sample size and the feature dimensionality are large. This is because the complexity for inverting Hessian matrices in our algorithm is affected by the sample size and the feature dimensionality. This is an expensive operation of time complexity and requires a lot of memory. We can solve this problem by using parallelization or graphics processing units or utilizing a gradient descent method that replaces the inversion of Hessian matrix with a constant or a variable varying with the iteration number.

We demonstrated the feasibility of privacy-preserving similarity search, and the experiments were conducted on a single machine (with different processes) to serve as a proof of concept. In practice, we need to deploy the algorithm in multiple computers, and that is a trivial task. We will execute this algorithm using secure multiparty computation such as in the Secure Multi-pArty Computation Grid LOgistic REgression [59] in future work.

We have also listed several limitations to consider for more elaborate future work. When constructing temporal sequences, it assumes the sequence events are sampled at the same frequency for simplicity, which means the temporal effect has not been represented in this work. We roughly determined parameters of projection dimension and decay factor, which might not be optimal. In our experiment, we used 3-digit ICD to show a proof of concept, but the granularity of the ICD code will affect the performance in real applications, especially if the interest is related to the rare ones.

Conclusions

We proposed a federated patient hashing framework and developed a privacy-preserving patient similarity learning algorithm. This technique allows to learn hash codes for each patient reflecting information of different sites without sharing patient-level data. Using MIMIC-III database, we conducted experiments to demonstrate the accuracy and usability of the proposed algorithm. By utilizing the multi-hash approach, our algorithm obtained more usable and practical results than the uni-hash approach. To avoid privacy leakage in patient similarity search, we also applied homomorphic encryption able to calculate the hamming distance without transmitting hash codes. As a result, we confirmed the same results without any privacy leakage.

Acknowledgments

This work was supported by NHGRI grants R00HG008175, R01HG008802, and R01HG007078, NIGMS R01GM114612, NLM grants R00LM011392, R21LM012060, and NHLBI grant U54HL108460. The work of FW is supported by NSF IIS-1650723 and IIS-1716432.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Privacy-preserving patient representation learning in a federated setting.

[PDF File (Adobe PDF File), 113KB - medinform_v6i2e20_app1.pdf]

Multimedia Appendix 2

Prediction performance of balanced class datasets.

[PDF File (Adobe PDF File), 130KB - medinform_v6i2e20_app2.pdf]

Multimedia Appendix 3

Prediction performance of imbalanced class datasets.

[PDF File (Adobe PDF File), 33KB - medinform_v6i2e20_app3.pdf]

Multimedia Appendix 4

Prediction performance of balanced class datasets.

[PDF File (Adobe PDF File), 41KB - medinform_v6i2e20_app4.pdf]

References

1. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 2001 Aug;23(1):89-109. [Medline: 11470218]

2. Roque FS, Jensen PB, Schmock H, Dalgaard M, Andreatta M, Hansen T, et al. Using electronic patient records to discover disease correlations and stratify patient cohorts. PLoS Comput Biol 2011 Aug;7(8):e1002141 [FREE Full text] [doi: 10.1371/journal.pcbi.1002141] [Medline: 21901084]
3. Savage N. Better medicine through machine learning. Commun ACM 2012 Jan 01;55(1):17 [FREE Full text] [doi: 10.1145/2063176.2063182]
4. Sun J, Hu J, Luo D, Markatou M, Wang F, Edabollahi S, et al. Combining knowledge and data driven insights for identifying risk factors using electronic health records. AMIA Annu Symp Proc 2012;2012:901-910 [FREE Full text] [Medline: 23304365]
5. Visweswaran S, Angus DC, Hsieh M, Weissfeld L, Yealy D, Cooper GF. Learning patient-specific predictive models from clinical data. J Biomed Inform 2010 Oct;43(5):669-685 [FREE Full text] [doi: 10.1016/j.jbi.2010.04.009] [Medline: 20450985]
6. 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]
7. Ebadollahi S, Sun J, Gotz D, Hu J, Sow D, Neti C. Predicting patient's trajectory of physiological data using temporal trends in similar patients: a system for near-term prognostics. AMIA Annu Symp Proc 2010 Nov 13;2010:192-196 [FREE Full text] [Medline: 21346967]
8. 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]
9. Saria S, Rajani AK, Gould J, Koller D, Penn AA. Integration of early physiological responses predicts later illness severity in preterm infants. Sci Transl Med 2010 Sep 08;2(48):48ra65 [FREE Full text] [doi: 10.1126/scitranslmed.3001304] [Medline: 20826840]
10. Sun J, Sow D, Hu J, Ebadollahi S. A system for mining temporal physiological data streams for advanced prognostic decision support. Presented at: IEEE 10th International Conference on Data Mining (ICDM); December 13-17, 2010; Sydney, NSW, Australia URL: http://ieeexplore.ieee.org/document/5694085/
11. Bennett C, Doub T, Selove R. EHRs connect research and practice: Where predictive modeling, artificial intelligence, and clinical decision support intersect. Health Policy Technol 2012 Jun;1(2):105-114 [FREE Full text] [doi: 10.1016/j.hlpt.2012.03.001]
12. Greengard S. A new model for healthcare. Commun ACM 2013 Feb 01;56(2):17-19 [FREE Full text] [doi: 10.1145/2408776.2408783]
13. Ramakrishnan N, Hanauer D, Keller B. Mining electronic health records. Computer 2010 Oct;43(10):77-81 [FREE Full text] [doi: 10.1109/mc.2010.292]
14. Romano MJ, Stafford RS. Electronic health records and clinical decision support systems: impact on national ambulatory care quality. Arch Intern Med 2011 May 23;171(10):897-903 [FREE Full text] [doi: 10.1001/archinternmed.2010.527] [Medline: 21263077]
15. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc 2013 Jun;20(e1):e147-e154 [FREE Full text] [doi: 10.1136/amiajnl-2012-000896] [Medline: 23531748]
16. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 2013 Dec;31(12):1102-1110 [FREE Full text] [doi: 10.1038/nbt.2749] [Medline: 24270849]
17. Platt R, Carnahan RM, Brown JS, Chrischilles E, Curtis LH, Hennessy S, et al. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:1-8 [FREE Full text] [doi: 10.1002/pds.2343] [Medline: 22262586]
18. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]
19. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
20. Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc 2009;16(5):624-630. [doi: 10.1197/jamia.M3191] [Medline: 19567788]
21. Ng K, Ghoting A, Steinhubl SR, Stewart WF, Malin B, Sun J. PARAMO: a PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records. J Biomed Inform 2014 Apr;48:160-170 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.012] [Medline: 24370496]
22. Gallego B, Walter SR, Day RO, Dunn AG, Sivaraman V, Shah N, et al. Bringing cohort studies to the bedside: framework for a 'green button' to support clinical decision-making. J Comp Eff Res 2015 May 11:1-7 [FREE Full text] [doi: 10.2217/cer.15.12] [Medline: 25959863]

23. Wu Y, Jiang X, Kim J, Ohno-Machado L. Grid Binary LOgistic REgression (GLORE): building shared models without sharing data. J Am Med Inform Assoc 2012;19(5):758-764 [FREE Full text] [doi: 10.1136/amiajnl-2012-000862] [Medline: 22511014]
24. Yu H, Jiang X, Jaideep V. Privacy-preserving SVM using nonlinear kernels on horizontally partitioned data. Presented at: The ACM symposium on Applied computing; April 23-27, 2006; Dijon, France.
25. Yu H, Vaidya J, Jiang X. Privacy-preserving SVM classification on vertically partitioned data. In: Ng WK, Kitsuregawa M, Li J, Chang K, editors. Advances in Knowledge Discovery and Data Mining. PAKDD 2006. Lecture Notes in Computer Science, vol 3918. Berlin, Heidelberg: Springer; 2006:647-656.
26. Lu CL, Wang S, Ji Z, Wu Y, Xiong L, Jiang X, et al. WebDISCO: a web service for distributed cox model learning without patient-level data sharing. J Am Med Inform Assoc 2015 Nov;22(6):1212-1219 [FREE Full text] [doi: 10.1093/jamia/ocv083] [Medline: 26159465]
27. Wang F, Sun J, Ebadollahi S. Integrating Distance Metrics Learned from Multiple Experts and its Application in Patient Similarity Assessment. Presented at: The SIAM International Conference on Data Mining; April 28, 2011; Mesa, Arizona. [doi: 10.1137/1.9781611972818.6]
28. Wang F, Hu J, Sun J. Medical prognosis based on patient similarity and expert feedback. Presented at: 21st International Conference on Pattern Recognition (ICPR) 2012; November 11-15, 2012; Tsukuba, Japan.
29. Sun J, Wang F, Hu J, Edabollahi S. Supervised patient similarity measure of heterogeneous patient records. SIGKDD Explor Newsl 2012 Dec 10;14(1):16-24. [doi: 10.1145/2408736.2408740]
30. Wang F, Sun J, Ebadollahi S. Composite distance metric integration by leveraging multiple experts' inputs and its application in patient similarity assessment. Stat Anal Data Min 2012 Feb 17;5(1):54-69. [doi: 10.1002/sam.11135]
31. Gentry C. Fully homomorphic encryption using ideal lattices. Presented at: Proceedings of the 41st Annual ACM Symposium on Theory of computing; May 31-June 2, 2009; Bethesda, MD. [doi: 10.1145/1536414.1536440]
32. Bloom BH. Space/time trade-offs in hash coding with allowable errors. Commun ACM 1970 Jul;13(7):422-426 [FREE Full text] [doi: 10.1145/362686.362692]
33. Schneier B. 2004 Aug. Schneier on Security: Cryptanalysis of MD5 and SHA: Time for a New Standard URL: https://www.schneier.com/essays/archives/2004/08/cryptanalysis_of_md5.html [accessed 2018-03-02] [WebCite Cache ID 6xcTiZQHw]
34. Wang J, Shen HT, Song J, Ji J. 2014 Aug. Hashing for Similarity Search: A Survey URL: http://arxiv.org/abs/1408.2927 [accessed 2018-03-03] [WebCite Cache ID 6xdEF5JTh]
35. Indyk P, Motwani R. Approximate nearest neighbors: towards removing the curse of dimensionality. Presented at: Proceedings of the Thirtieth Annual ACM Symposium on Theory of computing; May 24-26, 1998; Dallas, Texas. [doi: 10.1145/276698.276876]
36. Broder AZ, Charikar M, Frieze AM, Mitzenmacher M. Min-Wise Independent Permutations. J Comput Syst Sci 2000 Jun;60(3):630-659. [doi: 10.1006/jcss.1999.1690]
37. Liu W, Wang J, Chang SF, Kumar S. Hashing with graphs. Presented at: Proceedings of the 28th International Conference on Machine Learning; June 28, 2011; Bellevue, Washington URL: http://www.icml-2011.org/papers/6_icmlpaper.pdf
38. Gong Y, Lazebnik S. Iterative quantization: a procrustean approach to learning binary codes. Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 20-25, 2011; Colorado Springs, CO. [doi: 10.1109/CVPR.2011.5995432]
39. Kong W, Li WJ. Isotropic hashing. In: Bartlett P, Weinberger KQ, Burges CJ, Bottou L, Pereira FC, editors. Advances in Neural Information Processing Systems 25. Red Hook, NY, USA: Curran Associates, Inc; 2012.
40. Gong Y, Kumar S, Verma V, Lazebnik S. Angular quantization-based binary codes for fast similarity search. Presented at: Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS) - Volume 1; December 3-6, 2012; Lake Tahoe, Nevada URL: http://paperpile.com/b/4F2MYV/noQb
41. Wang J, Kumar S, Chang SF. Semi-supervised hashing for scalable image retrieval. Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010; June 13-18, 2010; San Francisco, CA. [doi: 10.1109/CVPR.2010.5539994]
42. Wang J, Kumar S, Chang SF. Semi-supervised hashing for large-scale search. IEEE Trans Pattern Anal Mach Intell 2012 Dec;34(12):2393-2406 [FREE Full text] [doi: 10.1109/TPAMI.2012.48] [Medline: 22331853]
43. Yang X, Fu H, Zha H, Barlow J. Semi-supervised Nonlinear Dimensionality Reduction. 2006 Presented at: The 23rd International Conference on Machine Learning; June 25-29, 2006; Pittsburgh, Pennsylvania. [doi: 10.1145/1143844.1143978]
44. Liu W, Wang J, Ji R, Jiang YG, Chang SF. Supervised Hashing with Kernels. Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2012; June 16-21, 2012; Providence, RI.
45. Wang J, Kumar S, Chang SF. Sequential projection learning for hashing with compact codes. Presented at: The 27th International Conference on Machine Learning (ICML) 2010; June 21-24, 2010; Haifa, Israel.
46. Bronstein MM, Bronstein AM, Michel F, Paragios N. Data fusion through cross-modality metric learning using similarity-sensitive hashing. Presented at: IEEE Computer Society Conference on Computer Vision and Pattern Recognition; June 13-18, 2010; San Francisco.
47. Kumar S, Udupa R. Learning hash functions for cross-view similarity search. Presented at: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI) 2011; July 16-22, 2011; Barcelona, Spain p. 1360.

48. Song J, Yang Y, Huang Z, Shen HT, Hong R. Multiple Feature Hashing for Real-time Large Scale Near-duplicate Video Retrieval. Presented at: Proceedings of the 19th ACM International Conference on Multimedia; November 28-December 1, 2011; Scottsdale, AZ.
49. Zhen Y, Yeung DY. A probabilistic model for multimodal hash function learning. Presented at: Proceedings of tThe 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 2012; August 12-16, 2012; Beijing, China.
50. Zhen Y, Yeung DY. Co-Regularized Hashing for Multimodal Data. Presented at: Proceedings of the 25th International Conference on Neural Information Processing Systems 2012 (NIPS 2012); December 3-6, 2012; Lake Tahoe, Nevada p. 1376-1384 URL: http://paperpile.com/b/4F2MYV/6Qrp
51. Song J, Yang Y, Yang Y, Huang Z, Shen HT. Inter-media hashing for large-scale retrieval from heterogeneous data sources. Presented at: Proceedings of the 2013 International Conference on Management of Data - SIGMOD '13; June 22-27, 2013; New York, NY, USA. [doi: 10.1145/2463676.2465274]
52. Zhu X, Huang Z, Shen HT, Zhao X. Linear cross-modal hashing for efficient multimedia search. Presented at: Proceedings of the 21st ACM International Conference on Multimedia; October 21-25, 2013; Barcelona, Spain p. 143-152 URL: http://paperpile.com/b/4F2MYV/LBef [doi: 10.1145/2502081.2502107]
53. Changpinyo S, Liu K, Sha F. Similarity component analysis. Presented at: Neural Information Processing Systems 2013 (NIPS 2013); October 2013; Lake Tahoe, Nevada p. 1511-1519 URL: http://www-scf.usc.edu/~kuanl/papers/nips13_sca.pdf
54. Horn RA, Johnson CR. Norms for vectors and matrices. In: Horn RA, editor. Matrix Analysis. Cambridge, England: Cambridge University Press; 1990:313-386.
55. Ortega JM, Rheinboldt WC. Iterative Solution of Nonlinear Equations in Several Variables. Philadelphia: SIAM; 1970.
56. Zehfuss G. Ueber eine gewisse determinante? J Appl Math Phys 1858;3:298-301.
57. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000 Jun 13;101(23):e215-e220 [FREE Full text] [doi: 10.1161/01.CIR.101.23.e215]
58. Han H, Wang WY, Mao BH. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. Presented at: Proceedings of the 2005 International Conference on Advances in Intelligent Computing (ICIC); August 23-26, 2005; Hefei, China p. 878-887.
59. Shi H, Jiang C, Dai W, Jiang X, Tang Y, Ohno-Machado L, et al. Secure Multi-pArty Computation Grid LOgistic REgression (SMAC-GLORE). BMC Med Inform Decis Mak 2016 Dec 25;16 Suppl 3:89 [FREE Full text] [doi: 10.1186/s12911-016-0316-1] [Medline: 27454168]

Abbreviations

AUC: area under the curve
CS: cloud server
CSP: crypto service provider
CV: cross validation
DC: data custodian
EHR: electronic health record
ICD-9: International Classification of Diseases, 9th revision
MIMIC-III: Multiparameter Intelligent Monitoring in Intensive Care-III
TPR: true positive rate

Edited by G Eysenbach; submitted 29.03.17; peer-reviewed by Y Luo, M Noman, K Marsolo; comments to author 22.07.17; revised version received 12.09.17; accepted 06.01.18; published 13.04.18.

Please cite as:
Lee J, Sun J, Wang F, Wang S, Jun CH, Jiang X
Privacy-Preserving Patient Similarity Learning in a Federated Environment: Development and Analysis
JMIR Med Inform 2018;6(2):e20
URL: http://medinform.jmir.org/2018/2/e20/
doi:10.2196/medinform.7744
PMID:29653917

Review

Reasons For Physicians Not Adopting Clinical Decision Support Systems: Critical Analysis

Saif Khairat[1,2], PhD, MPH; David Marc[3], PhD; William Crosby[2], BSN, RN; Ali Al Sanousi[4], MD, PhD

[1]Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[3]Health Informatics Graduate Program, College of Saint Scholastica, Duluth, MN, United States

[4]Hamad Medical Coperation, Doha, Qatar

Corresponding Author:
Saif Khairat, PhD, MPH
Carolina Health Informatics Program
University of North Carolina at Chapel Hill
428 Carrington Hall
Chapel Hill, NC, 27514
United States
Phone: 1 9198435413
Email: saif@unc.edu

Abstract

Background: Clinical decision support systems (CDSSs) are an integral component of today's health information technologies. They assist with interpretation, diagnosis, and treatment. A CDSS can be embedded throughout the patient safety continuum providing reminders, recommendations, and alerts to health care providers. Although CDSSs have been shown to reduce medical errors and improve patient outcomes, they have fallen short of their full potential. User acceptance has been identified as one of the potential reasons for this shortfall.

Objective: The purpose of this paper was to conduct a critical review and task analysis of CDSS research and to develop a new framework for CDSS design in order to achieve user acceptance.

Methods: A critical review of CDSS papers was conducted with a focus on user acceptance. To gain a greater understanding of the problems associated with CDSS acceptance, we conducted a task analysis to identify and describe the goals, user input, system output, knowledge requirements, and constraints from two different perspectives: the machine (ie, the CDSS engine) and the user (ie, the physician).

Results: Favorability of CDSSs was based on user acceptance of clinical guidelines, reminders, alerts, and diagnostic suggestions. We propose two models: (1) the user acceptance and system adaptation design model, which includes optimizing CDSS design based on user needs/expectations, and (2) the input-process-output-engagemodel, which reveals to users the processes that govern CDSS outputs.

Conclusions: This research demonstrates that the incorporation of the proposed models will improve user acceptance to support the beneficial effects of CDSSs adoption. Ultimately, if a user does not accept technology, this not only poses a threat to the use of the technology but can also pose a threat to the health and well-being of patients.

(JMIR Med Inform 2018;6(2):e24) doi:10.2196/medinform.8912

KEYWORDS

decision support systems, clinical; decision making, computer-assisted; attitude to computers

Introduction

The Agency for Healthcare Research and Quality [1] promotes a systems approach that aims "to catch human errors before they occur or block them from causing harm." Clinical decision support systems (CDSSs) are at the forefront of this aim. A CDSS provides alerts, reminders, prescribing recommendations, therapeutic guidelines, image interpretation, and diagnostic assistance. Although studies have shown that CDSSs reduce medical errors and improve outcomes, they also demonstrate that CDSSs fall short of their full potential [2-9]. Research has attempted to narrow in on the cause of this shortfall. Coiera [10]

identified provider's lack of willingness and ability to use the technological system as one of the primary reasons.

Wendt et al [11] discussed several factors that may be related to the acceptance of CDSSs, including the relevance of the information provided by the system, perceived validity of the system, and the work and time expended on using the system. These factors are similar to those defined by Davis [9] in the technology acceptance model (TAM) and later refined by Venkatesh et al [12] in the unified theory of acceptance and use of technology (UTAUT). These models offer a potential explanation for how expectations of performance, effort, social influences, and facilitating conditions are determinants of user acceptance and technology usage [12]. Using the TAM, Van Schaik et al [13] evaluated a gastroenterology referral CDSS. The system assisted primary care providers by suggesting an appropriate subspecialty referral (medical vs surgical), prioritizing urgency, and offering real-time booking [13]. They found that physicians rated acceptance based on the potential merits of the system rather than their experience with the computer system [13]. This finding was concordant with Venkatesh et al's [12] proposal of the UTAUT model, in which they demonstrated that performance expectancy is the strongest predictor of user acceptance of technology.

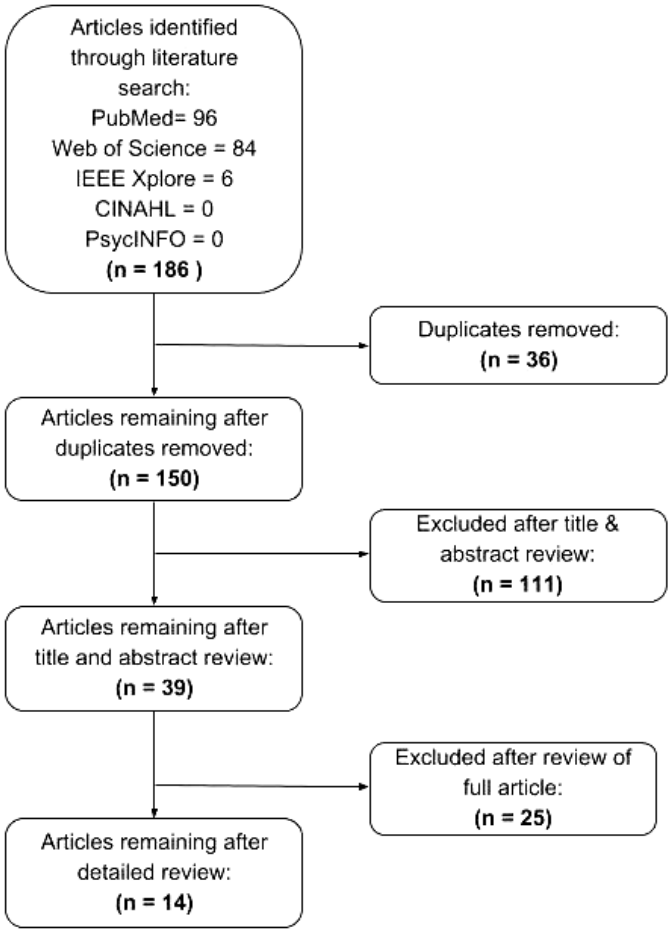
The theory behind user acceptance and its impact on the adoption of technology has been thoroughly described. The purpose of this paper was to conduct a review of the literature in order to evaluate our hypothesis that meaningful engagement of physicians in the design and development of CDSSs with transparent decision-making processes will result in higher acceptance rates.

Methods

Critical Review

A search of MEDLINE/PubMed, CINAHL, PsycInfo, IEEE Xplore, and Web of Science was conducted using the keywords "clinical decision support," "decision support acceptance," and "user acceptance." No timeframe limits were included for any database, and the language filters were set to English studies only. In our initial search, we found 186 papers. After removal of duplicates, 150 studies remained. To be included in this review, the papers had to match the following inclusionary criteria: investigate human interaction with a CDSS and evaluate user acceptance using the TAM questionnaire, focus groups, or interviews. Papers were excluded if the focus was on decision support systems that did not include clinical care or if they did not empirically investigate user acceptance. Title and abstract review eliminated 111 studies. The remaining 39 studies underwent a full-text review, resulting in a final count of 14 studies that met inclusion criteria. The search results are summarized in Figure 1.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Diagram.



Study findings were categorized as either showing favorable or unfavorable responses to CDSSs. The favorable and unfavorable categorization was based on interpretations of focus groups and interviews conducted by the researchers of the reviewed papers. Additionally, the type of CDSS was noted for each of the reviewed papers. If a study used the TAM questionnaire, the results were summarized separately.

Task Analysis

To gain a greater understanding of the problems associated with CDSSs, we conducted a task analysis. Using past research, the task analysis helped identify and describe the goals, user input, system output, knowledge requirements, and constraints from two different perspectives. We considered the perspective of the machine (ie, the CDSS engine). We also considered the perspective of the user (ie, the physician). The literature review and task analysis served as the basis for designing CDSS models that improved user acceptance.

Results

Critical Review

The results of 14 articles were evaluated. The 11 articles that qualitatively evaluated user acceptance of CDSSs can be found in Table 1 and the three articles that quantitatively evaluated user acceptance of CDSSs using TAM can be found in Table 2. Favorable and unfavorable responses for the aspects of clinical guidelines, reminders, and diagnostic CDSSs were recorded. Favorable responses were due to ease of system use, perceived time savings, and perceived usefulness of the systems in improving care delivery and overall patient health [14]. Users with higher computer skills were reported to have greater acceptance; however, the majority of users had an unfavorable acceptance response [15]. These unfavorable responses were often related to workflow interference, questionable validity of the systems, excessive disturbances caused by the systems, and lack of efficiency. More specifically, the workflow constraints were related to the CDSSs causing excessive alerts, increased time in computer handling, and decreased face-to-face time with patients [14,16,17].

Of the studies reviewed, three used the TAM questionnaire for assessing user acceptance of CDSSs (Table 2). The CDSSs in these studies included an evaluation of two different computerized clinical guideline systems and one that offered reminders and alerts for evidenced-based guidelines. The ranges on perceived usefulness all overlap, but on the CDSSs with the highest perceived usefulness it also has the highest perceived ease of use. Overall, the TAM questionnaire revealed moderate user acceptance on all scales. In terms of the relationship between user acceptance of CDSSs and patient safety, none of the reviewed papers evaluated this topic. However, Bergman and Fors [15] found that the use of the technology was relatively low when user acceptance was low.

Table 1. Summary of user acceptance related to clinical decision support systems (CDSSs) from previous studies (N=11).

Study	Favorable response to CDSS	Unfavorable response to CDSS	CDSS Description
Bergman & Fors (2005) [15]	Can save time and provide structure	Not suitable to workflow and there is the risk of becoming dependent	CDSS for medical diagnosis of psychiatric diseases
Curry & Reed (2011) [16]	Concept was supported	Interference with workflow and questionable validity	Prompts for adhering to diagnostic imaging guidelines
Gadd et al (1998) [18]	Easy to use, limits the need for data entry, accurate, and relevant	Benefits are lost because it takes so long to use	Internet-based system that interactively presents clinical practice guidelines at point of care
Johnson et al (2014) [19]	Longitudinal acceptance behavior, perceived ease of use, and perceived usefulness	Computer literacy, user satisfaction, and general optimism	Clinical reminders and alerts for patients with asthma, diabetes, hypertension, and hyperlipidemia
Rosenbloom et al (2004) [20]	Can improve efficiency and quality of care; enhances education	Senior physicians did not think it was necessary	CDSS for computerized order entry system
Rousseau et al (2003) [21]	Use of “active” CDSS can bridge the gap between own practice and best practice	Clinicians found it to be difficult to use and unhelpful clinically	CDSS for chronic disease in general practice
Shibl et al (2013) [22]	Performance expectancy, usefulness, and effort expectancy	Trust in CDSS and need for the system	No specified CDSS; responses based on past and present experiences with multiple CDSSs
Sousa et al (2015) [23]	Belief that the suggestions were good for the patient	Low confidence in the evidence	CDSS for nursing care plan
Terraz et al (2005) [24]	Ease of use and easy access to information	Information that is presented is already known	Guidelines for colonoscopies
Wallace et al (1995) [25]	Can improve patient outcomes	Alerts are ignored because there is not enough time to dedicate to forming an appropriate response	CDSS to standardize administration of supplemental oxygen
Zheng et al (2005) [17]	Improves performance leading to better care, easy to use, and efficient	Iterative advisories, lack of relevance, a lot of data entry, and disruptive	Clinical reminders for chronic diseases and preventive care

Table 2. Results of the technology acceptance model (TAM) questionnaire from prior studies evaluating user acceptance of CDSSs.

Study	Buenestado et al (2013) [26]	Heselmans et al (2012) [27]	Peleg et al (2009) [28]
CDSS description	Computerized clinical guidelines and protocols for asthma in children	Reminders and alerts for evidenced-based guidelines	Guideline-based decision support system for diabetic patient foot problems
Participant description	8 pediatricians	39 Dutch-speaking family physicians	8 family physicians
Likert scale score[a], mean (SD)	Seven-point scale	Seven-point scale	Five-point scale
Perceived usefulness	5.80 (1.24)	4.00 (1.37)	4.00 (0.71)
Perceived ease of use	6.17 (0.92)	5.02 (1.41)	4.40 (0.59)
Attitude toward using	6.21 (0.59)	4.84 (0.97)	N/A
Behavioral intention to use	5.71 (1.24)	5.91 (1.33)	4.88 (0.23)

[a]The scores are based on a Likert scale (1=totally disagree; 5 or 7=totally agree).

Task Analysis

Task analysis is conducted to stay updated with the changing professional practice (ie, health information technology) [29]. Task analysis applied to representative populations strengthen health systems by systematically evaluating the skills, knowledge, and behavior of clinicians that impact clinical practice [30]. The use of CDSS in health care has introduced new dynamics to practice and requires task analysis to understand the perception of users to this new technology. For that reason, conducting a task analysis will improve adoption levels. A task analysis includes goals, input, process, and output. The next sections discuss the purpose of each task analysis stage.

Goals

The goal of a CDSS is to supplement the physician as the sole information processor in clinical decision making and thereby aid in the reduction of medical errors. Yet, there is still much room for improvement. In part, this shortcoming may be due to the lack of physician acceptance of the CDSSs in supplementing their decision making. To get a better understanding of the challenges in creating clinical decision processes, we first consider what information goes into this process.

Input

A CDSS is based on an input-process-output (IPO) model. The inputs for the CDSS process include patient-specific information such as diagnoses, medications, symptoms, laboratory data, demographics, and other clinically relevant information. The inputs for knowledge-based CDSSs are often determined by clinical guidelines, whereas non-knowledge-based CDSSs use the most relevant information assessed by algorithm performance.

Process

The CDSS process takes two different forms: knowledge based and non-knowledge based [31]. Knowledge-based systems are governed by a set of rules. Non-knowledge-based systems, on the other hand, use a computer as the central processing unit to learn from historical information. As a result, these systems typically utilize machine learning algorithms.

When CDSSs offer clinical suggestions, the support, evidence, clinical guideline, or algorithm for those suggestions is not provided. The inputs for knowledge-based CDSSs are often determined by clinical guidelines, whereas non-knowledge-based CDSSs use the most relevant information assessed by algorithm performance. In both cases, the physician is not aware of the inputs or processes the CDSS utilizes. Thus, the CDSS is a black box to the physician.

Physicians make clinical decisions based on the same patient information in addition to social structures (acceptable behavior as determined by peer groups), institutions (the requirement to act according to mandated practices), and individual morality in decision making. One can conjecture that difficulties arise when automating such a complex network of inputs that could never by fully encapsulated or realized by a machine.

Output

The output from CDSSs and physicians are a result of the methods employed for processing the inputs. The output may be a diagnosis, procedure, prescription, etc. Ideally, in conditions where the computer and the physician are presented with the same information, the output from the CDSS should mirror the physician's decision.

The level of control the CDSS has with regards to the output is inversely related to the level of control the user has over the output. A CDSS can be passive in situations where they only "highlight" information for the user, but do not request acknowledgment or action [32]. An example would be presenting abnormal laboratory values in a red font and normal laboratory values in a black font. Active CDSSs act independently and provide suggestions to guide the physician's behavior [32]. An example would be a system that provides diagnostic assistance. The type of output then depends on the goal orientation of a task (eg, diagnoses, medication alerts, and clinical guidelines for preventive care).

Knowledge

Physicians are more likely to accept a CDSS if the system matches their own decision-making processes. Forster [33] described how humans quickly act on information by using bounded and ecological rationality. Bounded rationality is based on the use of simple heuristics, allowing for fast, real-time

decision making [34]. Ecological rationality is based on rational beliefs of things in a given environmental setting where conditions are fluid. Forster [33] argued that both bounded and ecological rationality need to be present in machine learning to mimic the human decision processes.

Incorporating these two approaches into CDSSs can be challenging. Even though the heuristics that mediate decision processes are simple; the complexity of the cognitive infrastructure underlying heuristic operations can be difficult to implement. Still, Forster [33] argued that machine learning algorithms can be improved by incorporating the principles of bounded and ecological rationality. To carry out this task, Forster [33] suggested that a decision-making machine should have (1) a set of ad hoc rules (or biases) to act on and (2) a set of ecologically viable environmental factors to consider.

Clark [35] extended this idea of mediating decision processes by bounded and ecological rationality through a concept he referred to as scaffolding. Clark [35] posited that human reasoning involves three aspects: (1) individual reasoning cast by some form of fast, pattern-completing style of computation (ie, bounded rationality); (2) substantial problem-solving work offloaded onto external structures and processes (eg, social and institutional structures); and (3) public language used as a means of coordinating social structures and mediating individual thought. Thus, decision making and cognition are largely dependent on the capacity to dissipate reasoning throughout the environment to reduce individual workload.

Holland and colleagues [36] added additional elements that can be useful to understand physician decision making. These elements provide a cognitive framework for problem solving, which includes two distinct schemas: pragmatic reasoning schema and problem schema. Pragmatic reasoning schemas are clusters of abstract inferential rules that characterize relations over general classes of object kinds, event relationships, and problem goals. Problem schemas are used by experts to solve routine problems, where an expert retrieves an appropriate problem schema and provides it with problem-specific parameters.

The system must also have two types of knowledge structures: mental models and condition-action rules. Holland and colleagues [36] assert that "mental models are transient, dynamic representations of particular, unique situations. They exist only implicitly, corresponding to the organized, multifaceted description of the current situation and the expectations that flow from it." A condition-action rule can be thought of as an IF (condition)...THEN (action) statement. Together, these knowledge structures allow the mental schemas to operate in order to solve problems.

To successfully implement and use CDSSs, these mental models have to be identified. Hayek [37] stated that knowledge is not given to anyone in its entirety. This statement legitimizes why CDSSs are so important. In theory, CDSSs lessen the cognitive resources a physician needs to make decisions.

Constraints

The major limitation of CDSSs is that scaffolding cannot be fully captured by computers. The environmental, clinical, and social constraints in which physicians practice are difficult to include as inputs into a CDSS. In addition, reproducing a physician's tacit knowledge through mental models and condition-action rules is a formidable objective. Additionally, physicians must be able to support their decision and are skeptical of recommendations or claims that lack supporting evidence or transparency. The fact that CDSSs do not reveal how output decisions are made may be a driving force behind the lack of users' acceptance.

Discussion

Means for Solving User Acceptance of Clinical Decision Support Systems

Studies have revealed that responses to CDSSs can be unfavorable when resulting improvements in patient outcomes are inconsistent [6]. Also, some studies have reported incidents of patient harm associated with CDSS implementation [38]. Despite these findings, limited research has formally evaluated the impact of user acceptance. Based on our comprehensive review of the literature, we have found both favorable and unfavorable user acceptance to CDSSs.

If a user finds a product frustrating or perceives that the purpose of the product is to limit autonomy, the user may not use the product or do so inappropriately [39]. Vashitz et al [40] explains the consequence of loss of autonomy as reactance. *Reactance* is an unpleasant motivational state whereby people react to situations to retain freedom and autonomy. Reactance may exist when physicians feel threatened by clinical reminders for fear that they are losing autonomy and freedom of choice in the presence of such systems. Physicians may have the perception that these systems are meant to replace or degrade their clinical duties. Vashitz et al [40] describe how unsolicited advice may lead to a reactance state if the advice contradicts a person's original impression of choice options.

Based on the UTAUT, user expectations need to be taken into consideration for technology to be accepted [12]. Therefore, in the design of CDSSs, the human element cannot be ignored. Reminders and alerts should be presented in such a way that the user does not find them threatening or obtrusive. User needs and expectations of a CDSS should be evaluated early and throughout the development lifecycle. For instance, Gadd [18] observed enhanced usability and usefulness by implementing usability testing in the early phases of CDSS development. They evaluated an evolving prototype of the system and observed user interactions over a 3-month period. In a series of sessions, they focused on evaluating user interactions with different sets of system features such as screen layout, input/output, and links to educational materials. Finally, they considered the user feedback on system recommendations in the design process. Compelling suggestions for system enhancements made by users during the earlier sessions influenced system development of features that were evaluated in later sessions.

Peleg et al [28] discussed the development process of their CDSS, where clinically knowledgeable users worked alongside the developers to design and implement the CDSS. They also used a lifecycle model user-centered design and evaluation

process for evaluating the users' goals/expectations, workflow, environmental constraints, and tasks. Finally, they conducted usability testing (ie, heuristic evaluation of user-interface, keystroke-level modeling, and cognitive walkthroughs) prior to implementation.

Developers of CDSSs have attempted to bottle-up the decision-making capacity of physicians and place that knowledge into a computer. Current methods to achieve this feat employ rules and machine learning algorithms. However, the lack of user acceptance has impeded CDSS use. Research has shown that consideration of users' needs and expectations in the design of the CDSS may help overcome this obstacle. We argue that this approach is only part of the solution.

We propose that CDSSs move away from the black-box process to a more transparent method within the IPO model. Simply put, tell the physician how the computer is making the decision. If the computer can become part of scaffolded knowledge, the physician may view the computer as an aid rather than a threat or hindrance. Research supports the idea that the rules governing alerts be specified to practitioners and the information be presented based on users' needs and expectations [41].

Proposal of Models to Gain User Acceptance

We propose two models to improve CDSSs development that may lead to increased utilization resulting in improved patient outcomes. First, is the user acceptance and system adaptation design (UASAD) model that aims to involve end users early in the design and throughout the development of CDSSs. Second, is replacing the current IPO model of CDSS development with the input-process-output-engage (IPOE) model that serves to "engage" the physician through CDSS process transparency.

The UASAD model demands early end-user involvement in CDSS development. User needs and expectations need to be fully realized prior to the development of a CDSS. Another consideration is to evaluate system preparedness to ensure that users can trust the security and privacy of the system. Prototypic designs should undergo an iterative design process following rigorous usability testing in a laboratory and natural setting (ie, pilot study) to ensure that the system works within the cognitive and environmental constraints with which the user functions.

Finally, user acceptance should be evaluated to ensure that the system is used appropriately. If user acceptance is not achieved above a predefined threshold, the CDSS should be reevaluated from the point of view of user needs and expectations. It should also be subjected to adaptive redesign. This process should iterate until user acceptance exceeds a predefined threshold. To illustrate this process, we have developed a UASAD model (Figure 2). The purpose of the model is to include the user as the focal point of the design process of CDSS.

The IPOE model offers users a window into the black-box IPO process. Through "engage" physicians will see how the CDSS is making decisions. The IPOE window will be called "engage" because it will present users with the rules that the machine followed to generate the output (Figure 3). Therefore, the user can make informed decisions when determining to accept or deny outputs. "Engage" will display the input, process, and output that led to the CDSS's decision. The physician will then be able to evaluate the relevancy, validity, supporting evidence, and strength of a recommendation. Therefore, this system becomes a component of the physician's scaffolded knowledge and enables them to act more confidently in accepting the technology and its role in their decision-making processes.

A limitation of the IPOE model is that in order for the model to work successfully, the physician has to understand the process. Processes that utilize a machine learning algorithm, such as neural networks, do not provide rules. Therefore, it is challenging to make all processes transparent.

Why Do We Make Bad Decisions?

Physicians' tendencies to incorrectly process challenging decisions usually lead to bad clinical decisions. Most practicing physicians tend to make decisions out of their own medical experience, whereas others pursue medical consultations and filtering through the jargon of relevant research. The most effective physician, though, is the one who has the ability to utilize his clinical judgment coupled by the computerized decision support tools to leverage the power of CDSSs. Most clinicians exhibit bias when it comes to medical information that they know and, therefore, they typically focus on things that would agree with the specific clinical outcome that they want to see in their patients. Therefore, the context of using a CDSS is mandated by the efforts to decrease medical errors by utilizing existing knowledge and technology. These systems are a result of long-term scientific research to build efficient tools for physicians to supplement their clinical experience. Physicians should look at CDSSs as an added value to make the best decisions in their day-to-day practice and to better serve their patients. These systems seek to reduce medical errors by enabling the practicing physicians to make informed decisions that are both accurate and precise.

Figure 2. A user acceptance and system adaptation design (UASAD) model. CDSS: clinical decision support system; UTAUT: unified theory of acceptance and use of technology.

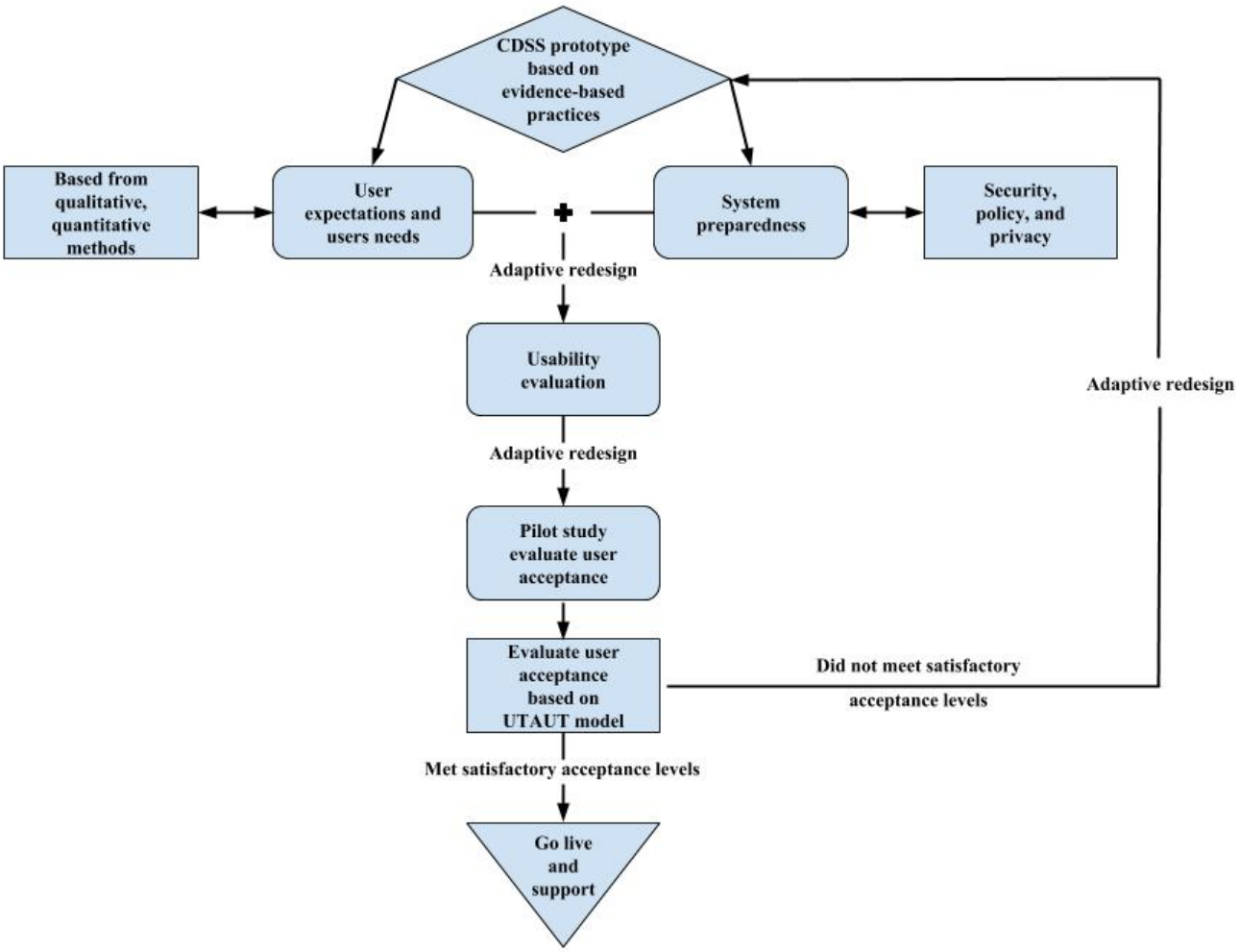
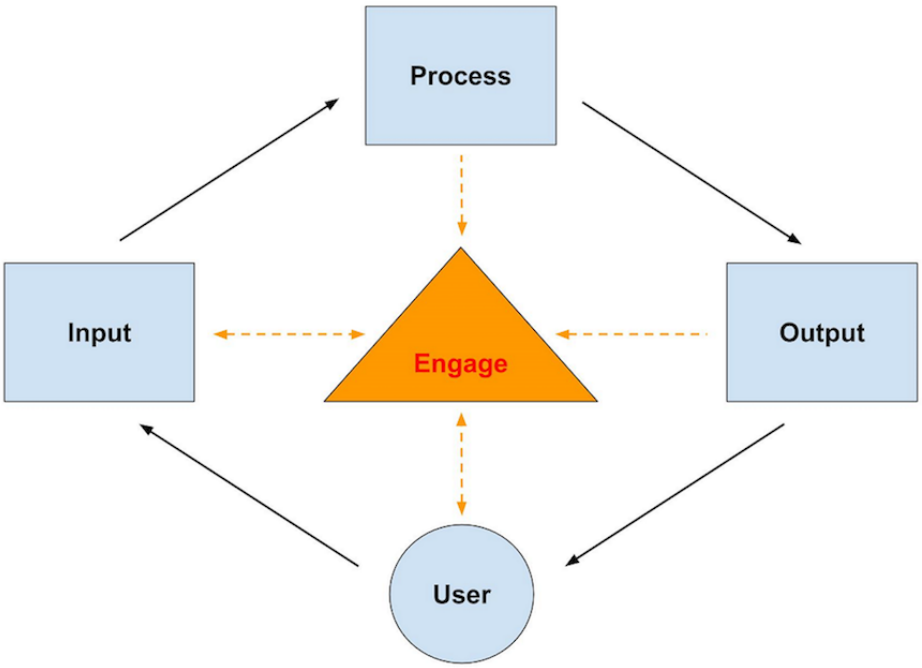


Figure 3. The input-process-output-engage (IPOE) model.



Conclusion

Implementation of CDSSs has demonstrated increased efficiency, reduced medical errors, and improved outcomes, but they continue to fall short of their full potential [2-9]. We believe this key shortcoming may partly be due to the lack of physician acceptance. In the past, CDSS designs have not incorporated input from physicians and do not reveal their decision-making processes. Consequently, many physicians are hesitant to accept CDSSs leading to suboptimal implementation. Here we propose two models for designing CDSSs with the goal of improving efficacy and physician acceptance. One model, UASAD, focuses on including the physician in the design process by examining user needs and expectations and usability of prototypic designs. The other model, IPOE, extends the existing IPO framework by adding an "engage" stage that displays the CDSS process to the physician. This approach allows the physician to include the CDSS as a component of their decisions while maintaining professional autonomy. There is still considerable work to be done for validating these models, yet user acceptance appears to be pertinent for successful CDSS use. Ultimately, if a physician does not accept the technology, it not only poses a threat to the use of the technology but can also pose a threat to the health and well-being of patients.

Acknowledgments

The authors would like to express appreciation to Dr Marla Broadfoot, PhD, and Ms Victoria Rand for their contribution to this paper.

Conflicts of Interest

None declared.

References

1. Patient Safety Network. Patient safety primer: systems approach URL: https://psnet.ahrq.gov/primers/primer/21/systems-approach [WebCite Cache ID 6xRITvchJ]
2. Belard A, Buchman T, Forsberg J, Potter BK, Dente CJ, Kirk A, et al. Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care. J Clin Monit Comput 2017 Apr;31(2):261-271. [doi: 10.1007/s10877-016-9849-1] [Medline: 26902081]
3. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma 2015;5:4 [FREE Full text] [doi: 10.1186/s13336-015-0019-3] [Medline: 25834725]
4. Damberg C, Timbie J, Bell D, Hiatt L, Smith A, Schneider E. RAND Corporation. 2012. Developing a framework for establishing clinical decision support meaningful use objectives for clinical specialties URL: https://www.rand.org/pubs/technical_reports/TR1129.html [accessed 2018-02-23] [WebCite Cache ID 6xRIkZoVv]
5. Eberhardt J, Bilchik A, Stojadinovic A. Clinical decision support systems: potential with pitfalls. J Surg Oncol 2012 Apr 01;105(5):502-510. [doi: 10.1002/jso.23053] [Medline: 22441903]
6. Jaspers MW, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011 May 01;18(3):327-334 [FREE Full text] [doi: 10.1136/amiajnl-2011-000094] [Medline: 21422100]
7. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 02;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
8. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. Arch Intern Med 2003 Jun 23;163(12):1409-1416. [doi: 10.1001/archinte.163.12.1409] [Medline: 12824090]
9. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int J Man Mach Stud 1993 Mar;38(3):475-487 [FREE Full text] [doi: 10.1006/imms.1993.1022]
10. Coiera E. Clinical decision support systems. In: Guide to Health Informatics. 2nd ed. Boca Raton, FL: CRC Press; 2003:331-344.
11. Wendt T, Knaup-Gregori P, Winter A. Decision support in medicine: a survey of problems of user acceptance. Stud Health Technol Inform 2000;77:852-856. [Medline: 11187675]
12. Venkatesh, Morris, Davis, Davis. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
13. Van Schaik P, Flynn D, Van Wersch A, Douglass A, Cann P. The acceptance of a computerised decision-support system in primary care: A preliminary investigation. Behav Info Technol 2004 Sep;23(5):321-326. [doi: 10.1080/01449290410006699 41]
14. Zheng K, Padman R, Johnson MP, Engberg J, Diamond HH. An adoption study of a clinical reminder system in ambulatory care using a developmental trajectory approach. Stud Health Technol Inform 2004;107(Pt 2):1115-1119. [Medline: 15360986]

15. Bergman LG, Fors UGH. Computer-aided DSM-IV-diagnostics - acceptance, use and perceived usefulness in relation to users' learning styles. BMC Med Inform Decis Mak 2005 Jan 07;5:1 [FREE Full text] [doi: 10.1186/1472-6947-5-1] [Medline: 15638940]
16. Curry L, Reed MH. Electronic decision support for diagnostic imaging in a primary care setting. J Am Med Inform Assoc 2011 May 01;18(3):267-270 [FREE Full text] [doi: 10.1136/amiajnl-2011-000049] [Medline: 21486884]
17. Zheng K, Padman R, Johnson MP, Diamond HS. Understanding technology adoption in clinical care: clinician adoption behavior of a point-of-care reminder system. Int J Med Inform 2005 Aug;74(7-8):535-543. [doi: 10.1016/j.ijmedinf.2005.03.007] [Medline: 16043083]
18. Gadd CS, Baskaran P, Lobach DF. Identification of design features to enhance utilization and acceptance of systems for Internet-based decision support at the point of care. Proc AMIA Symp 1998:91-95 [FREE Full text] [Medline: 9929188]
19. Johnson M, Zheng K, Padman R. Modeling the longitudinality of user acceptance of technology with an evidence-adaptive clinical decision support system. Decision Support Systems 2014 Jan;57:444-453 [FREE Full text] [doi: 10.1016/j.dss.2012.10.049]
20. Rosenbloom ST, Talbert D, Aronsky D. Clinicians' perceptions of clinical decision support integrated into computerized provider order entry. Int J Med Inform 2004 Jun 15;73(5):433-441. [doi: 10.1016/j.ijmedinf.2004.04.001] [Medline: 15171984]
21. Rousseau N, McColl E, Newton J, Grimshaw J, Eccles M. Practice based, longitudinal, qualitative interview study of computerised evidence based guidelines in primary care. BMJ 2003 Feb 08;326(7384):314 [FREE Full text] [Medline: 12574046]
22. Shibl R, Lawley M, Debuse J. Factors influencing decision support system acceptance. Decis Support Syst 2013 Jan;54(2):953-961. [doi: 10.1016/j.dss.2012.09.018]
23. Sousa VEC, Lopez KD, Febretti A, Stifter J, Yao Y, Johnson A, et al. Use of simulation to study nurses' acceptance and nonacceptance of clinical decision support suggestions. Comput Inform Nurs 2015 Oct;33(10):465-472 [FREE Full text] [doi: 10.1097/CIN.0000000000000185] [Medline: 26361268]
24. Terraz O, Wietlisbach V, Jeannot J, Burnand B, Froehlich F, Gonvers J, et al. The EPAGE internet guideline as a decision support tool for determining the appropriateness of colonoscopy. Digestion 2005;71(2):72-77 [FREE Full text] [doi: 10.1159/000084522] [Medline: 15775674]
25. Wallace CJ, Metcalf S, Zhang X, Kinder AT, Greenway L, Morris AH. Cost effective computerized decision support: tracking caregiver acceptance at the point of care. Proc Annu Symp Comput Appl Med Care 1995:810-813 [FREE Full text] [Medline: 8563404]
26. Buenestado D, Elorz J, Pérez-Yarza EG, Iruetaguena A, Segundo U, Barrena R, et al. Evaluating acceptance and user experience of a guideline-based clinical decision support system execution platform. J Med Syst 2013 Apr;37(2):9910. [doi: 10.1007/s10916-012-9910-7] [Medline: 23377779]
27. Heselmans A, Aertgeerts B, Donceel P, Geens S, Van de Velde S, Ramaekers D. Family physicians' perceptions and use of electronic clinical decision support during the first year of implementation. J Med Syst 2012 Dec;36(6):3677-3684. [doi: 10.1007/s10916-012-9841-3] [Medline: 22402980]
28. Peleg M, Shachak A, Wang D, Karnieli E. Using multi-perspective methodologies to study users' interactions with the prototype front end of a guideline-based decision support system for diabetic foot care. Int J Med Inform 2009 Jul;78(7):482-493. [doi: 10.1016/j.ijmedinf.2009.02.008] [Medline: 19328739]
29. Fullerton JT. A task analysis of American nurse-midwifery practice. J Nurse Midwifery 1987;32(5):291-296. [Medline: 3681470]
30. Oshio S, Johnson P, Fullerton J. The 1999-2000 task analysis of American nurse-midwifery/midwifery practice. J Midwifery Womens Health 2002;47(1):35-41. [Medline: 11874090]
31. Berner E, La Lande T. Overview of clinical decision support systems. In: Berner ES, editor. Clinical Decision Support Systems: Theory and Practice. New York: Springer Science+ Business Media; 2007:3-22.
32. Musen M, Shortliffe E, Cimino J. Clinical decision-support systems. In: Biomedical Informatics. New York: Springer Science; May 06, 2006:698-736.
33. Forster M. How do simple rules fit to reality in a complex world? Mind Mach 1999;9(4):543-564. [doi: 10.1023/A:1008304819398]
34. Simon H. Models of Bounded Rationality: Behavioral Economics and Business Organization. Volume 2. Cambridge, MA: The MIT Press; 1982.
35. Clark A. Being There: Putting Brain, Body, and World Together Again. Cambridge, MA: MIT press; 1998.
36. Holland J, Holyoak K, Nisbett R, Thagard P. A framework for induction. In: Induction: Processes of Inference, Learning, and Discovery. Cambridge, MA: MIT press; 1989:1-28.
37. Hayek F. The use of knowledge in society. Am Econ Rev 1945;35(4):519-530 [FREE Full text]
38. Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care--an interactive sociotechnical analysis. J Am Med Inform Assoc 2007;14(5):542-549 [FREE Full text] [doi: 10.1197/jamia.M2384] [Medline: 17600093]

39. Walter Z, Lopez MS. Physician acceptance of information technologies: role of perceived threat to professional autonomy. Decis Support Syst 2008 Dec;46(1):206-215. [doi: 10.1016/j.dss.2008.06.004]
40. Vashitz G, Meyer J, Parmet Y, Peleg R, Goldfarb D, Porath A, et al. Defining and measuring physicians' responses to clinical reminders. J Biomed Inform 2009 Apr;42(2):317-326 [FREE Full text] [doi: 10.1016/j.jbi.2008.10.001] [Medline: 19000935]
41. Stultz JS, Nahata MC. Computerized clinical decision support for medication prescribing and utilization in pediatrics. J Am Med Inform Assoc 2012;19(6):942-953 [FREE Full text] [doi: 10.1136/amiajnl-2011-000798] [Medline: 22813761]

Abbreviations

CDSS: clinical decision support systems
IPO: input-process-output
IPOE: input-process-output-engage
TAM: technology acceptance model
UASAD: user acceptance and system adaptation design
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 06.09.17; peer-reviewed by Y Gong, R Agrawal, L Zhou, T Saheb, KC Wong, R Robinson; comments to author 23.11.17; revised version received 02.03.18; accepted 19.03.18; published 18.04.18.

Please cite as:
Khairat S, Marc D, Crosby W, Al Sanousi A
Reasons For Physicians Not Adopting Clinical Decision Support Systems: Critical Analysis
JMIR Med Inform 2018;6(2):e24
URL: http://medinform.jmir.org/2018/2/e24/
doi:10.2196/medinform.8912
PMID:29669706

Original Paper

Nurses' Experience With Health Information Technology: Longitudinal Qualitative Study

Inga M Zadvinskis[1,2], PhD, RN; Jessica Garvey Smith[3], BS; Po-Yin Yen[3,4,5], PhD, RN

[1]Riverside Methodist Hospital, OhioHealth, Columbus, OH, United States

[2]Helene Fuld Health Trust National Institute for Evidence-Based Practice in Nursing and Healthcare, College of Nursing, The Ohio State University, Columbus, OH, United States

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

[4]Institute for Informatics, Department of Medicine, Washington University, St Louis, MO, United States

[5]Goldfarb School of Nursing, Barnes Jewish College, BJC Healthcare, St. Louis, MO, United States

Corresponding Author:
Po-Yin Yen, PhD, RN
Institute for Informatics
Department of Medicine
Washington University
4444 Forest Park Avenue, Suite 6318
St Louis, MO, 63108-2212
United States
Phone: 1 3142732313
Email: yenp@wustl.edu

Abstract

Background: Nurses are the largest group of health information technology (HIT) users. As such, nurses' adaptations are critical for HIT implementation success. However, longitudinal approaches to understanding nurses' perceptions of HIT remain underexplored. Previous studies of nurses' perceptions demonstrate that the progress and timing for acceptance of and adaptation to HIT varies.

Objective: This study aimed to explore nurses' experience regarding implementation of HIT over time.

Methods: A phenomenological approach was used for this longitudinal qualitative study to explore nurses' perceptions of HIT implementation over time, focusing on three time points (rounds) at 3, 9, and 18 months after implementation of electronic health records and bar code medication administration. The purposive sample was comprised of clinical nurses who worked on a medical-surgical unit in an academic center.

Results: Major findings were categorized into 7 main themes with 54 subthemes. Nurses reported personal-level and organizational-level factors that facilitated HIT adaptation. We also generated network graphs to illustrate the occurrence of themes. Thematic interconnectivity differed due to nurses' concerns and satisfaction at different time points. Equipment and workflow were the most frequent themes across all three rounds. Nurses were the most dissatisfied approximately 9 months after HIT implementation. Eighteen months after HIT implementation, nurses' perceptions appeared more balanced.

Conclusions: It is recommended that organizations invest in equipment (ie, wireless barcode scanners), refine policies to reflect nursing practice, and improve systems to focus on patient safety. Future research is necessary to confirm patterns of nurses' adaptation to HIT in other samples.

(JMIR Med Inform 2018;6(2):e38) doi:10.2196/medinform.8734

KEYWORDS

health IT; electronic health record; barcode medication administration; qualitative research; adaptation

Introduction

Background

Health information technology (HIT or Health IT) is a broad concept that includes a variety of technologies, including computer equipment, system software, and infrastructure that records, stores, protects, and retrieves clinical, administrative, or financial information [1]. In the United States, by 2015 96% of all hospitals had adopted a certified electronic health record (EHR) [2]. An EHR is a repository of patient data that is stored and exchanged securely among multiple authorized users to support quality integrated health care [3]. While providers and patients appreciate the positive benefits of HIT through improved care [4], there are also problems with HIT [5]. HIT problems are noteworthy because they may result in delayed care, altered clinical decision-making, and modified care processes, which affect patient outcomes [5]. A few unintended adverse consequences of EHRs are incomplete information, usability issues leading to frustrating user experiences, and patient privacy breaches [6].

Longitudinal Qualitative Studies About Clinicians' Perceptions of HIT

A myriad of individual and organizational factors influence HIT, contributing to its complexity and multi-dimensional nature [7]. A socio-technical model offers insights for studying HIT in complex adaptive health care systems through eight dimensions, including: (1) hardware and software computing infrastructure; (2) clinical content; (3) human-computer interface; (4) people; (5) workflow and communication; (6) internal organizational policies, procedures, and culture; (7) external rules, regulations, and pressures; and (8) system measurement and monitoring [8]. Researchers have studied provider perceptions of HIT over time, describing the mixed (both positive and negative) effects of HIT. For example, a two-year prospective, longitudinal survey of attending physicians in three clinical areas experienced the change from a homegrown EHR to a vendor EHR [9]. Safety perceptions dropped during the first six months, but then began to rise [9]. Physicians reported that the EHR created additional work and their satisfaction dropped [9]. In a similarly-designed study, ophthalmologists did not report a significant change in overall job satisfaction over time (3, 7, 13, and 24 months post-EHR implementation), but they expressed concern about the EHR's effect on interactions with patients and their ability to create quality documentation [10]. In primary care, even two years after EHR implementation, the EHR learning curve and computer knowledge remained challenging for interprofessional staff [11].

Longitudinal perceptions of EHRs have also been studied in nurses. Intensive care unit (ICU) nurses completed two cross-sectional survey questionnaires at 3 months and 12 months after EHR implementation and reported greater acceptance of the EHR at 12 months compared to 3 months [12]. In contrast, US nurses working on inpatient units within an academic medical center reported less positive attitudes toward the EHR 18 months after implementation compared to preimplementation and 6 months postimplementation [13]. For nurses, the timeline of adaptation for new HIT varies. A study conducted in Taiwan reported that nurses needed 3 months to understand the functionalities and benefits of an EHR [14]. In the United States, ICU nurses perceived the EHR as useful (through access to up-to-date information) at 3 months, but perceived that usefulness was not as relevant for acceptance at 12 months [12]. The authors explained that other EHR functionalities may have greater precedence over time to influence acceptance [12]. Positive computer attitudes are a significant predictor of fast adaptation [15]. Although studies exist regarding clinicians' adaptation to HIT, there is insufficient qualitative evidence regarding facilitators, hindrances, and a longitudinal timeline of adaptation, especially among clinical nurses.

Levels of Expectation Regarding HIT

In our prior study, we found that nurses' expectations regarding HIT can be stratified from personal-level (human-computer interaction) to organizational-level (quality of care) [16]. The five levels were: (1) how easy the system is to use (ie, equipment and system), (2) nurses' workflow and task performance, (3) collaboration within the nursing unit, (4) nurses' communication across hospital disciplines and departments, and (5) the effects of HIT on quality of care (ie, patient safety and nurse/patient satisfaction). Each level expands from individual user concerns to the team and the broader organization, and each level increases in complexity. In early implementation, we found that HIT users may be more concerned with lower-level expectations, but the timeframe for nurses' expectations, and thus adaptation, is unknown.

Objective

We investigated HIT adaptation, which is, "a process of modifying existing conditions in an effort to achieve alignment" [17] involving workflow redesign, user training, and technology maintenance [18]. This was a longitudinal qualitative study that explored medical-surgical nurse perceptions of HIT implementation over time. We interviewed nurses three times after EHR and bar code medication administration (BCMA) implementation to capture evolving adaptation of their perceptions and behaviors in this specific job role. The objective of the study was to explore nurses' experience of HIT implementation, and how they adapted their perceptions and behavior to HIT upgrades and optimization over time.

Methods

Overview

We conducted a phenomenological qualitative study at a large Midwestern academic medical center. The medical center implemented a customized commercialized EHR system (EPIC platform [19]) in October 2011, which included computerized provider order entry, electronic charting, and BCMA. Prior to 2011, a few nursing units (ie, critical care) used some electronic documentation. In 2011, all nursing units within the medical center began using the new EHR system. In this study, participants were from a medical-surgical unit that used paper charts prior to implementation.

Sample

We used purposive sampling to recruit staff nurses who worked on a medical-surgical unit that used EHRs and BCMA and had a minimum of two years of working experience in the organization. We approached participants either face-to-face or via email and hosted private face-to-face interviews in a location away from the clinical area to ensure privacy and avoid disruption.

Data Collection

We used a semi-structured interview guide with additional probes (Multimedia Appendix 1) to clarify and discover in-depth information. Field notes were taken during and after the interviews. We interviewed participants over three time points based on convenience, including Round 1 (R1) 3 months post-EHR implementation (winter 2012), Round 2 (R2) 9 months post-EHR implementation (summer 2012), and Round 3 (R3) 18 months post-EHR implementation (summer 2013). The length of each interview ranged from 20 to 60 minutes. All interviews were audio-recorded and professionally transcribed verbatim. We also collected basic demographic information, including age, gender, position, education, and years of experience.

Data Analysis

Three researchers (IZ, JGS, and PY) read the transcripts independently and located relevant statements in the transcripts that expressed units of meaning. The researchers generated common themes by synthesizing the meaning units. Themes reflected a general description of the nurse participants' experience with EHRs and BCMA. The structure for the first five themes was derived from previous work that delineated confirmed nurses' expectations ranging from personal-level to organizational-level [16]. Dimensions of the socio-technical model, such as hardware, people, workflow communication, and policies, informed coding structure during qualitative analysis. After iterative discussion and refinement, a codebook was generated. We developed a total of 61 themes: 7 at the first level and 54 at the second level. We also kept a detailed record of our codes and their definitions as we updated them over time (Multimedia Appendix 2). Each quote was classified from one to four themes, as one quote could contain multiple units of meaning (themes), which affected thematic frequency; we defined this as theme cooccurrence. We used NVivo 10 [20], a qualitative research analysis tool, for data management and analysis. In addition, we used Gephi, a graph visualization and exploration software [21], to illustrate the cooccurrence relationships among themes, where nodes represented themes and lines signified the cooccurring relationships. The network graph was illustrated in a forced layout.

Intercoder Reliability

We assessed and established our intercoder reliability using Cohen's Kappa. Cohen's Kappa has been commonly used to assess intercoder reliability, with recommendations greater than 0.7 for semi-structured interviews, especially with multiple complex codes [22-24]. We coded 10% of the transcripts together to establish consensus, as experts recommend [22,24]. For each independently-coded transcript, we calculated Cohen's Kappa by averaging the Kappas of all themes through a multiple independent coding comparison process [22,24,25]. In our analysis, two researchers (IZ and JGS) independently coded randomly selected transcripts and discussed discrepancies; a third researcher (PY) mediated the discussion to confirm final coding. After iterative discussion, we reached a Cohen's Kappa of 0.82.

Results

Principal Findings

Nineteen nurses participated in the study, with some nurses participating in multiple rounds. We conducted a total of 30 interviews: 9 from R1, 11 from R2, and 10 from R3. We were unable to interview all nurses across all 3 rounds because some nurses were unavailable (ie, schedule conflicts, transferred to new positions, declined participation). Among the 9 nurses that participated in R1, 7 participated in R2, and 3 participated in R3. Eleven nurses participated in only one interview, and 8 nurses participated in more than one interview. Among the 19 nurses, 17 were female. The age of the nurses ranged from 22 to 52 years old, with 3 to 25 years of working experience, and 58% (11/19) worked the day shift. Most nurses (15/19, 79%) were Bachelor of Science in Nursing-prepared and all worked as staff nurses. All nurses owned home computers and 79% (15/19) owned smartphones. Across all rounds, nurses rated themselves an average of 4 out of 5 in computer competency, with 1 meaning *not competent* and 5 meaning *very competent*. We used information saturation to determine the number of participants.

We assembled themes into a table format to review nurse perceptions over time (Multimedia Appendix 2), categorizing them into 7 main themes with 54 subthemes. Details regarding development of the first five primary themes, E1 to E5, are described elsewhere [16]. The longitudinal approach led to the discovery of two additional themes: *adaptation* and *organizational factors*. *Adaptation* explained both internal and external resources that influenced acclimatization to technology over time. *Organizational factors* discussed communication or HIT decisions made by organizational leaders that influenced nursing work. Quote examples and frequencies can be found in Multimedia Appendix 3.

Thematic Findings

E1: Nurses' Interaction with HIT

The E1 subthemes involved *system* (software) and *equipment* components. Equipment addressed workstations on wheels (WOWs), scanners, and wires (wired mouse/scanners and electrical cords). Nurses perceived equipment negatively because of noise, occasional (battery) power loss, and challenging use in semi-private patient rooms and narrow, crowded hallways. By R3, the information technology (IT) department installed well-liked wireless BCMA scanners. System functionality concerns included lengthy login, program shutdowns during medication administration, and BCMA scanning issues. Shutdowns occurred if users forgot to plug in the WOW when not in use, leading to power loss, or sometimes for unknown reasons. Nurses reported difficult system navigation in all

rounds, especially during emergency documentation. In later rounds, nurses acknowledged easier system navigation because EHR flowsheets were updated with head-to-toe organization that reflected how nurses conducted physical assessments.

E2: Nursing Performance Regarding Task Accomplishment

Nurses viewed documentation as time-consuming and arduous, yet thorough. Nurses valued feedback on performance from visual indicators (green and red dots) for complete or incomplete documentation. In R1 and R2, nurses thought documentation was inefficient and not streamlined because it contained elements irrelevant to their population, required too much scrolling, appeared chaotic, and took longer than paper charting. However, by R3 some nurses expressed that documentation was streamlined and efficient since the IT department had updated the EHRs to be more compact. Nonetheless, documenting rare events caused nurses stress and confusion. For instance, blood administration involved scanning multiple barcodes in a strange pattern and emergency documentation had complicated screen layouts.

E3: Unit-Specific Teamwork

Since all nurses experienced glitches with EHRs, they relied on nursing unit collaboration for assistance. Collaboration impacted patient safety because nurses appreciated the ability to view all records when administering medicine to other nurses' patients. Teamwork was also associated with improved nurses' satisfaction. Due to strong teamwork, nurses relied on one another more than IT staff for resolving system concerns.

E4: Interdisciplinary Teamwork

E4 themes concerned communication between departments and disciplines. Interprofessional notes helped nurses understand the care plan promoting integrated and better care. Nurses reviewed patient transfer information before arrival. Unexpectedly, some departments stopped telephone handoffs, leading to potential missed information (ie, last pain medication). Nurses identified unequal standards between departments, such as not scanning all medications or omitting parts of admission documentation. Documentation standards varied and some prescribers did not enter orders immediately after patient assessment. Unequal standards existed in the context of shifting responsibilities, such as doctors asking nurses to input their orders. This factor frustrated nurses, and they proposed standardized documentation classes and policies to improve interprofessional communication.

E5: Quality of Care

E5 themes related to quality of care and nurse satisfaction. Although BCMA reduced some medication errors, the potential for error remained because BCMA did not verify multi-dose medication containers (like insulin) and prescribers could still input orders incorrectly. Additional potential for errors during physical assessment documentation could occur due to the repetitive nature of electronic documentation (mouse clicking), leading to distraction and loss of attention. Patient-nurse interactions were also altered because nurses had to position their backs to patients to document on WOWs. In addition, occasionally WOWs logged nurses off or malfunctioned, leading to pain medication delays that negatively impacted care. During R1 and R2, nurses were so frustrated with learning the system that they did not appreciate potential improvements in patient care quality. However, by R3, after nurses adapted to EHRs, they frequently mentioned better quality of patient care through access to patient history, notes from all disciplines, task/documentation reminders, and improved patient safety. Perceptions of care quality improved in R3 compared to R1 and R2.

When the EHR system was first implemented, some nurses felt scared or intimidated, although eventually it met expectations or appeared that it would in the future. The most common expectation was reduced documentation that would allow for more time with patients, which did not happen, leading to disappointment. Nurses' dissatisfaction and satisfaction were mentioned with similar frequencies in R1 and R3, although in R2 nurses were more dissatisfied than satisfied. The greatest sources of nurse dissatisfaction were equipment, system functionalities, inefficient documentation, and lengthy logins. Conversely, nurses were satisfied with BCMA error reduction, workflow simplification, patient protection, better care, and documentation thoroughness and reminders. By R3, after nurses adapted, some nurses expressed that the EHR system offered more time for higher-quality patient care.

Adaptation

Adaptation affected nurses' acclimatization to new technology over time. Nurses discussed self-learning through personal motivation, practice, and long-term use. Self-learning occurred through the EHR playground, where nurses could explore the EHR layout. However, training was a major concern with rushed, fast, and overwhelming classes that were provided too far in advance of implementation. Nurses commented that training classes did not reflect nurses' workflow; they only showed system design and navigation.

Organizational Factors

Organizational factors, such as policies, requirements, and decisions made by leadership, were frequently mentioned with nurses' dissatisfaction. Clinicians expected leaders to explain rationale for HIT decisions that would impact clinical practice. For example, nurses were not aware of the rationale for policies not allowing a copy/paste function, or real-time documentation which expected nurses to chart assessments immediately after care. Nurses expected hospital leadership to be more aware of bedside nursing workflow and resolve issues quickly. Leadership added requirements or responsibilities but did not retire old/unnecessary requirements, which added to nursing work and complicated workflow. Nurses expected leadership to advocate for system features to improve nursing workflow, so that they could spend more time at the bedside.

Nurses' suggestions for improvement grew with each round and were often related to the system and equipment. Some suggestions were incorporated by R3, such as an exact time stamp and wireless scanners. Nurses were thankful that nursing management filtered information regarding important EHR updates that affected nursing work. Nurses valued leadership's feedback on their performance regarding percent of scanned

medications, and advised leadership to be patient, remain supportive, provide resources, and answer questions.

Visualization of Theme and Subtheme Interrelationships

We used Gephi to create a network (relationship) graph of themes for each round. Each quote contained one to four themes, while the edges/lines revealed theme cooccurrence within the same quote (Figure 1). Larger size nodes (bubbles) illustrate more frequent themes; thicker lines illustrate stronger relationships/cooccurrence. The numbers within nodes are a labeling mechanism, and do not represent the frequency of the quotes. All graphs used the same scale to facilitate thematic comparisons between rounds. Themes are organized by color, with E1 as green, E2 as blue, E3 as red, E4 as yellow, E5 as purple, adaptation as orange, and organizational factors as black.

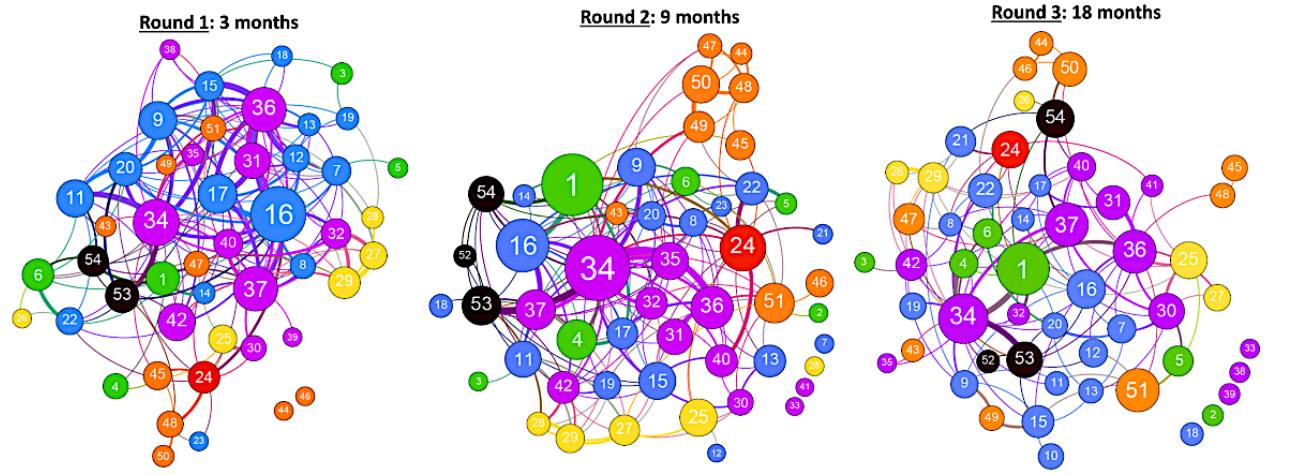
Nurses' dissatisfaction was related to equipment and documentation. *(#1): Equipment* was a great concern in R2 and R3 and was strongly associated with *(#34): Nurses' dissatisfaction* in R3, as evidenced by the thick interrelationship line. *(#9): Documentation inefficient* and *(#11): Documentation not streamlined* decreased with each round, especially in R3, and they were associated with *(#34): Nurses' dissatisfaction*. *(#35): Nurses' expectations* of EHR strongly reflected their *(#34): Dissatisfaction* in R2.

There were trends in collaboration, communication, quality of care, and workflow. *(#24): Nursing collaboration* was greatest in R2. *(#25): Communication across disciplines* was increasingly connected to *(#30): Better care* and *(#36): Nurses' satisfaction*. *(#36): Nurses' satisfaction* was consistently concurrent with *(#31): Error reduction*. As nurses began to adapt to HIT, the *(#16): Impact on workflow* decreased by R3.

The longitudinal approach led to discovery of two additional themes: *adaptation* and *organizational factors*. *(#51): Users adapt* became more prominent in later rounds compared to earlier rounds. *(#53): Policies* were strongly connected to *(#34): Nurses' dissatisfaction* in all rounds.

Due to the research approach, nearly all quotes could be related to *(#34): Nurses' dissatisfaction* or, conversely, *(#36): Nurses' satisfaction*. Aside from these themes, *(#1): Equipment* and *(#16): Impact on workflow* were most frequent across all three rounds. In R1, the most common themes were *(#16): Impact on workflow* and *(#37): Patient experience/satisfaction*. In R2, the most common themes were *(#1): Equipment*, *(#16): Impact on workflow*, and *#(24): Nursing unit collaboration*. In R3, *(#1): Equipment* remained as the top theme, followed by *(#37): Patient experience/satisfaction*. Themes grew and changed at each time point and were interconnected differently based on the nurses' main concerns and satisfactions at different time points. Visually, the R3 graph was the most balanced as nurses' perceptions seemed to be equally scattered across all themes compared to R1 and R2.

Figure 1. Visualization of theme and subtheme interrelationships.



E1 Nurses' Interaction with HIT	E2 Nursing Performance regarding Task Accomplishment			E3 Unit-specific Teamwork	E4 Interdisciplinary Teamwork	E5 Quality of Care		Adaptation		Organizational Factors
1: Equipment	7: Documentation Efficient	13:Documentation Thoroughness	19: Med Admin Overriding and Linking	24: Nursing Unit Collaboration	25: Communication Across Disciplines	30: Better Care	37: Patient Experience and Satisfaction	43: Clinicians Involvement with Design	49: Technology Proficiency	52: Leadership Suggestions
2: System Changing Layouts	8: Documentation General	14: Environment	20: Physical Assessment		26: Missed Communication	31: Error Reduction	38: Patient Experience Modern Healthcare	44: EHR Playground	50: Training	53: Policies
3: System Errors due to Disorganized Layout	9: Documentation Inefficient	15: Feedback on Performance	21: Rare Events Blood Admin		27: Patient Transfers	32: Negative Impact on Care	39: Patient Experience Patients Adjust	45: IT Support Team	51: Users Adapt	54: Suggestions for Improvement
4: System Functionality	10: Documentation New Features	16: Impact on Workflow	22: Rare Events Emergency		28: Shifting Responsibility	33: No Change in Care	40: Patient Safety	46: Self-learning		
5: System Easy Navigation	11: Documentation Not Streamlined	17: Med Admin General	23: Rare Events General		29: Unequal Standards	34: Nurses' Dissatisfaction	41: Patient Safety Catching Deterioration	47: Staffing		
6: System Difficult Navigation	12: Documentation Streamlined	18: Med Admin Timing				35: Nurses' Expectations	42: Potential for Error	48: Super Users		
						36: Nurses' Satisfaction				

*The numbers associated with the themes do not represent their frequency, the numbers are only a coding mechanism to find them easily on the graph. The size of the bubbles represents the frequency of the quotes.
*Themes that were not mentioned by nurses in each round do not appear on the graphs. In R1, 6 themes were not mentioned including 10:Documentation New Features, 52:Leadership Suggestions, 33:No Change in Care, 41:Patient Safety Catching Deterioration, 21:Rare Events Blood Admin, and 2:System Changing Layouts. In R2, 3 themes were not mentioned including 10:Documentation New Features, 38:Patient Experience Modern Healthcare, and 39:Patient Experience Patients Adjust. In R3 one theme was not mentioned, 23:Rare Events General.

Discussion

Overview

This study explored the trajectory of change through a qualitative analysis of nurses' experiences after EHR and BCMA implementation at three time points: 3, 9, and 18 months. We used the socio-technical model [8] to guide our discussion. Two dimensions of the socio-technical model are not addressed ("external rules, regulations, and pressures" and "system measurement and monitoring") because nurses did not express opinions related to these dimensions.

Hardware and Software Computing Infrastructure

Nurses shared opinions regarding HIT equipment. Workstation preferences varied: some nurses liked working with portable workstations so they could store medications in locked drawers, while other nurses desired fixed workstations in patient rooms or portable tablets. Mobile devices may promote nurses' ability to document at the bedside and point-of-care [26]. Nurses appreciated equipment that improved functionality, such as fingerprint login scanners and wireless input devices (scanners).

Clinical Content

Nurses offered multiple system suggestions: they wanted illuminated new and abnormal laboratory results, parameters for holding medication (eg, blood pressure), and unacknowledged orders to be highlighted in red. Nurses suggested system recognition and display of insidious abnormal trends, such as increasing white blood cell counts or decreasing hemoglobin. Physical assessment customization, such as removal of irrelevant fields or adding fields/drop-down boxes, would improve nurses' documentation. EHR navigation could be improved with a help sheet or search tab with a glossary or key words. A patient calendar to track scheduled tests would help prepare patients and improve workflow, communication, and possibly patient satisfaction.

Human-Computer Interface

Clinician involvement in HIT design is a potential strategy for successful adaptation [27]. Nurses expressed frustration that clinician input was not adequately integrated into the system. Communication regarding end-user requests facilitates success [28]. Previous research indicates that when nurses were involved in refining the usability of an existing EHR, end-user satisfaction increased and nurse-sensitive quality indicators improved, including fewer catheter-associated urinary tract infections, improved documentation of the presence of pressure ulcers, and fewer restraints [29]. Nurses offered suggestions for workflow redesign and EHR modifications via email request, discussion with the Nurse Manager, and during IT meetings.

People (Training and Peer Support)

Training and competency are sociotechnical factors that affect HIT adaptation [30]. Nurses may improve their competency and technology proficiency through practice. In this study, EHR training began months prior to implementation with nurses attending four one-hour educational training sessions. During training and early implementation, nurses were encouraged to practice documentation in the EHR playground. Although this sample of nurses rated their proficiency with technology as competent (4 out of 5), nurses working in other settings may feel less proficient or less IT-literate. Low IT literacy is a known barrier to EHR implementation [31,32], and it may be helpful to improve clinicians' computer literacy prior to implementation. Training programs may be customized to different nurses' learning needs to promote adaptation. Nurses viewed super users as instrumental for adapting to HIT. This finding is similar to another study in which clinicians viewed super users as supportive, familiar, and knowledgeable regarding day-to-day work [33], but different from another study that showed having super users did not contribute to meaningful HIT use [34]. Effective super users may be characterized by being proactive, providing comprehensive explanations, using positive framing, and freely sharing information [33].

Workflow and Communication

In the socio-technical model for studying HIT, the clinical workflow involved with operating HIT systems must be consistent with internal policies and procedures [8]. Lack of congruency between policy and practice was a source of nursing frustration in this study. For example, prior to EHR implementation, nurses working 12 hours would reassess patients and write, "unchanged from previous assessment" in the paper chart, which was supported by policy. After EHR implementation, nurses had to redocument a complete physical assessment every 8 hours, because IT developers set the system for eight-hour tours of duty for nurses, which was inconsistent with practice. Therefore, numerous nurses expressed the desire for a copy and paste function, but this can inadvertently lead to clinical harm [35]. Organizations interested in using a copy and paste function may benefit from adopting best practices from the Partnership for Health IT Patient Safety [35]. Eventually, the physical documentation assessment policy in the EHR was modified to reflect nursing workflow.

Internal Organizational Policies, Procedures, and Culture

Nurses voiced that a reduced patient load (better nurse staffing) was very helpful for adapting to the new system. During the first day of EHR implementation, medical-surgical nurses cared for only one or two patients. By R3, nurses returned to caring for four to five patients, but some nurses continued to struggle with completing documentation requirements. In the future, augmented clinical HIT (where nurse staffing decisions are based on patient volume, acuity levels, etc [36]) may become more commonplace.

A 2013 integrative review found that strong leadership ensures that the team works toward successful HIT implementation [37]. Modification of hospital policies and environment may be necessary, especially when HIT changes are directed by a vendor [38]. Leaders are expected to communicate a clear vision and expectations related to HIT for clinicians [39]. Users' expectations are a considerable psychological factor that affects EHR adoption [40]. Leadership may need to develop expectations for staff accountabilities [39] to maintain HIT in good working order (eg, plug in workstations when not in use, report broken equipment).

Timing of Adaptation to HIT

Adaptation to HIT over time may be explained in part by the Gartner Hype Cycle [41], which describes maturity and adoption of new technologies and applications. The Hype Cycle includes five phases: (1) innovation trigger, (2) peak of inflated expectations, (3) trough of disillusionment, (4) slope of enlightenment, and (5) plateau of productivity. The peak of nurses' dissatisfaction occurred in R2, approximately nine months after HIT implementation. This finding may be due to nurses' high expectations not being met at that time, corresponding with the Hype Cycle "trough of disillusionment." Nurses began to grow impatient regarding supposed EHR benefits. Previous research indicates that nurses' expectations for HIT include availability, speed, decreased work load, and ease of use [42]. In this study, nurses were able to provide suggestions and feedback to align HIT with their expectations regarding nursing workflow, training, and technology. However, the effectiveness of such communication is unclear. Research also indicates that nurses want to improve HIT through suggestions, but low communication levels and lack of feedback are barriers to enhancing system performance [43]. Variables related to nurses' acceptance of EHRs include training/education, facilitating conditions, social influences, observability, and job relevance [44].

Comparison of Nurses' Experience With HIT With Other Clinicians

Previous research indicates that clinician satisfaction with HIT is mixed, which may be related to clinical documentation practice, workload, and productivity [45]. Physicians have reported dissatisfaction with template-based HIT documentation [45], possibly due to their high autonomy needs to prioritize work tasks [46]. In the current study, nurses also expressed dissatisfaction with template-based physical assessment because documentation was not streamlined. Nurses may be more accustomed to template-based documentation than physicians due to the use of standardized nursing care plans and clinical care pathways in the profession. However, if the HIT template does not match nurses' mental model for accomplishing tasks, they will be dissatisfied.

Physicians and nurses may differ in perceptions of productivity after HIT implementation. While physicians may experience better productivity due to increased charges, improved work relative value units, and less time writing orders [45], nurses may not experience HIT value because there is no direct link in billing systems between individual nurses and patients [47]. Nurses' productivity (hours of nursing care per patient day) is built into hospital room and board charges, so HIT may not be able to capture the work value of individual users like nurses [47]. Conversely, nurses' experience with HIT may be similar to other clinicians' experience. A qualitative study that aimed to understand physicians' and nurses' experience with EHRs found no major perceived differences based on profession [48]. A commonality among all professions appears to be the need for communication when implementing new HIT [46].

Study Limitations

Generalizability of these findings is limited due to sampling nurses from one unit within an academic medical center. Self-selection bias may have occurred from voluntary participation. The timing of interview rounds was based upon interviewer availability rather than change theory. Although we conducted 30 interviews, some nurses participated over multiple time periods and the group comprised a small sample size (n=19). Despite the small sample size, recurrence of similar themes across multiple individuals established information saturation and data quality.

Conclusion

A longitudinal qualitative approach for studying HIT adaptation facilitated understanding of thematic relationships over time. Although thematic interconnectivity differed due to nurses' concerns and satisfaction at different time points, some trends were noted. Nurses appeared the most dissatisfied in R2, but many sources of dissatisfaction may be rectified, such as new equipment, refined policies, and improved systems to focus on patient safety. Approximately 18 months after HIT implementation, nurses' perceptions appeared more balanced, as indicated by more consistent thematic frequencies and weaker cooccurrences in the Gephi chart. Balanced thematic distribution and interconnectivity within Gephi charts may be a visual indicator of HIT adaptation progress. Future research is necessary to confirm if researchers can replicate these findings in other samples.

Acknowledgments

The authors would like to thank the computer science student Xiaonan Ji for providing the Gephi graph. We thank Esther Chipps, Laura Droll, and Kelly Torrence at The Ohio State University Wexner Medical Center for coordinating the recruitment, and all participating nurses for their time and effort.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Semistructured interview guide.

[PDF File (Adobe PDF File), 328KB - medinform_v6i2e38_app1.pdf]

Multimedia Appendix 2

Theme codebook and definitions.

[PDF File (Adobe PDF File), 38KB - medinform_v6i2e38_app2.pdf]

Multimedia Appendix 3

Themes with quote examples.

[PDF File (Adobe PDF File), 394KB - medinform_v6i2e38_app3.pdf]

References

1. Office of the National Coordinator for Health Information Technology. Washington, DC: US Department of Health and Human Services; 2017 Sep 15. Benefits of health IT URL: https://www.healthit.gov/topic/health-it-basics/glossary [accessed 2018-05-23] [WebCite Cache ID 6zdC06Wt2]
2. Office of the National Coordinator for Health Information Technology. Washington, DC: US Department of Health and Human Services; 2016. Non-federal acute care hospital health IT adoption and use: health IT dashboard URL: https://dashboard.healthit.gov/apps/hospital-health-it-adoption.php [accessed 2018-05-23] [WebCite Cache ID 6zdC5HQaV]
3. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]
4. Buntin M, Burke M, Hoaglin M, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
5. Kim M, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. J Am Med Inform Assoc 2017 Dec 01;24(2):246-250. [doi: 10.1093/jamia/ocw154] [Medline: 28011595]
6. Sittig DF, Wright A, Ash J, Singh H. New unintended adverse consequences of electronic health records. Yearb Med Inform 2016 Nov 10(1):7-12 [FREE Full text] [doi: 10.15265/IY-2016-023] [Medline: 27830226]
7. Kukafka R, Johnson SB, Linfante A, Allegrante JP. Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform 2003 Jun;36(3):218-227 [FREE Full text] [Medline: 14615230]
8. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
9. Hanauer D, Branford G, Greenberg G, Kileny S, Couper M, Zheng K, et al. Two-year longitudinal assessment of physicians' perceptions after replacement of a longstanding homegrown electronic health record: does a J-curve of satisfaction really exist? J Am Med Inform Assoc 2017 Apr 01;24(e1):e157-e165. [doi: 10.1093/jamia/ocw077] [Medline: 27375291]
10. Ehrlich J, Michelotti M, Blachley T, Zheng K, Couper M, Greenberg G, et al. A two-year longitudinal assessment of ophthalmologists' perceptions after implementing an electronic health record system. Appl Clin Inform 2016 Dec 12;7(4):930-945 [FREE Full text] [doi: 10.4338/ACI-2016-05-RA-0075] [Medline: 27730248]
11. Terry A, Brown J, Bestard DL, Thind A, Stewart M. Perspectives on electronic medical record implementation after two years of use in primary health care practice. J Am Board Fam Med 2012;25(4):522-527 [FREE Full text] [doi: 10.3122/jabfm.2012.04.110089] [Medline: 22773720]
12. Carayon P, Cartmill R, Blosky M, Brown R, Hackenberg M, Hoonakker P, et al. ICU nurses' acceptance of electronic health records. J Am Med Inform Assoc 2011;18(6):812-819 [FREE Full text] [doi: 10.1136/amiajnl-2010-000018] [Medline: 21697291]
13. Laramee A, Bosek M, Shaner-McRae H, Powers-Phaneuf T. A comparison of nurse attitudes before implementation and 6 and 18 months after implementation of an electronic health record. Comput Inform Nurs 2012 Oct;30(10):521-530. [doi: 10.1097/NXN.0b013e3182573b04] [Medline: 22584880]
14. Chang C, Lee T, Liu C, Mills M. Nurses' experiences of an initial and reimplemented electronic health record use. Comput Inform Nurs 2016 Apr;34(4):183-190. [doi: 10.1097/CIN.0000000000000222] [Medline: 26886680]
15. Colligan L, Potts H, Finn C, Sinkin R. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. Int J Med Inform 2015 Jul;84(7):469-476. [doi: 10.1016/j.ijmedinf.2015.03.003] [Medline: 25868807]
16. Zadvinskis I, Chipps E, Yen P. Exploring nurses' confirmed expectations regarding health IT: a phenomenological study. Int J Med Inform 2014 Feb;83(2):89-98. [doi: 10.1016/j.ijmedinf.2013.11.001] [Medline: 24321515]
17. Majchrzak A, Rice R, Malhotra A, King N. Technology adaptation: the case of a computer-supported inter-organizational virtual team. MIS Quarterly 2000;24(4):569-600. [doi: 10.2307/3250948]

18. Cooper R, Zmud R. Information technology implementation research: a technological diffusion approach. Manag Sci 1990 Feb;36(2):123-139. [doi: 10.1287/mnsc.36.2.123]
19. EPIC Systems Corporation. 2016. Epic with the patient at the heart: basics homepage URL: https://www.epic.com/ [accessed 2018-06-03] [WebCite Cache ID 6zttIF93o]
20. QSR International. 2018. What is NVivo? URL: http://www.qsrinternational.com/nvivo/what-is-nvivo [accessed 2018-06-03] [WebCite Cache ID 6zttNSxif]
21. The Gephi Consortium.: Gephi; 2017. The open graph viz platform URL: https://gephi.org/ [accessed 2018-05-23] [WebCite Cache ID 6zdCpxH5U]
22. Campbell J, Quincy C, Osserman J, Pedersen O. Coding in-depth semistructured interviews. Sociol Methods Res 2013 Aug 21;42(3):294-320. [doi: 10.1177/0049124113500475]
23. Hruschka D, Schwartz D, St. John DC, Picone-Decaro E, Jenkins RA, Carey JW. Reliability in coding open-ended data: lessons learned from HIV behavioral research. Field Methods 2016 Jul 24;16(3):307-331. [doi: 10.1177/1525822X04266540]
24. MacPhail C, Khoza N, Abler L, Ranganathan M. Process guidelines for establishing intercoder reliability in qualitative studies. Qual Res 2015 Apr 20;16(2):198-212. [doi: 10.1177/1468794115577012]
25. Burla L, Knierim B, Barth J, Liewald K, Duetz M, Abel T. From text to codings: intercoder reliability assessment in qualitative content analysis. Nurs Res 2008;57(2):113-117. [doi: 10.1097/01.NNR.0000313482.33917.7d] [Medline: 18347483]
26. Kohle-Ersher A, Chatterjee P, Osmanbeyoglu HU, Hochheiser H, Bartos C. Evaluating the barriers to point-of-care documentation for nursing staff. Comput Inform Nurs 2012 Mar;30(3):126-133. [doi: 10.1097/NCN.0b013e3182343f14] [Medline: 22024972]
27. Lavin MA, Harper E, Barr N. Health information technology, patient safety, and professional nursing care documentation in acute care settings. Online J Issues Nurs 2015 Apr 14;20(2):6 [FREE Full text] [Medline: 26882425]
28. Collins SA, Alexander D, Moss J. Nursing domain of CI governance: recommendations for health IT adoption and optimization. J Am Med Inform Assoc 2015 May;22(3):697-706. [doi: 10.1093/jamia/ocu001] [Medline: 25670752]
29. Page CAK, Schadler A. A nursing focus on EMR usability enhancing documentation of patient outcomes. Nurs Clin North Am 2014 Mar;49(1):81-90. [doi: 10.1016/j.cnur.2013.11.010] [Medline: 24485189]
30. Yen P, McAlearney AS, Sieck CJ, Hefner JL, Huerta TR. Health information technology (HIT) adaptation: refocusing on the journey to successful HIT implementation. JMIR Med Inform 2017 Sep 07;5(3):e28 [FREE Full text] [doi: 10.2196/medinform.7476] [Medline: 28882812]
31. Takian A, Sheikh A, Barber N. We are bitter, but we are better off: case study of the implementation of an electronic health record system into a mental health hospital in England. BMC Health Serv Res 2012 Dec 31;12:484 [FREE Full text] [doi: 10.1186/1472-6963-12-484] [Medline: 23272770]
32. Verma M, Gupta S. Problems faced by nurses in use of electronic health records during clinical practice. Stud Health Technol Inform 2016;225:985-986. [Medline: 27332444]
33. Yuan CT, Bradley EH, Nembhard IM. A mixed methods study of how clinician 'super users' influence others during the implementation of electronic health records. BMC Med Inform Decis Mak 2015 Apr 10;15:26 [FREE Full text] [doi: 10.1186/s12911-015-0154-6] [Medline: 25889076]
34. Shea C, Reiter K, Weaver M, Albritton J. Quality improvement teams, super-users, and nurse champions: a recipe for meaningful use? J Am Med Inform Assoc 2016 Nov;23(6):1195-1198 [FREE Full text] [doi: 10.1093/jamia/ocw029] [Medline: 27107442]
35. Tsou AY, Lehmann CU, Michel J, Solomon R, Possanza L, Gandhi T. Safe practices for copy and paste in the EHR. Systematic review, recommendations, and novel model for health IT collaboration. Appl Clin Inform 2017 Dec 11;8(1):12-34 [FREE Full text] [doi: 10.4338/ACI-2016-09-R-0150] [Medline: 28074211]
36. Sharma L, Chandrasekaran A, Boyer K, McDermott C. The impact of Health Information Technology bundles on Hospital performance: An econometric study. Journal of Operations Management 2016 Jan;41:25-41. [doi: 10.1016/j.jom.2015.10.001]
37. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
38. Topaz M, Ronquillo C, Peltonen LM, Pruinelli L, Sarmiento RF, Badger MK, et al. Nurse informaticians report low satisfaction and multi-level concerns with electronic health records: results from an international survey. AMIA Annu Symp Proc 2016;2016:2016-2025 [FREE Full text] [Medline: 28269961]
39. Ingebrigtsen T, Georgiou A, Clay-Williams R, Magrabi F, Hordern A, Prgomet M, et al. The impact of clinical leadership on health information technology adoption: systematic review. Int J Med Inform 2014 Jun;83(6):393-405. [doi: 10.1016/j.ijmedinf.2014.02.005] [Medline: 24656180]
40. Najaftorkaman M, Ghapanchi AH, Talaei-Khoei A, Ray P. A taxonomy of antecedents to user adoption of health information systems: a synthesis of thirty years of research. J Assn Inf Sci Tec 2014 May 09;66(3):576-598 [FREE Full text] [doi: 10.1002/asi.23181]
41. Gartner Inc. Stamford, CT: Gartner; 2008. Understanding Gartner's hype cycles URL: https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp [accessed 2018-06-03] [WebCite Cache ID 6ztrjfxek]

42. Gürsel G. User expectations: nurses' perspective. Stud Health Technol Inform 2016;225:897-898. [Medline: 27332398] [doi: 10.3233/978-1-61499-658-3-897]
43. Khalifa M. Evaluating nurses acceptance of hospital information systems: a case study of a tertiary care hospital. Stud Health Technol Inform 2016;225:78-82. [Medline: 27332166] [doi: 10.3233/978-1-61499-658-3-78]
44. Strudwick G, McGillis Hall L. Nurse acceptance of electronic health record technology: a literature review. J Res Nurs 2015 Nov 25;20(7):596-607. [doi: 10.1177/1744987115615658]
45. Nguyen L, Bellucci E, Nguyen L. Electronic health records implementation: an evaluation of information system impact and contingency factors. Int J Med Inform 2014 Nov;83(11):779-796. [doi: 10.1016/j.ijmedinf.2014.06.011] [Medline: 25085286]
46. Barrett A. Electronic health record (EHR) organizational change: explaining resistance through profession, organizational experience, and EHR communication quality. Health Commun 2018 Apr;33(4):496-506. [doi: 10.1080/10410236.2016.1278506] [Medline: 28157382]
47. Welton JM, Harper EM. Measuring nursing care value. Nurs Econ 2016;34(1):7-14 [FREE Full text] [Medline: 27055306]
48. Rathert C, Porter T, Mittler J, Fleig-Palmer M. Seven years after meaningful use: physicians' and nurses' experiences with electronic health records. Health Care Manage Rev 2017 Jun 13. [doi: 10.1097/HMR.0000000000000168] [Medline: 28614166]

Abbreviations

BCMA: bar code medication administration
E: theme
EHR: electronic health record
HIT: health information technology
ICU: intensive care unit
IT: information technology
R: round (of data collection)
WOW: workstation on wheels

Edited by G Eysenbach; submitted 11.08.17; peer-reviewed by D Hanauer, A Khan; comments to author 30.11.17; revised version received 08.02.18; accepted 07.04.18; published 26.06.18.

Please cite as:
Zadvinskis IM, Garvey Smith J, Yen PY
Nurses' Experience With Health Information Technology: Longitudinal Qualitative Study
JMIR Med Inform 2018;6(2):e38
URL: http://medinform.jmir.org/2018/2/e38/
doi:10.2196/medinform.8734
PMID:29945862

Original Paper

Validation of a Natural Language Processing Algorithm for Detecting Infectious Disease Symptoms in Primary Care Electronic Medical Records in Singapore

Antony Hardjojo[1], PhD; Arunan Gunachandran[1], MPH; Long Pang[1], BSc (Hons); Mohammed Ridzwan Bin Abdullah[1], RN; Win Wah[1], MPH; Joash Wen Chen Chong[1], B Biomedical Science; Ee Hui Goh[1], MPH; Sok Huang Teo[2], MPH; Gilbert Lim[3], PhD; Mong Li Lee[3], PhD; Wynne Hsu[3], PhD; Vernon Lee[1], PhD; Mark I-Cheng Chen[1,4], PhD; Franco Wong[2,5], MMed; Jonathan Siung King Phang[2,5], MMed

[1]Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore, Singapore

[2]National Healthcare Group Polyclinics, Singapore, Singapore

[3]School of Computing, National University of Singapore, Singapore, Singapore

[4]National Centre for Infectious Diseases, Singapore, Singapore

[5]National University Polyclinics, Singapore, Singapore

Corresponding Author:
Mark I-Cheng Chen, PhD
Saw Swee Hock School of Public Health
National University Health System
National University of Singapore
MD1, #10-01
12 Science Drive 2
Singapore, 117549
Singapore
Phone: 65 65165781
Email: mark_chen@nuhs.edu.sg

Abstract

Background: Free-text clinical records provide a source of information that complements traditional disease surveillance. To electronically harness these records, they need to be transformed into codified fields by natural language processing algorithms.

Objective: The aim of this study was to develop, train, and validate Clinical History Extractor for Syndromic Surveillance (CHESS), an natural language processing algorithm to extract clinical information from free-text primary care records.

Methods: CHESS is a keyword-based natural language processing algorithm to extract 48 signs and symptoms suggesting respiratory infections, gastrointestinal infections, constitutional, as well as other signs and symptoms potentially associated with infectious diseases. The algorithm also captured the assertion status (affirmed, negated, or suspected) and symptom duration. Electronic medical records from the National Healthcare Group Polyclinics, a major public sector primary care provider in Singapore, were randomly extracted and manually reviewed by 2 human reviewers, with a third reviewer as the adjudicator. The algorithm was evaluated based on 1680 notes against the human-coded result as the reference standard, with half of the data used for training and the other half for validation.

Results: The symptoms most commonly present within the 1680 clinical records at the episode level were those typically present in respiratory infections such as cough (744/7703, 9.66%), sore throat (591/7703, 7.67%), rhinorrhea (552/7703, 7.17%), and fever (928/7703, 12.04%). At the episode level, CHESS had an overall performance of 96.7% precision and 97.6% recall on the training dataset and 96.0% precision and 93.1% recall on the validation dataset. Symptoms suggesting respiratory and gastrointestinal infections were all detected with more than 90% precision and recall. CHESS correctly assigned the assertion status in 97.3%, 97.9%, and 89.8% of affirmed, negated, and suspected signs and symptoms, respectively (97.6% overall accuracy). Symptom episode duration was correctly identified in 81.2% of records with known duration status.

Conclusions: We have developed an natural language processing algorithm dubbed CHESS that achieves good performance in extracting signs and symptoms from primary care free-text clinical records. In addition to the presence of symptoms, our algorithm can also accurately distinguish affirmed, negated, and suspected assertion statuses and extract symptom durations.

(JMIR Med Inform 2018;6(2):e36) doi:10.2196/medinform.8204

KEYWORDS

natural language processing; communicable diseases; epidemiology; surveillance; syndromic surveillance; electronic health records

Introduction

Study Background and Rationale

The world continues to be vulnerable to the threat from infectious diseases. This includes novel emerging infections, changes in the incidence or severity of common circulating pathogens, as well as the potential use of infectious agents in bioterrorism. There is thus an interest in developing infectious disease surveillance systems that can detect outbreaks, as well as provide adequate advanced warning of possible surges in incidence or hospitalization burden so as to enlist appropriate public health response efficiently [1].

At present, surveillance of infectious diseases in Singapore, such as in many jurisdictions, is largely passive in nature. In Singapore, this occurs through a central agency, Ministry of Health, which collates information via notifications of key infectious diseases by clinicians and laboratories and also performs weekly retrospective analysis of health care data using broad diagnostic groups [2]. The existing surveillance system with its traditional reliance on physician and laboratory diagnoses and reports has several limitations that may lead to delays in the recognition and notification of an outbreak. These include a dependence on timely recognition and reporting by clinicians, challenges faced by clinicians in recognizing the unexpected presentations of novel pathogens, and delays in obtaining laboratory results for agent identification [3-5]. For novel infections, in particular, the failure to suspect a case, order a laboratory test, or in some instances the unavailability of an accurate diagnostic laboratory assay may all contribute to delays in detection. Moreover, the retrospective nature and coarse grouping of conditions by diagnoses codes with use of only simple thresholds on counts of cases can miss more subtle but important signals that take into account the spatial and contextual relationships between clusters of infectious cases and possible changes in incidence, clinical presentation, or severity, even for commonly circulating pathogens. Singapore currently has universal uptake of electronic health records among its public sector health care providers, and syndromic surveillance systems leveraging on electronic medical records (EMRs) to identify syndromes may help to overcome some of these limitations by providing surveillance data that complement our existing methods for surveillance [5,6]. By grouping symptoms identified into specific syndromes based on the presentation of the illness, we may potentially identify illness clusters that would not otherwise be suspected [7,8], particularly when leveraging off other routinely available information in electronic health records, such as demographic and geolocation data [9]. However, to capture clinical presentation as syndromes requires additional intervention. We could request that doctors remember to and comply with the burden of entering additional data alongside their clinical duties as predefined syndromes (as is currently done for monitoring of influenza-like illness [10]). However, this has several drawbacks, including a need to predefine syndromes with consequent practical limits to the number of case definitions that could be in use, the need to educate all reporting parties on the case definitions, variations in interpretations of these case definitions, and potentially poor compliance. Approaches have also been developed to map diagnoses into syndromes for surveillance [11,12], but these have in some instances been found to be inadequate to detect outbreaks on their own. For instance, Lusigna and colleagues [13] found that an ontological approach to define gastrointestinal disease using all the terms and codes was better than using International Statistical Classification of Diseases and Related Health Problems-10th revision (ICD-10) alone. Another alternative to these approaches would be to rely on natural language processing (NLP) algorithms to extract from free-text information what would be routinely documented by practicing clinicians and transform such data into codified information [8]. However, free-text clinical narratives are rife with abbreviations or shorthand forms, misspellings, synonyms, and contextual information, which poses a challenge to accurately extract clinical information [8,14]. As such, NLP algorithms need to be trained and validated to achieve optimal performance.

Aims and Objectives of the Study

The aim of the study was to describe in detail the process of creating a rule-based NLP algorithm called Clinical History Extractor for Syndromic Surveillance (CHESS) that extracts signs and symptoms associated with infectious diseases outbreaks. We also trained and validated CHESS's performance against a manually coded reference standard and present the results in this paper.

Methods

Study Setting and Algorithm Development

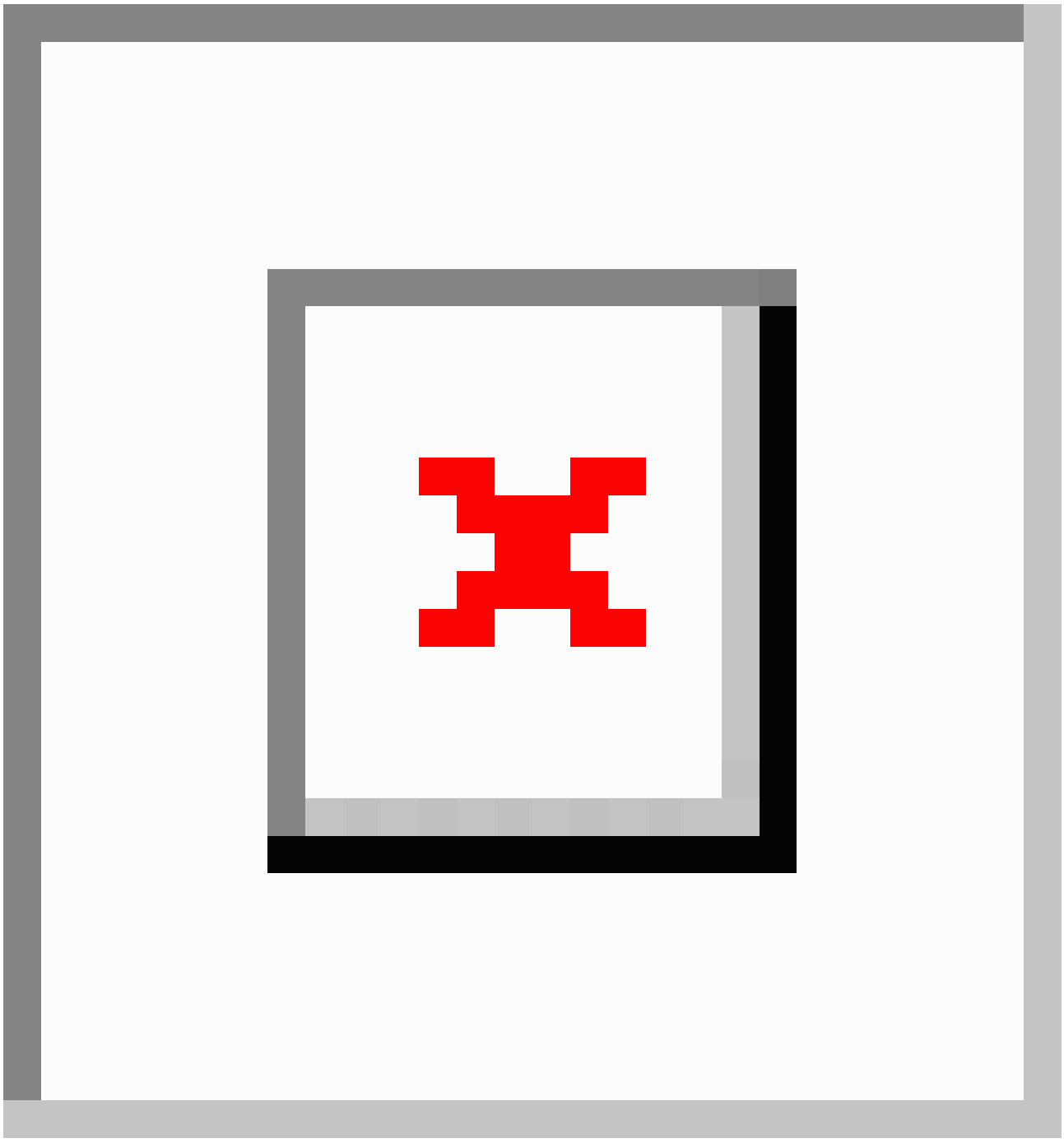
We developed CHESS that adopts concept extraction using a rule-based approach. The tool uses part of speech tagging, prefixes, and regular expressions and incorporates ontology and grammar-based analysis to extract signs and symptoms from free-text notes. We chose the rule-based approach as it was simpler to operate and easier to create and understand than other systems based on machine learning. We felt this would thus also be an appropriate benchmark for the development of iterations of algorithms based on machine learning, which are likely to be developed in the future. Furthermore, with a good keyword dictionary adapted to local context, this tool can be easily updated to incorporate various new features and adapted to other clinical contexts. CHESS targets 48 signs and symptoms of interest from four different syndrome categories: (1) gastrointestinal infection syndromes, (2) respiratory infection syndromes, (3) constitutional signs and symptoms typically present during infectious diseases, and (4) other signs and symptoms not belonging to the former three groups (with the

full categorization displayed in Multimedia Appendix 1). The choice of symptoms were based on infectious disease diagnoses categories currently monitored in Singapore [2] and were sufficiently detailed to give a flexibility to combine symptoms to construct case definitions for detecting possible future outbreaks. The mapping of symptoms to syndromes was modeled after the Centers for Disease Control and Prevention Electronic Surveillance System for the Early Notification of Community-Based Epidemics II framework [11,15].

The process began by constructing a library of keywords associated with the signs and symptoms of interest. We downloaded the 2014AB version of United Medical Language System (UMLS) [16] and identified key medical concepts. We started with the UMLS metathesaurus as it was free to use and had a comprehensive database of over 3 million medical concepts from over 150 libraries including Systematized Nomenclature of Medicine-Clinical Terms and ICD-10-clinical modification, the latter being commonly used in Singapore. The NLP module was built with ANother Tool for Language Recognition (ANTLR), which is an open source Java-based parser generator. This has been modified to include various components as per our requirements. In the first iteration of CHESS, the tool had an overall recall value of 65.4% when tested with a random dataset, indicating that a huge number of terms went undetected (false negatives). This was attributed to shorthand forms, which were common locally and misspellings within free-text notes [14,17] that were not accounted for in the UMLS metathesaurus. To broaden the dictionary and include these terms, CHESS was trained ad-hoc with two small pilot local health care datasets made available to us for preliminary developmental work before further training and validating the process on National Healthcare Group Polyclinics (NHGP) datasets as described in this paper. Training included manual addition of possible terms based on clinical notes.

In Singapore, clinical free-text information is usually short, with each new finding separated by line breaks. In the ontology analysis, the clinical visit free text is separated into phrases by line breaks. Phrases are recursively parsed into tokens for easier categorization according to a set of lexer rules for patterns. These tokens are broadly categorized as symptoms, assertion status, and duration (Figure 1, top portion). Each symptom that is identified from the dictionary has a relational database to incorporate common misspellings, abbreviations, and synonyms. Assertion status is identified by specific terms that determined if there is a negation modifier (eg, *no*, *denies*, and *nil*). In addition, we used another set of terms indicating *suspected* status (eg, *claims* and *?<symptom>*). If these terms were present in the phrase, they will change the assertion status of symptoms in that phrase to *negated* or *suspected*, respectively. Otherwise, symptoms are identified to be *affirmed*. Negation modifiers reverse assertion status of symptom; for example *not afebrile* will be fever affirmed. Conjunction terms such as *and*, *or*, *commas* (ie,) and *slash* (ie/) are used to chain a list of signs and symptoms together in the same phrase. A stopword dictionary was built to remove nonessential words (eg, *over*, *on*, and *before*) that will interfere with exact string matches. In the grammar-based analysis phase, relationships between tokens produced in the ontology analysis are built up to make sense of the sentence (Figure 1, top portion). Patterns and grammar rules were initially built up from the UMLS and modified with inputs from domain experts. In addition, duration tokens were normalized by comparing with a duration dictionary. Temporal attributes are identified by rules that are set to associate duration to appropriate symptoms in proximity to the duration token and by taking into consideration conjunction terms. Instances where a specific onset date was given (either a calendar date or with reference to the date of consultation, for example, today and yesterday) were converted to duration terms, with onset on the day of consultation counted as 0 day and onset yesterday as 1 day. The algorithm was implemented in the ANTLR, which generates a Java implementation from a grammar file.

Figure 1. Ontology and grammar-based analysis of the rule-based natural language processing (NLP) algorithm. Signs and symptoms and information on assertion status and duration are captured and tokenized in the ontology analysis. Relationships between tokens are built up in the grammar-based analysis. C/o: complain of; ST; sore throat.



Symptoms were then manually coded at the phrase level, and this information was used to create episode level symptom coding (Figure 2). The purpose of episode level output was to identify unique symptoms, with useful information on presence and duration from multiple entries of the same symptom in each clinical record. After phrase level symptoms were accurately identified, we utilized a set of rules to achieve episode level output. In infectious disease surveillance, the presence of a symptom (affirmed) in a patient is likely more important information than the similar symptom noted as not being present (negated) during the documentation of the same episode. Thus, the affirmation of a particular symptom was given priority over negation of that symptom recorded elsewhere in the same record. In instances where the presence of a symptom is suspected, this symptom is made void when the same symptom is negated or affirmed elsewhere in the clinical record of that episode, as we considered to this to be less certain than affirmation or negation. For symptom duration, both the manual coding and the NLP tool would identify multiple instances of symptom duration occurring at the phrase level within the same episode. Symptoms specified to have lasted for “few days” were considered to be unspecified but acute symptoms. Symptoms lasting more than 7 days, or indicated as beginning “last week” are grouped together as chronic (>1 week) symptoms. To simplify the analysis, we chose to summarize the data using the symptom

with the earliest onset (ie, the longest duration) at the episode level, which we then compared against the reference standard.

Dataset Used and Training and Validation Process

Our data was obtained from the NHGP, a major public sector chain of clinics estimated to provide about 10% of the primary care in Singapore. To facilitate batch extraction, we chose three clinics (one each from the West, North, and Central regions), then performed a stepwise random sampling of the records across the period from which EMR was available from the middle of 2009 to June 2014. For each clinic, we randomly selected 10 dates that did not fall on a Sunday or public holiday; these dates were evenly divided into 10 half-yearly periods from across the 5-year period for which the EMR was available. Subsequently, for each selected date, 56 records with at least three lines of free-text notes were randomly selected, thus giving a total of 560 records of consultation clinical records from each of the three clinics across the 5-year period. As free-text notes could potentially contain identifiable information, to comply with personal data protection regulations, every record was vetted (and where necessary redacted) by an internal staff member of NHGP before it was shared with the wider collaborative research team (including those from other institutions) for further analysis, and this process limited the total number of records that could be extracted and shared.

Figure 2. Sample set of clinical notes and transformation following phrase-level manual coding and episode-level coding. Abd: abdominal; NA: not applicable; NKDA: no known drug allergy; PMHX: past medical history; RIF: right iliac fossa.

Phrase number	Clinical notes	Phrase level manual coding	Episode level manual coding
1	NKDA		
2			
3	PMHX		
4	thyroid disease		
5	admitted for fever feb 2012	5| Fever|Not Related|NA	
6			
7	Came in due to:		
8			
9	RIF pain 2/7	9| Abdominal pain|Affirmed|2 Days	9| Abdominal pain|Affirmed|2 Days
10	fever noted in clinic	10| Fever|Affirmed|0 Days	11| Fever|Affirmed|0 Days
11	No vomiting episodes	11| Vomiting|Negated|NA	11| Vomiting|Negated|NA
12	No dysuria / hematuria	12| Dysuria|Negated|NA 12| Hematuria|Negated|NA	12| Dysuria|Negated|NA 12| Hematuria|Negated|NA
13	? fever	13| Fever|Suspected|NA	
14	Abd pain unbearable today	14| Abdominal pain|Affirmed|0 Days	

Figure 3. Flowchart of process for creating reference standard.

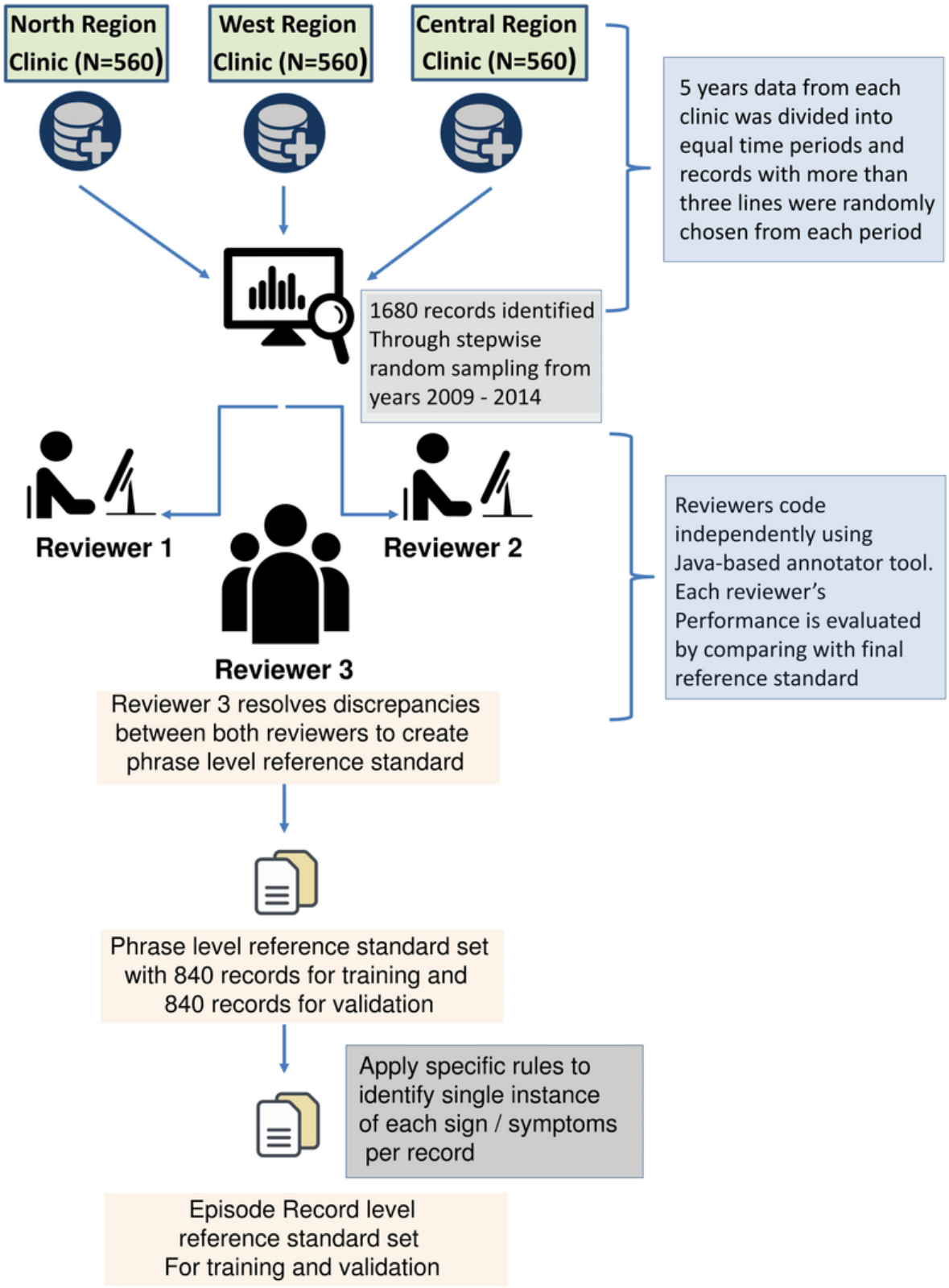


Figure 3 describes the process by which we used manual review by human coders on all extracted records to create a reference standard to train and validate CHESS's algorithm. A Java-based annotator interface tool was created to improve manual coding methodology and prevent mistakes. Two independent human reviewers, who were health care workers with substantial experience in clinical research and interpreting medical case notes, read through the clinical records and then annotated each line of the record for the presence of signs and symptoms, the assertion status, and duration of symptoms experienced; this

allowed us to capture multiple instances where a sign or symptom appeared within each record. Then a third reviewer, a clinician who has practiced in the primary care setting, served as the adjudicator in instances where the two reviewers were not in agreement.

Training of the algorithm was conducted with 840 manually annotated records to improve CHESS's performance by identifying new terms, misspellings, and shorthand forms to be updated into the pattern and grammar library, or removing keywords that caused significant false detections. We repeated several rounds of training until we achieved satisfactory performance with the training dataset. Following training, CHESS's performance was validated on the remaining 840 notes that were independent from the training set to test the algorithm's robustness in correctly identifying signs and symptoms.

Analysis of Natural Language Processing Performance Compared With Manual Coding Performance and Reference Standard

CHESS's performance was assessed by its precision, recall, and F-measure for detecting signs and symptoms in comparison with the adjudicated reference standard. This was performed at both the phrase level and episode level. The precision, recall, and F-measure were defined by the following formulae, where true positive refers to signs and symptoms that were accurately identified by CHESS, false positive refers to signs and symptoms incorrectly identified, and false negatives refers to signs and symptoms missed:

- Precision = (True Positive) / (True Positive + False Positive)
- Recall = (True Positive) / (True Positive + False Negative)
- F–measure = (2 × [Precision × Recall]) / (Precision + Recall)

Precision is the frequency with which symptoms identified by the tool are relevant (positive predictive value). Recall is the frequency with which relevant symptoms are identified (sensitivity). We used F-measure, which is a weighted harmonic mean of precision and recall to give an overall picture of the tool's performance. F-measure is applicable for our situation as we do not have true negatives, yet had a reference standard to compare with [18]. Again, because we do not have a true negative, we could not use Cohen Kappa statistic to review interrater reliability, and so the same metrics were also used to assess each manual coder's performance against the final reference standard.

To assess the performance of CHESS for symptom identification, assertion status, and duration, we utilized the dictionary after it was trained with 840 records on the validation set. The performance of CHESS in capturing specific symptoms was assessed for all symptoms and also stratified by individual symptoms and visualized in a bubble chart. Performance in assigning the correct assertion status to symptoms was assessed on all true positive symptoms that were identified by both NLP and reference standard. A matrix plot was created to see where the capacity to identify assertion status was lacking.

Finally, in syndromic surveillance, an episode level onset date is helpful in characterizing the temporality of an infection, and this can be imputed if the duration of symptoms is known at the time of consultation. We present the proportion of episodes in the reference standard with a valid episode level duration (based on the earliest symptom) that were correctly identified, with additional stratification by episode duration for acute symptoms.

We also conducted a qualitative review of instances where the NLP algorithm failed to correctly identify symptoms, assertion statuses, and symptom duration and describe the potential areas for improvement.

Results

Description of Data Source and Common Symptoms Identified

For the 1680 primary care clinical records extracted, there were no significant differences on the genders of the patients across the North, West, and Central clinics. However, a significantly higher proportion of Chinese ethnicity (compared with Malay and Indian) and significantly older population was observed in the clinic from the North. This was in concordance with the overall population distribution within the three districts based on national demographic surveys [19]. Consequent to the older case mix, the clinic from the North also had more consultations for chronic diseases than the other two clinics.

Table 1 shows the frequencies of the 10 most commonly detected signs and symptoms from the 1680 records reviewed by human coders (full list of signs and symptoms displayed in Multimedia Appendix 1). Overall, fever was detected most frequently (12.05% [928/7703] of all instances of symptoms detection) but was in the large majority of instances "negated." Other common signs and symptoms detected within the clinical records were those associated with upper respiratory tract infections such as cough, sore throat, rhinorrhea, and sputum, and these were in the majority of instances affirmed (between 70.38% and up to 86.82%).

Comparison of Natural Language Processing Against Human Coders in Identifying Signs and Symptoms in Free Text

Table 2 shows that for phrase level output, both human coders have good agreement with the final adjudicated output used as the reference standard other than for a slightly lower recall for coder 2 (because of differences in interpretation of clinical examination findings and abbreviations). The final round of training led to sufficient performance, with CHESS having a precision of 95.3% and recall of 96.2% with the training set; levels which were fairly similar to those of the human clinical coders. CHESS also achieved a precision and recall of 94.2% and 90.4% with the validation set, with the lower performance because of our limitations in identifying (through the training dataset) all relevant phrase-level terms present in the validation dataset. Results for episode-level analysis (Table 3) were better, with the performance again being comparable with the human coders, with a precision and recall of 96.7% and 97.6% in the training dataset and 96.0% and 93.1% in the validation dataset, respectively.

Table 1. Frequency of the 10 most commonly detected signs and symptoms within 1680 primary care clinical records by human coders.

Symptoms sorted by frequency of symptom mention in episode level	All instances (N=7703), n (%)[a]	Instance of symptom affirmation, n (%)[b]
Fever	928 (12.04)	228 (24.6)
Cough	744 (9.66)	646 (86.8)
Sore throat	591 (7.67)	416 (70.4)
Rhinorrhea	552 (7.17)	435 (78.8)
Altered state of consciousness	376 (4.88)	7 (1.9)
Vomiting	347 (4.50)	75 (21.6)
Rash	345 (4.48)	72 (20.7)
Dyspnea	286 (3.71)	31 (10.8)
Diarrhea	271 (3.52)	137 (50.6)
Sputum	256 (3.32)	212 (82.8)

[a]Column percentages, with the denominator being all instances (N=7703).

[b]Row percentages, with the denominator being the instances where the symptom in that row appears (eg, for Fever, n=928).

Table 2. Phrase level precision, recall, and F-measure of human coders and Clinical History Extractor for Syndromic Surveillance (CHESS) outputs compared against instances of symptom occurrences in reference standard.

Performance against reference standard	Comparison of coder 1 versus coder 2 (N=8861 instances)		CHESS performance for training set (after training of dictionary, N=4282 instances)	CHESS performance for validation set (after training of dictionary, N=4578 instances)
	Coder 1	Coder 2		
Precision, %	98.52	96.06	95.24	94.15
Recall, %	96.93	84.30	96.17	90.39
F-measure, %	97.72	89.80	95.70	92.23

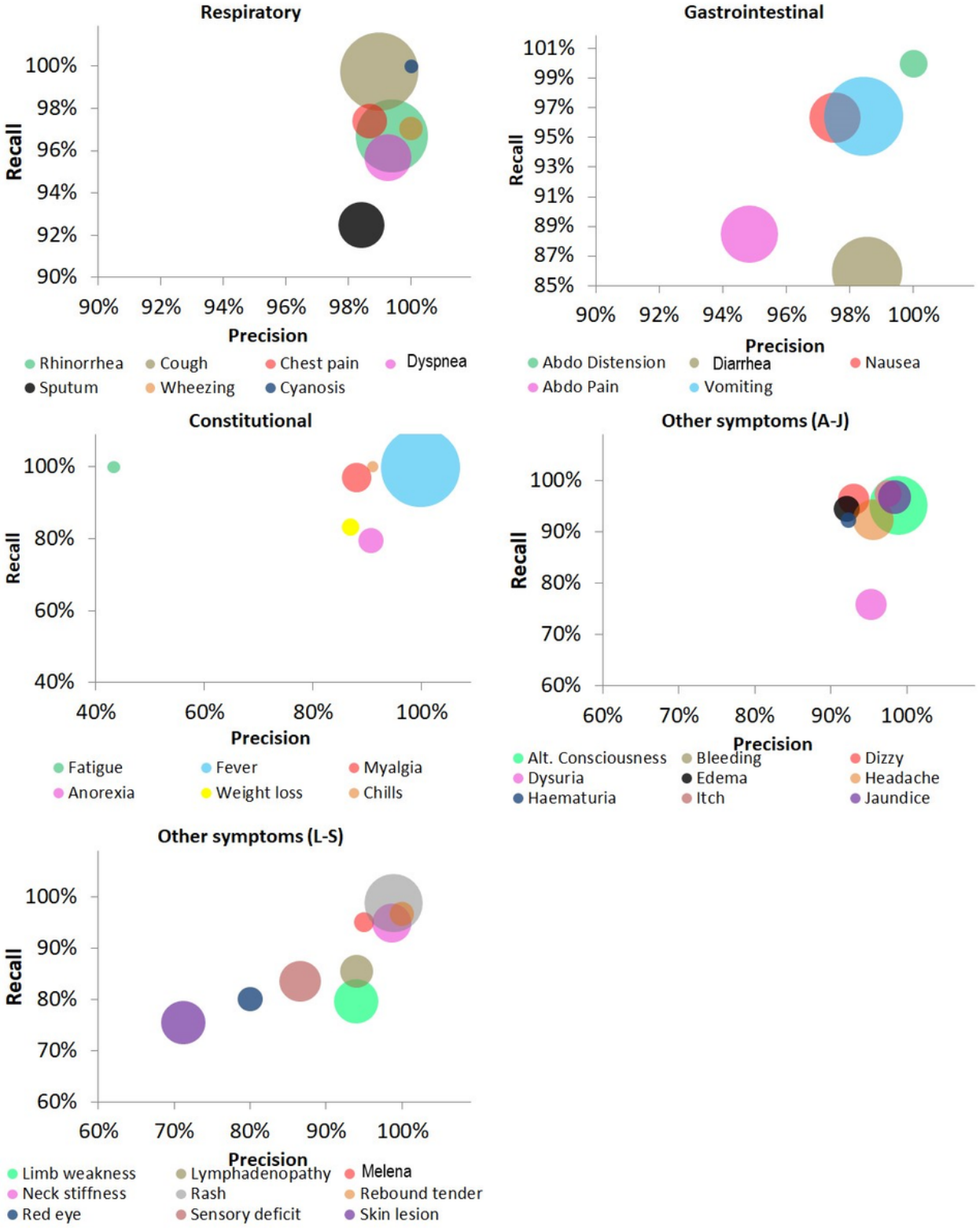
[a]CHESS: Clinical History Extractor for Syndromic Surveillance.

Table 3. Episode level precision, recall, and F-measure of human coders and Clinical History Extractor for Syndromic Surveillance (CHESS) outputs compared against instances of symptom occurrences in reference standard.

Performance against reference standard	Comparison of coder 1 versus coder 2 (N=7703 instances)		CHESS[a] performance for training set (after training of dictionary, N=3738 instances)	CHESS performance for validation set (after training of dictionary, N=3965 instances)
	Coder 1	Coder 2		
Precision, %	98.91	97.13	96.74	95.97
Recall, %	97.46	88.47	97.65	93.06
F-measure, %	98.18	92.58	97.19	94.49

Figure 4 gives CHESS's performance for specific signs and symptoms that occur in more than 1% of the medical records (see supplementary table E2 and E3 for detailed breakdown) using the validation dataset. High precision and recall of >90% were achieved for most signs and symptoms associated with respiratory and gastrointestinal syndromes. "Diarrhea" and "abdominal pain" had slightly lower recall (<90%) in the validation set, but this was limited to records where these were "negated"; recall was 97.7% and 90.6%, respectively, when diarrhea and abdominal pain was "affirmed" vs only 71.0% and 85.0% when "negated." This was because of clinicians entering misspelled words (eg, "supropubic" pain) and new terminologies (eg, RIF) that CHESS was not able to identify resulting in high false negatives. Bleeding had the poorest recall of the symptoms with only 60.8%. This is because the word "blood" was intentionally omitted from CHESS's list of keywords because of the generic use of the word for unrelated tests and measurements (eg, blood test and blood pressure). As such, adding "blood" into the list would have generated many false positives leading to an even worse precision for bleeding. On the other hand, fatigue was found to have a poor precision of 45.5%. This was because of the word "weakness," also commonly used to describe limb weakness. This resulted in false positives for fatigue and false negatives for limb weakness. However, the overall recall for limb weakness was still above 80% because of the large number of true positive instances (n=175).

Figure 4. Bubble chart of the Clinical History Extractor for Syndromic Surveillance's (CHESS's) precision and recall for each sign and symptom in episode level analysis for the validation dataset. Each bubble denotes a single symptom categorized into symptom types: respiratory, gastrointestinal, constitutional, and others. Bubble size is proportional to the number of cases identified by humans (true positive + false negative). Symptoms present in less than 1% of records are not presented.



Accuracy of Natural Language Processing in Identifying Assertion Status and Duration of Symptoms

CHESS also performed well in assigning the correct assertion status to the signs and symptoms correctly identified (ie, true positives). Of 3690 instances of true positives in the validation dataset, 1728 (46.83%), 1937 (52.49%), and 25 (0.68%) were determined as affirmed, negated, and suspected, respectively, in the reference standard (Figure 5). CHESS correctly assigned the assertion status of signs and symptoms for 96.9% of instances when they were affirmed, 97.5% when they were

negated, and 92.0% when they were suspected, with an overall accuracy of 97.2%. Sources of error mainly arose in three ways. First, as our tool relied on using line breaks to separate out phrases, when the whole visit was entered without any appropriate conjunction keywords in one line instead of multiple lines, the assertion status would be deemed by the NLP to apply to the all the symptoms in that line. Although this was rare given the prevailing styles of clinical text data entry, it did result in a few instances of misclassification for assertion statuses. Second, misclassification by CHESS of a symptom as affirmed occasionally occurred when doctors advised a patient of future symptoms to watch out for. Third, the keywords learned from our training dataset to identify instances where a symptom as "suspected" were not exhaustive for all the instances found in the validation dataset.

Of 778 records with at least one sign or symptom detected in the validation dataset, 583 (75.0%) included information on the duration of the episode (Figure 6). The majority (53.3%) of these had acute onset within the past 2 days. In terms of accuracy, CHESS had an overall accuracy of 83% for detecting and assigning the correct duration. Performance was degraded largely because the rules devised based on the training dataset were not exhaustive. There were many abbreviations such as "y" for years and instances such as "2days" where the words and numbers occurred together (without an intervening space), as well as misspellings, most of which were apparent only on reviewing classification errors for the validation dataset.

Figure 5. Clinical History Extractor for Syndromic Surveillance's (CHESS's) accuracy in identifying assertion status of symptoms within episode level analysis based on the validation dataset.

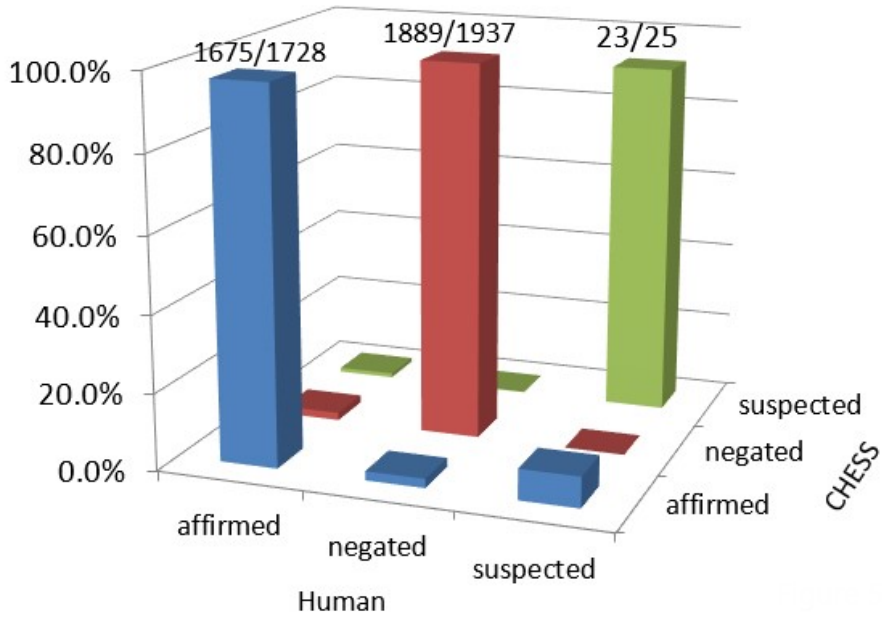
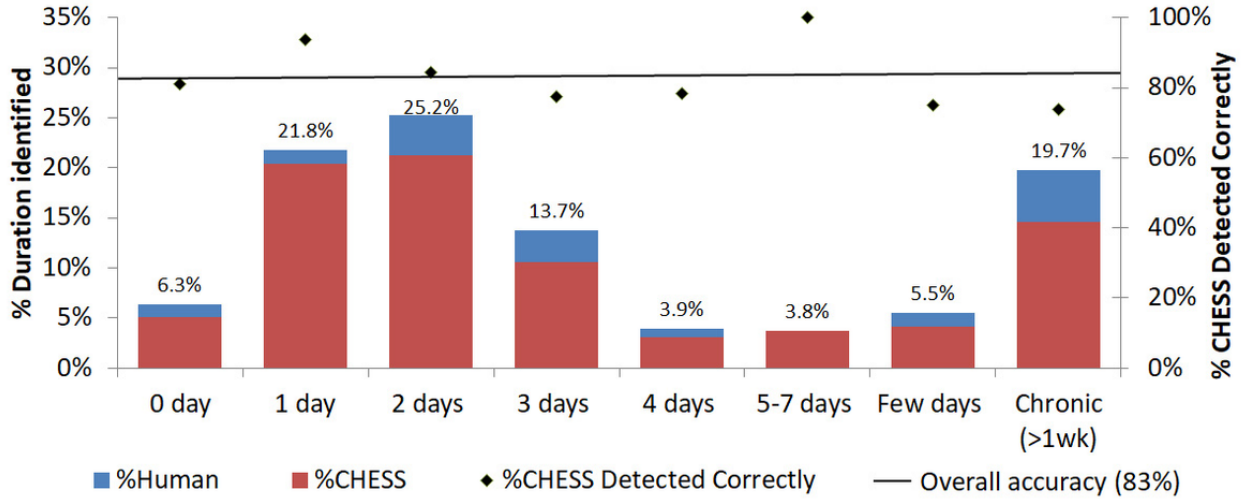


Figure 6. Episode level analysis on the distribution of symptom episode duration in instances detected by human coders (blue) among all the National Healthcare Group Polyclinics (NHGP) records and the distribution of durations detected by Clinical History Extractor for Syndromic Surveillance (CHESS; red) based on the validation dataset. Diamonds give the proportion of records where CHESS correctly identifies and assigns the duration information stratified by episode duration (based on the reference standard), with the horizontal line giving the aggregated accuracy for detection of symptom duration for all records analyzed.



Discussion

Principal Findings

In this work, we have described the design and performance of CHESS, a rule-based NLP algorithm that we showed is reasonably accurate in extracting information on symptoms, assertion status, and the duration of symptoms from free-text clinical notes in a set of EMR from a large primary care provider.

The performance of CHESS is comparable with results from other systems in the extant literature. For instance, a system developed by McRae et al [20] to identify influenza-like-illness from unstructured primary care notes reported a precision of 87.8% and recall of 90.2% on their validation set relative to a human reference standard. Another system, the Multi-Threaded Clinical Vocabulary Server (MCVS), was used by Matheny et al to identify specific symptoms suggestive of tuberculosis, hepatitis, and influenza from clinical notes [21]. That MCVS-based system, which can also identify assertion status, had an overall performance of 91.2% precision and 83.5% recall for detecting symptoms. The system was able to correctly identify 84.7% of positive assertions, 75.1% of negative assertions, and 0.7% of uncertain assertions. Elkin and colleagues also reported that the MCVS achieved a sensitivity of 92.9% and a specificity of 34.6% in identifying influenza infection in patients [22]. MCVS uses medical concepts from SNOMED-CT terminology, and it was noted in one study that ontology based on SNOMED-CT have a precision of 99.8% and recall of 99.7% in identifying medical problems [23].

The ability of CHESS to accurately discern whether the signs and symptoms were affirmed, negated, or suspected with very high accuracy (97.2% for validation set) is an important feature when monitoring primary care records for specific case definitions associated with particular infections. In our study, we noted that a large portion of the symptoms identified in the clinical notes were of negated status as the clinician was eliminating symptoms of the key differential diagnoses. For example, “fever” was the most frequently detected symptom, but it was far more often negated than affirmed. Failure to distinguish negation from affirmation could potentially lead to significant background noise that may mask the signal from a real outbreak.

The other key novel capability of CHESS compared with other NLP systems worth highlighting is the extraction of symptom duration in addition to the assertion status. This is particularly critical in primary care data, given that consultations for infectious disease conditions such as upper respiratory tract infections would be more common on Monday than on other days of the week and be the lowest during weekends when not all clinics are open [24]. This day-of-week variation in visitation rates potentially necessitates setting of higher daily thresholds for signaling an outbreak that again may reduce our sensitivity to detect outbreak signals [25]. Extracting symptom duration allows us to impute an estimated day of onset instead of relying only on the day of consultation. This potentially reduces day-of-week effects, with consequent improvements in temporal resolution for EMR-based surveillance.

Application of Clinical History Extractor for Syndromic Surveillance (CHESS) to Infectious Disease Surveillance

We intentionally designed CHESS to extract individual signs and symptoms rather than predefined syndromes. Such a design facilitates the use of specific case definitions involving combinations of individual symptoms. For instance, in 2016, an outbreak of Zika virus infections in Singapore was detected when an astute primary care physician reported a cluster of patients presenting with fever, rash, and joint pains [26]. Our system would have the flexibility of including additional symptoms for a Zika virus case definition, such as conjunctivitis, which was also associated with Zika virus infections. We can also tailor case definitions to new emerging infections of concern, then monitor for unexpected clusters of such cases anywhere within the reach of our EMR systems. Other applications could include surveillance for changes in incidence or severity of commonly circulating infections of concern, such as influenza. In such an application, we could track incidence of a syndrome comprising acute onset of fever, cough (for which our algorithm performed fairly well), and a body temperature ≥38 C (which is a coded field in NHGP EMR) that has reasonable discriminatory value for influenza in primary care [24,27]. Combining this with hospital admissions for influenza can potentially allow us to assess age-stratified incidence and severity, which has been known to differ between influenza epidemic as well as influenza pandemic strains [28]. Applications in these areas will require further validation for specific syndromes of interest, such as by comparing disease incidence estimated from primary care data, in this case through EMR, to other independent methods [29]. Such validation work should also look into approaches that combine free text based with codified information such as diagnosis codes and incorporating other sources of information such as laboratory data and procedural data to see if this adds value to detection and monitoring of infectious disease epidemics beyond what is currently possible through our current surveillance modalities.

Future Work and Limitations

Other future work on CHESS to consider would include incorporating qualitative descriptions of severity. Such terms, either at the overall episode level, or in association with specific symptoms, could potentially add value to surveillance or even diagnosis of infectious (and possibly noninfectious) conditions. The current tool had components such as general condition (well, good, fair, poor, and alert) and appearance (toxic and nontoxic), but the primary care records available for analysis did not have sufficient data to allow us to validate this function. Validation would require implementing CHESS on a larger set of records and a more diverse set of free-text notes from primary care as well as emergency departments. Such expanded coverage would likely enhance our ability to discern signals from infectious disease outbreaks. It occurs to us that, having validated the algorithm at the level of phrases containing various symptoms, the work also sets the foundation for NLP tools to be used outside the confines of syndromic surveillance. For instance, new symptoms can easily be added to the dictionary to expand the application of the tool to noncommunicable disease–related conditions, to attempt what has been done using

other systems, for instance, to classify clinical problem lists and detect postoperative complications [23,30]. However, such applications may need recognition of symptoms described by anatomy and be able to interpret other terms expressing uncertainty in the assertion status. This would require improvements to our current tool, including contextual learning modules that identifies terms based on where they are placed in the clinical record.

Furthermore, several limitations in our work should be acknowledged. First, we have described some of the weaknesses in CHESS's algorithms from our qualitative review of those instances where misclassification occurred. We have already added keywords and misspellings identified in the validation set to the current version of CHESS, although a more generalizable way of dealing with misspellings would be ideal. Other improvements needed include an algorithm to identify different sentences within a single line of text (which we have since implemented) and a module to distinguish instances when the doctor advises the patient of future symptoms from currently reported symptoms (that we are now building). There were insufficient instances of these occurrences in primary care notes to allow us to validate these enhancements, but such advice upon discharge is likely to occur at much higher frequencies in emergency department as compared with primary care EMR.

Furthermore, it must be noted that although NHGP is a major primary care provider in Singapore and currently has the information systems to allow near-time access to their EMR for the NLP algorithm to be viably implemented in their context, it is unclear if the infrastructure for other primary care providers and emergency departments can support real-time surveillance. Singapore currently has a National Electronic Health Record system that receives contributions several times daily from various providers, including both primary care providers and emergency departments, and we are currently exploring the feasibility of using that as a platform to implement CHESS. However, in doing so, we must also recognize that CHESS was trained and validated only on NHGP notes. We expect additional shorthand forms, misspellings, and terminologies should our system be extended to other primary care systems, or to records from emergency departments (though preliminary testing of CHESS on a set of emergency department notes showed a good albeit slightly lower performance for our primary care notes with 93.2% precision and 86.3% recall). Even for use within NHGP, we acknowledge that the introduction of additional words and terms because of factors such as staff turnover or new methods of documentation may cause degradation of performance. Prospective implementation would hence require periodic revalidation, with retraining instituted should the performance drop below a satisfactory standard. These weaknesses are inherent in the keyword-based approach we adopted, where an exact match for a specific string of characters is required for detection by the parser; any new terms thus has to be manually added into the algorithm dictionary to be detected.

Currently there is a trend toward using automated approaches to NLP to identify new ontology and improve detection sensitivity, and this would be an alternative to manually adding keywords. However, there are several potential issues with such approaches. Topic modeling, for instance, requires large numbers of medical notes to come up with concept similarities within unstructured data. Although at a glance, this method of building the ontology may appear to be simple, it still needs to be verified and manually supervised. Moreover, in a study by Arnold et al [31], it was noted that the Latent Dirichlet allocation method of identifying topics resulted in lesser number of interpretable topics than a primary physician could identify. This was attributed to the fact that the Latent Dirichlet allocation method needs to identify topics from a highly specialized collection with a large vocabulary of related medical terminologies, which is not feasible without supervision. In another study, it was noted that the corpus for training NLP had almost 30% redundancy, where the doctor copies and pastes previous medical histories of a single patient. Redundancy can also occur when doctors at a hospital use a template for data entry (or in some cases for standard advice given to patients with a particular set of diagnoses). In such cases, as the NLP is trained using topic modeling, an inherent bias is created because of the increased probability of the co-occurrence of specific words [32]. Other automated NLP systems such as SimStat also require manual input to create an inclusion and exclusion dictionary from the list of words most frequently found and may thus also not be time-efficient [33]. Furthermore, clinical notes are full of spelling mistakes, abbreviations, and multi-word phrases, which makes it harder for automated NLP tools to identify patterns of occurrence. In this particular instance, the NHGP clinical notes were mostly short with an average of 4.7 words per phrase (maximum of 35 words per phrase), and each record had an average of 10 phrases. This was likely because of the high workload in the primary care setting, where clinicians had less time for more extensive documentation. Notes were hence to the point but rife with abbreviations and misspellings, and it is hence uncertain how an automated NLP technique might have performed. Therefore, although our method of manual coding to identify keywords was time-consuming, it proved to have sufficient performance, and we see it as a necessary step to serve as a benchmark algorithm for future work using automated NLP techniques. Moreover, the dictionary of terms used in local clinical practice that we compiled, though certainly not exhaustive, is an invaluable resource that can be exported into other systems to improve detection rates. For instance, IDEAL-X, an online machine-learning tool, requires a list of control vocabulary terms to improve on its statistical models of automated NLP [34].

Conclusions

In conclusion, we have described the process of developing and validating CHESS, an NLP algorithm to extract information on signs and symptoms, along with information on assertion status and symptom duration from free-text primary care notes that we intend to make available for free download for researchers to access and build on. This simple rule-based concept extraction NLP tool could achieve good precision and recall approaching that for manual identification of symptoms and accurately identified most of the common infectious disease-related symptoms. Problems with performance were mainly because of the instances where we wanted to reduce false positives while improving sensitivity for a small proportion of situations where

the documentation style was unusual or not found in our training dataset. Future steps would be to implement CHESS on a larger set of records and develop approaches to combine free text based with codified information such as diagnosis codes while comparing the outputs of such approaches with those from existing surveillance systems. There is also a need to test CHESS on a more diverse set of free-text notes from primary care as well as emergency departments, as expanded coverage would likely enhance our ability to discern signals from infectious disease outbreaks. We should also simultaneously test if newer approaches based on machine learning can serve as a more efficient and similarly effective way of updating our NLP algorithms.

Acknowledgments

This project is funded by the Singapore National Medical Research Council (NMRC/CS-IRG/1384/2014; CIRG13nov041).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Three supplementary tables, which include frequency of symptom in 1680 episodes, performance of CHESS algorithm on training dataset, and performance on validation dataset.

[PDF File (Adobe PDF File), 398KB - medinform_v6i2e36_app1.pdf]

References

1. Hope K, Durrheim DN, d'Espaignet ET, Dalton C. Syndromic surveillance: is it a useful tool for local outbreak detection? J Epidemiol Community Health 2006 May;60(5):374-375 [FREE Full text] [Medline: 16680907]
2. Ministry of Health, Singapore. 2018. Communicable Diseases Surveillance in Singapore 2016 URL: https://www.moh.gov.sg/content/moh_web/home/Publications/Reports/2017/communicable-diseases-surveillance-in-singapore-2016.html [accessed 2018-02-05] [WebCite Cache ID 6x7cX1g61]
3. Morse SS. Public health surveillance and infectious disease detection. Biosecur Bioterror 2012 Mar;10(1):6-16 [FREE Full text] [doi: 10.1089/bsp.2011.0088] [Medline: 22455675]
4. Levin JE, Raman S. Early detection of rotavirus gastrointestinal illness outbreaks by multiple data sources and detection algorithms at a pediatric health system. AMIA Annu Symp Proc 2005:445-449 [FREE Full text] [Medline: 16779079]
5. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V, CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR Recomm Rep 2004 May 7;53(RR-5):1-11 [FREE Full text] [Medline: 15129191]
6. Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004;11(2):141-150 [FREE Full text] [doi: 10.1197/jamia.M1356] [Medline: 14633933]
7. Chapman WW, Dowling JN, Baer A, Buckeridge DL, Cochrane D, Conway MA, et al. Developing syndrome definitions based on consensus and current use. J Am Med Inform Assoc 2010;17(5):595-601 [FREE Full text] [doi: 10.1136/jamia.2010.003210] [Medline: 20819870]
8. Kimia AA, Savova G, Landschaft A, Harper MB. An introduction to natural language processing: how you can get more from those electronic notes you are generating. Pediatr Emerg Care 2015 Jul;31(7):536-541. [doi: 10.1097/PEC.0000000000000484] [Medline: 26148107]
9. Smith S, Elliot AJ, Mallaghan C, Modha D, Hippisley-Cox J, Large S, et al. Value of syndromic surveillance in monitoring a focal waterborne outbreak due to an unusual Cryptosporidium genotype in Northamptonshire, United Kingdom, June - July 2008. Euro Surveill 2010 Aug 19;15(33):19643 [FREE Full text] [Medline: 20738999]
10. World Health Organization. 2018 Feb. WHO global technical consultation: global standards and tools for influenza surveillance URL: http://apps.who.int/iris/bitstream/10665/70724/1/WHO_HSE_GIP_2011.1_eng.pdf [WebCite Cache ID 6yeAmfUKG]
11. Centers for Disease Control and Prevention. Syndrome definitions for diseases associated with critical bioterrorism-associated agents URL: https://emergency.cdc.gov/surveillance/syndromedef/pdf/syndromedefinitions.pdf [WebCite Cache ID 6x7bSHoIV]
12. Lombardo JS, Burkom H, Pavlin J. ESSENCE II and the framework for evaluating syndromic surveillance systems. MMWR Suppl 2004 Sep 24;53:159-165 [FREE Full text] [Medline: 15714646]
13. de Lusignan S, Shinneman S, Yonova I, van Vlymen J, Elliot AJ, Bolton F, et al. An ontology to improve transparency in case definition and increase case finding of infectious intestinal disease: database study in English general practice. JMIR Med Inform 2017 Sep 28;5(3):e34 [FREE Full text] [doi: 10.2196/medinform.7641] [Medline: 28958989]

14. Travers DA, Haas SW. Using nurses' natural language entries to build a concept-oriented terminology for patients' chief complaints in the emergency department. J Biomed Inform 2003;36(4-5):260-270 [FREE Full text] [Medline: 14643721]
15. Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Loschen W, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J Urban Health 2003 Jun;80(2 Suppl 1):i32-i42 [FREE Full text] [Medline: 12791777]
16. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
17. Ruch P, Baud R, Geissbühler A. Evaluating and reducing the effect of data corruption when applying bag of words approaches to medical records. Int J Med Inform 2002 Dec 04;67(1-3):75-83 [FREE Full text] [Medline: 12460633]
18. Powers D. 2015 Mar 22. What the F-measure doesn't measure: Features, Flaws, Fallacies and Fixes URL: http://arxiv.org/abs/1503.06410 [accessed 2018-02-10] [WebCite Cache ID 6x7cAR0Um]
19. Department of Statistics, Singapore. 2018. Statistics Singapore - Geographic Distribution Internet URL: https://www.singstat.gov.sg/statistics/browse-by-theme/geographic-distribution [accessed 2018-04-13] [WebCite Cache ID 6yeB07Zi4]
20. MacRae J, Love T, Baker MG, Dowell A, Carnachan M, Stubbe M, et al. Identifying influenza-like illness presentation from unstructured general practice clinical narrative using a text classifier rule-based expert system versus a clinical expert. BMC Med Inform Decis Mak 2015 Oct 06;15:78 [FREE Full text] [doi: 10.1186/s12911-015-0201-3] [Medline: 26445235]
21. Matheny ME, Fitzhenry F, Speroff T, Green JK, Griffith ML, Vasilevskis EE, et al. Detection of infectious symptoms from VA emergency department and primary care clinical documentation. Int J Med Inform 2012 Mar;81(3):143-156 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.11.005] [Medline: 22244191]
22. Elkin PL, Froehling DA, Wahner-Roedler DL, Brown SH, Bailey KR. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. Ann Intern Med 2012 Jan 03;156(1 Pt 1):11-18. [doi: 10.7326/0003-4819-156-1-201201030-00003] [Medline: 22213490]
23. Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc 2006 Jun;81(6):741-748. [doi: 10.4065/81.6.741] [Medline: 16770974]
24. Jiang L, Lee VJ, Lim WY, Chen MI, Chen Y, Tan L, et al. Performance of case definitions for influenza surveillance. Euro Surveill 2015 Jun 04;20(22):21145 [FREE Full text] [Medline: 26062645]
25. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. Proc Natl Acad Sci U S A 2003 Feb 18;100(4):1961-1965 [FREE Full text] [doi: 10.1073/pnas.0335026100] [Medline: 12574522]
26. Singapore Zika Study Group. Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis. Lancet Infect Dis 2017 Aug;17(8):813-821. [doi: 10.1016/S1473-3099(17)30249-9] [Medline: 28527892]
27. Pattie DC, Atherton MJ, Cox KL. Electronic influenza monitoring: evaluation of body temperature to classify influenza-like illness in a syndromic surveillance system. Qual Manag Health Care 2009;18(2):91-102. [doi: 10.1097/QMH.0b013e3181a0274d] [Medline: 19369852]
28. Goh EH, Jiang L, Hsu JP, Tan LW, Lim WY, Phoon MC, et al. Epidemiology and relative severity of influenza subtypes in Singapore in the post-pandemic period from 2009 to 2010. Clin Infect Dis 2017 Nov 13;65(11):1905-1913. [doi: 10.1093/cid/cix694] [Medline: 29028950]
29. Lee VJ, Chen MI, Yap J, Ong J, Lim W, Lin RT, et al. Comparability of different methods for estimating influenza infection rates over a single epidemic wave. Am J Epidemiol 2011 Aug 15;174(4):468-478 [FREE Full text] [doi: 10.1093/aje/kwr113] [Medline: 21719743]
30. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. J Am Med Assoc 2011 Aug 24;306(8):848-855. [doi: 10.1001/jama.2011.1204] [Medline: 21862746]
31. Arnold CW, Oh A, Chen S, Speier W. Evaluating topic model interpretability from a primary care physician perspective. Comput Methods Programs Biomed 2016 Feb;124:67-75 [FREE Full text] [doi: 10.1016/j.cmpb.2015.10.014] [Medline: 26614020]
32. Cohen R, Elhadad M, Elhadad N. Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. BMC Bioinformatics 2013 Jan 16;14:10 [FREE Full text] [doi: 10.1186/1471-2105-14-10] [Medline: 23323800]
33. Duz M, Marshall JF, Parkin T. Validation of an improved computer-assisted technique for mining free-text electronic medical records. JMIR Med Inform 2017 Jun 29;5(2):e17 [FREE Full text] [doi: 10.2196/medinform.7123] [Medline: 28663163]
34. Zheng S, Lu J, Ghasemzadeh N, Hayek S, Quyyumi A, Wang F. Effective information extraction framework for heterogeneous clinical reports using online machine learning and controlled vocabularies. JMIR Med Inform 2017 May 09;5(2):e12. [doi: 10.2196/medinform.7235] [Medline: 28487265]

Abbreviations

ANTLR: ANother Tool for Language Recognition
CHESS: Clinical History Extractor for Syndromic Surveillance
EMR: electronic medical record
ICD-10: International Statistical Classification of Diseases and Related Health Problems-10th revision
MCVS: Multi-Threaded Clinical Vocabulary Server
NHGP: National Healthcare Group Polyclinics
NLP: natural language processing
UMLS: United Medical Language System

Edited by G Eysenbach; submitted 14.06.17; peer-reviewed by P Elkin, C Fu; comments to author 22.11.17; revised version received 14.02.18; accepted 19.03.18; published 11.06.18.

Please cite as:
Hardjojo A, Gunachandran A, Pang L, Abdullah MRB, Wah W, Chong JWC, Goh EH, Teo SH, Lim G, Lee ML, Hsu W, Lee V, Chen MIC, Wong F, Phang JSK
Validation of a Natural Language Processing Algorithm for Detecting Infectious Disease Symptoms in Primary Care Electronic Medical Records in Singapore
JMIR Med Inform 2018;6(2):e36
URL: http://medinform.jmir.org/2018/2/e36/
doi:10.2196/medinform.8204
PMID:29907560

Original Paper

Adverse Drug Event Reporting From Clinical Care: Mixed-Methods Analysis for a Minimum Required Dataset

David Peddie[1,2*], BEng, MA; Serena S Small[1,2*], MA; Katherin Badke[3*], BSc (Pharm), ACPR, PharmD; Chantelle Bailey[1,4*], MSc, PhD; Ellen Balka[1,2*], PhD; Corinne M Hohl[1,4,5*], MHSc, MD, FRCP(C)

[1]Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Research Institute, Vancouver, BC, Canada

[2]School of Communication, Simon Fraser University, Burnaby, BC, Canada

[3]Vancouver General Hospital, Department of Pharmaceutical Sciences, Vancouver, BC, Canada

[4]Department of Emergency Medicine, University of British Columbia, Vancouver, BC, Canada

[5]Vancouver General Hospital, Emergency Department, Vancouver, BC, Canada

[*]all authors contributed equally

Corresponding Author:
Serena S Small, MA
Centre for Clinical Epidemiology and Evaluation
Vancouver Coastal Health Research Institute
Research Pavilion, 7th Floor
828 West 10th Avenue
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 604 875 4111 ext 55219
Fax: 1 604 875 5179
Email: Small@sfu.ca

Abstract

Background: Patients commonly transition between health care settings, requiring care providers to transfer medication utilization information. Yet, information sharing about adverse drug events (ADEs) remains nonstandardized.

Objective: The objective of our study was to describe a minimum required dataset for clinicians to document and communicate ADEs to support clinical decision making and improve patient safety.

Methods: We used mixed-methods analysis to design a minimum required dataset for ADE documentation and communication. First, we completed a systematic review of the existing ADE reporting systems. After synthesizing reporting concepts and data fields, we conducted fieldwork to inform the design of a preliminary reporting form. We presented this information to clinician end-user groups to establish a recommended dataset. Finally, we pilot-tested and refined the dataset in a paper-based format.

Results: We evaluated a total of 1782 unique data fields identified in our systematic review that describe the reporter, patient, ADE, and suspect and concomitant drugs. Of these, clinicians requested that 26 data fields be integrated into the dataset. Avoiding the need to report information already available electronically, reliance on prospective rather than retrospective causality assessments, and omitting fields deemed irrelevant to clinical care were key considerations.

Conclusions: By attending to the information needs of clinicians, we developed a standardized dataset for adverse drug event reporting. This dataset can be used to support communication between care providers and integrated into electronic systems to improve patient safety. If anonymized, these standardized data may be used for enhanced pharmacovigilance and research activities.

(JMIR Med Inform 2018;6(2):e10248) doi:10.2196/10248

KEYWORDS

adverse drug event; adverse drug reaction; data fields; dataset; reporting; pharmacovigilance; mixed-methods; clinician-informed design

Introduction

Patients commonly transition between health care locations and care providers. Yet, electronic medical records containing critical information about a patient's medical care are usually confined to one health care sector within a geographic location (eg, a hospital) or to a group of care providers who share a common office or the same specialty (eg, a family physician group practice) [1]. National and international health care accreditation bodies and patient safety organizations have emphasized the importance of transferring accurate medication histories to avoid unintentional errors and patient harm when transitions between care locations or care providers occur [2,3]. These are reflected in the established standards and goals for obtaining and documenting the best possible medication histories [2,4-7].

Despite significant progress in this area, information sharing about adverse drug events (ADEs) remains inadequate, even though these are the leading cause of emergency department visits and hospitalizations [8-10]. Emerging evidence suggests that inadequate information sharing about ADEs across health care sectors and between care providers may lead to unintentional re-exposures of patients to medications that previously caused harm [11]. In a recent large cohort study of elderly patients in Ontario, Canada, 55% of patients hospitalized for a fall-related injury while on high-risk medications were re-exposed to the same medication—a benzodiazepine or a neuroleptic—within 180 days, most within only 90 days [12]. This study also found that 38% of the elderly who were admitted for hypoglycemia while on glyburide were restarted on the same medication, despite the medication-associated risk of hypoglycemia in this age group and the availability of safer treatment options [12]. These data indicate that a gap exists in information continuity about medication safety risks that has the potential to cause harm when patients transition between care locations and providers [13].

Electronic systems could enable more accurate and complete documentation of medication safety risks and play a pivotal role in electronically communicating this information to other care providers to close this gap, but they are presently underutilized for this purpose [14]. Many existing electronic medical record systems include data fields for allergy documentation. However, the structure of allergy data collection modules is inappropriate for the documentation of many common ADEs (eg, drug-disease state interactions). Even when broader fields are available to document ADEs, these are most commonly in free text format and, therefore, unstructured and nonstandardized, making them prone to misinterpretation when read by other care providers. Electronic ADE reporting systems in use by pharmacovigilance organizations (eg, Health Canada's MedEffect program) contain more structured and standardized data entry fields, but are burdensome to clinicians as they are external to electronic medical record systems and request clinically irrelevant information (eg, lot numbers of vaccines). ADE reporting within these systems is designed solely for drug regulatory purposes and is disconnected from clinical care activities, such that many clinicians do not access the systems at all [15-17].

Our objective was to develop a set of standardized data fields that clinicians could use to document and share information about ADEs from the point-of-care to address the information needs of clinicians and the limitations of existing systems. A secondary objective was to understand how electronic ADE documentation could be integrated into clinical activities to minimize the burden of documentation while improving reporting consistency, accuracy, and quantity.

Methods

Design and Setting

This was a mixed-methods study completed in British Columbia (BC), Canada, between March 2014 and April 2016 using a phased approached. As we have previously published our research methodology [15] and the results of some individual phases of this work [16,18], in this manuscript, we have focused on the research results used to develop, refine, and prioritize a minimum required dataset for ADE reporting.

In the first phase, we completed a systematic review of the existing ADE reporting systems [16]. We used information derived from our systematic review to develop a preliminary dataset that we presented to clinicians in iterative workshops in order to understand which data fields should be integrated into the minimum required dataset, their priority for integration, as well as their reporting sequence [15]. In parallel, we completed qualitative fieldwork to inform our understanding of the clinical nature of ADEs, clinicians' workflow in diagnosing ADEs, and challenges related to their documentation [19-21]. This informed our design decisions and will be integral to the successful implementation of the set of data fields [21]. We then pilot-tested a paper-based ADE data collection form in two clinical settings and refined the final dataset [18].

The University of British Columbia Research Ethics Board reviewed and approved the study protocol. Workshop participants provided implied consent, and care providers observed during workplace observations and pilot testing provided verbal consent. Consolidated criteria for reporting qualitative research informed the reporting of study findings.

Systematic Review

We began our work by completing a qualitative systematic review to synthesize data fields from existing ADE reporting systems worldwide [15,16]. We worked with a professional librarian to complete a systematic electronic bibliographic reference database and electronic gray literature search to identify ADE reporting systems worldwide. We developed, piloted, and refined a standardized data collection form to extract data about the reporting concepts and data fields used in each reporting system.

After identifying ADE reporting concepts and data fields, we imported them into the visual thinking software Inspiration 9.2 (Multimedia Appendix 1). We represented each individual data field with a bubble. Three research assistants (CB, DP, and MW) removed duplicates for identical data fields (eg, labeled "suspect medication") and summarized their frequency by indicating the number of instances that the data field was used by all reporting systems. We then sorted the remaining bubbles

into categories representing broad reporting concepts. In the third step, we collapsed nearly identical data fields (eg, "suspect medication" and "suspected medication") and identified the relationships and hierarchies between reporting concepts and data fields. This allowed us to create a preliminary reporting form containing all data elements and concepts used in ADE reporting internationally.

Qualitative Fieldwork

In order to understand the limitations of and existing means of documenting ADEs in clinical practice, we completed qualitative observations of care providers. Trained research assistants (CB, DP, SSS, and MW) observed clinical pharmacists and physicians in emergency departments and wards in 4- to 8-hour shifts at various times of the day and days of the week. We recruited a convenience sample of participants via email, word of mouth, and personal connections of care providers on the research team. Research assistants observed the process of patient care, which included clinicians managing patients' medications and occasionally investigating, documenting, and treating ADEs. We sought to understand the real-world clinical experiences related to ADEs, recognizing that retrospective accounts of ADEs may gloss over important characteristics, challenges, and work activities. We aimed to produce nuanced accounts of clinicians' workflows—patterns in their activities and interactions with patients and other clinicians and artifacts in the care setting (eg, electronic medical records, paper charts, forms). We used the findings from our observations to inform design decisions, particularly in relation to the implementation of the set of data fields. Two research assistants (DP and SSS) coded the field notes from the observations using qualitative data analysis software (NVivo 11). Following initial inductive coding, the team met regularly to discuss emergent results and finalize a coding structure.

Workshops

To obtain feedback on our preliminary ADE data field set, we created a preliminary reporting form using Microsoft Visual Basic for Applications to resemble a screenshot from a computer (Figures 1 and 2). The preliminary form contained all of the reporting concepts identified in the systematic review. We developed data formats and value sets for different data fields by drawing on existing standards.

We presented this preliminary form to groups of clinicians in workshops scheduled during lunchtime rounds for clinicians practicing in hospital settings and scheduled in the evening for clinicians practicing outside of hospitals. We targeted groups of hospital- and community-based pharmacists, emergency department physicians, general practitioners, and hospitalists as these individuals commonly diagnose, treat, or follow up patients with acute ADEs. We recruited prospective participants using posters, email invitations, and in-person conversations with colleagues. The sessions were led or co-led by a practicing physician (CMH) or clinical pharmacist (KB) on the team and were attended by research assistants (DP and SSS) who took field notes during the sessions. We informed participants that our principal goal was to design a novel system to document and report ADEs and to obtain their feedback on our preliminary form. We presented ADE cases that we had observed in prior qualitative fieldwork to facilitate the discussion. We asked participants to identify information about the event that they felt was, or was not, required and how and where the information should be documented within the form. We asked participants to contemplate the required information needs from the perspectives of someone needing to document the information as well as receiving the report in order to balance the need to minimize documentation while ensuring that the required information was available.

Figure 1. Sample screenshot from a preliminary adverse drug event (ADE) reporting form created using Microsoft Visual Basic for Applications.

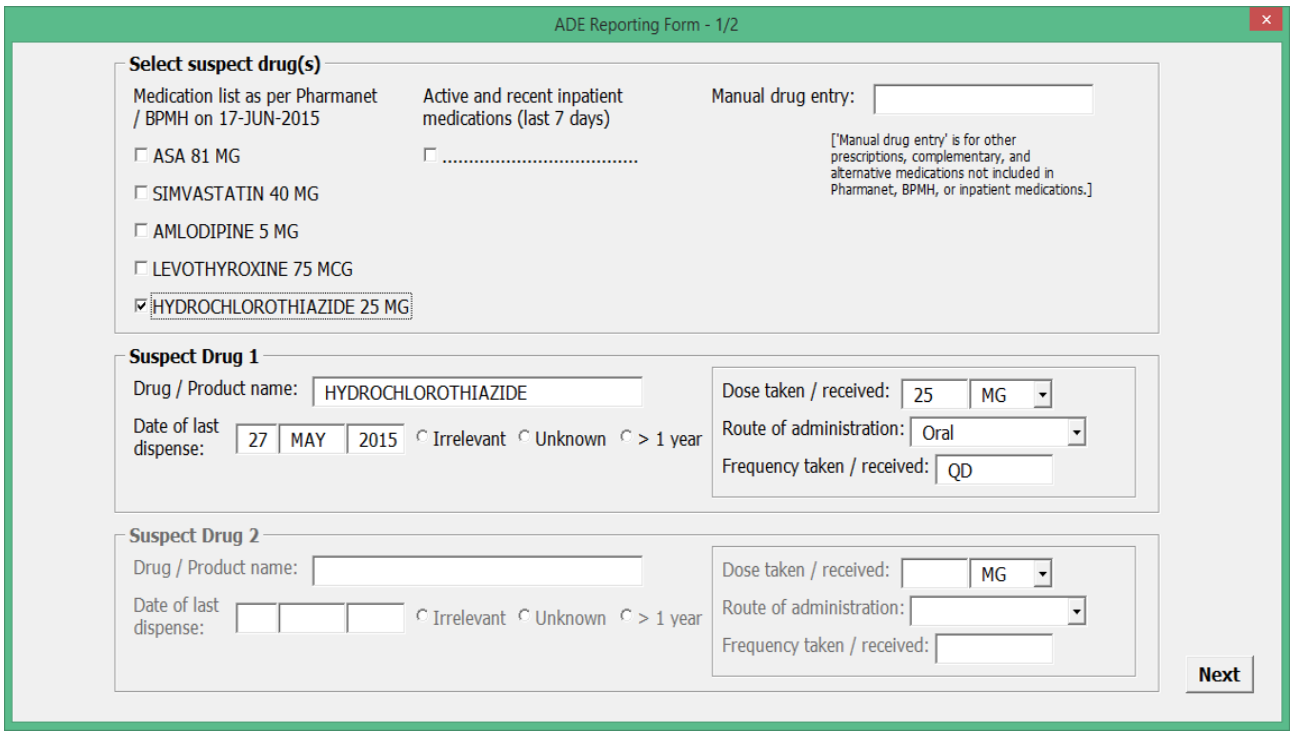
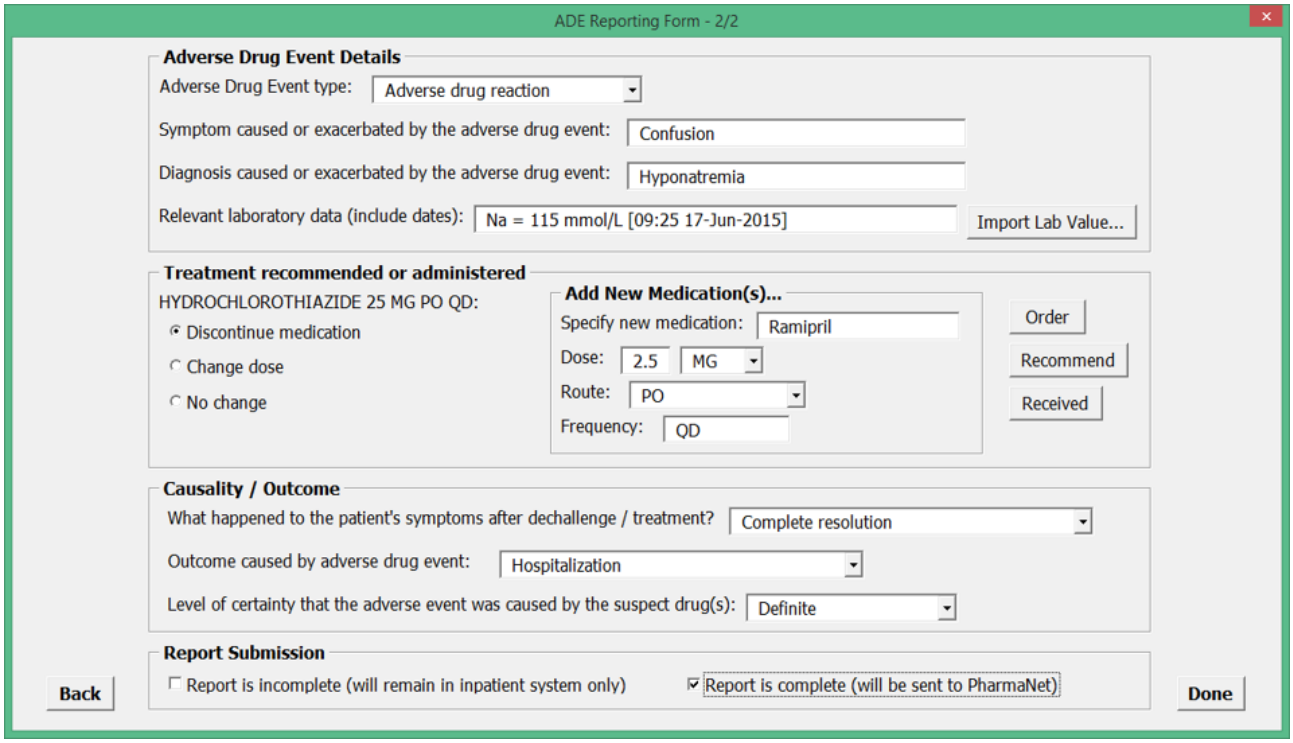


Figure 2. Sample screenshot from a preliminary adverse drug event (ADE) reporting form created using Microsoft Visual Basic for Applications.



Following each workshop, our team revised the data fields to incorporate feedback, producing a refined data field set for the next meeting or group. We maintained a log of changes that were made to the form, including the rationale for each change. Two research assistants (DP and SSS) coded the field notes from the workshops using the same approach as the qualitative observations. We considered the form to be a draft data field set when no novel suggestions or concerns were raised.

Pilot Testing

A research assistant (AC) piloted the form in two clinical settings to test it for content and functionality prior to its planned computerization [18]. During paper-based pilot testing, we sought to understand the electronic linkages that would be required in other systems to facilitate reporting by observing which information sources clinicians accessed when they completed the form. The trained research assistant observed clinicians using "lightweight ethnography" supplemented with semistructured interviews. Lightweight ethnography is a method that enables the collection of specific relevant information while acknowledging that a complete comprehension of the work setting is not possible or necessary [22]. We recruited a convenience sample of clinical pharmacists through email invitations and in-person conversations with colleagues as our prior work had demonstrated that pharmacists regarded ADE identification, documentation, and reporting to be central to their role, whereas physicians referred patients with ADEs to pharmacists for these tasks. The research assistant shadowed clinical pharmacists in 2- to 4-hour shifts. The research assistant provided the pharmacists a paper version of the ADE reporting form at the beginning of their shifts and asked them to complete it when they encountered an event. The research assistant collected data on the process of completing the form, as well as additional information about the ADE and the relevant workflow.

Results

We identified 108 active ADE reporting systems worldwide through our systematic review containing 1782 unique data fields [16]. We sorted the data fields into 33 reporting concepts that described the reporter, information about the patient, the ADE, and the suspect and concomitant drugs [16]. We completed 238 hours of observations of clinical pharmacists and 27 hours of observations of physicians in emergency departments, during which care providers encountered 65 possible ADEs [21]. We conducted 12 workshops with over 120 care providers: 6 with hospital pharmacists, 1 with community pharmacists, 2 with emergency department physicians, 1 with general practitioners, and 2 with hospitalists. We completed 25 hours of clinical observations during the pilot-testing phase, which included the documentation of 24 ADEs [18].

Table 1 summarizes the set of data fields that clinicians considered necessary when communicating information about ADEs, along with the formats, value sets, and data sources that they felt were most appropriate or expedient. Some value sets, including those for the data fields "Practitioner Role" and "Level of Certainty," were drawn from Fast Healthcare Interoperability Resources, which provides standards for data elements created by Health Level Seven International, a health care standards organization. Another source for the value sets was the Medical Dictionary for Regulatory Activities (MedDRA), an international dictionary for a medical terminology that has been clinically validated, applied using the Preferred Terms, which are descriptors for symptoms, diagnosis, and indication. MedDRA allows terms to be mapped to another internationally recognized standard, SNOMED CT, using the Unified Medical Language System metathesaurus.

Table 1. Data fields deemed relevant and necessary for adverse drug event reporting by clinicians.

Data Field	Format	Description or Value Set
Patient Information		
Date of birth	Alpha or Numeric	Autopopulate from EMR[a] or other electronic system, DD-MMM YYYY (eg, 01-FEB 1997)
Gender	Value set	Autopopulate from EMR or other electronic system (Male or Female or Other or Unknown) as per FHIR[b]
Name	Alpha	Autopopulate from EMR or other electronic system
Personal Health Number	Numeric	Autopopulate from EMR or other electronic system
Reporter Information		
Name	Alpha	Autopopulate from EMR login, or entered by clinician
Practitioner role	Value set	Autopopulate from EMR login, or entered by clinician (Doctor or Nurse or Pharmacist) as per FHIR
Hospital name and department	Alpha	Autopopulate from EMR, abbreviated
ADE[c] Information		
Date of report	Alpha or Numeric	Autopopulate from EMR, DD-MMM YYYY (eg, 01-FEB 1997)
ADE type	Value set	(Adverse drug reaction, Allergy, Incorrect drug, Subtherapeutic dose, Supratherapeutic dose, Treatment failure, Drug withdrawal, Drug interaction, Nonadherence, Other) derived from results of 4 prior prospective studies [8,23-25]
Symptom caused or exacerbated by ADE	Value set	Predictive entry from MedDRA[d] Preferred Terms
Diagnosis caused or exacerbated by ADE	Value set	Predictive entry from MedDRA Preferred Terms
Relevant tests or lab data (include dates)	Free text	Option for clinician to import from EMR (ideal) or enter manually
Outcome caused by ADE	Value set	(Death, Permanent disability, Exacerbated pre-existing condition, Congenital anomaly, Hospitalization, Emergency Department visit, Other, Unknown) derived from Health Canada Adverse Drug Reaction reporting standards, amended to reflect qualitative results)—not mutually exclusive
What happened after dechallenge or treatment?	Value set	(Resolved, Recovering, Ongoing, Resolved with Sequelae, Fatal, Unknown) as per FHIR
Level of certainty that the adverse event was caused by the suspect drug(s)	Value set	(Certain, Probably or Likely, Possible, Unlikely, Conditional or Classified, Unassessable or Unclassifiable, Refute) as per FHIR
ADE Treatment Information		
Suspect drug actions	Value set	(Discontinue, Change dose, No change)
Add new medication		Multiple fields (suspect drug or product name, dose, route, frequency, other information)—see Health Product data fields below
Treatment Status	Value set	(Ordered, Recommended, Received)
Health Product		
Suspect drug or product name(s)	Value set	Option to select from patient's medication list in EMR (ideal) or predictive entry from Canadian Clinical Drug Dataset combined with the Licensed National Health Products Database. Drugs included in the provincial formulary prioritized in search results. Drugs must also be searchable using a DIN[e] or NPN[f]. Multiple products may be entered as suspect drugs for the same event.
Dose taken or received	Alpha or Numeric or Special	Manual entry
Dose unit	Value set	(g, mg, mcg, IU[g], Units)
Route of administration	Value set	(Oral, SC[h], IM[i], IV[j], Topical)
Frequency taken or received	Alpha or Numeric or Special	Manual entry

Data Field	Format	Description or Value Set
Indication for drug	Value set	Prescription indication for use subset developed by Canada Health Infoway [26]
Other dosing information	Free text	Manual entry
Other		
Additional information (important details or context, timelines, follow-up)	Free text	For clinicians to provide additional information about any of the above, specify follow-up.

[a]EMR: electronic medical record.

[b]FHIR: Fast Healthcare Interoperability Resources.

[c]ADE: adverse drug event.

[d]MedDRA: Medical Dictionary for Regulatory Activities.

[e]DIN: Drug Identification Number.

[f]NPN: Natural Product Number.

[g]IU: International Unit.

[h]SC: subcutaneous.

[i]IM: intramuscular.

[j]IV: intravenous.

Clinicians discussed the tradeoffs of various data formats, noting that structured documentation eliminated confusing shorthand and led to more succinct, comprehensible reports and analyzable data. However, they also noted that they were unwilling to use forms where the data formats forced them to enter inaccurate or incomplete information. Clinicians expressed frustration with value sets that were incomplete or where only one option could be selected when several were relevant (eg, if they had to choose a single symptom or ADE type that did not accurately reflect their patient's situation or the use of an allergy field for documenting a drug-disease state interaction). Many health outcomes are not mutually exclusive. Therefore, clinicians are able to select more than one health outcome from the list (see "Outcome caused by ADE" field in Table 1). Clinicians noted that many events were not straightforward and required free text to provide important details, context, and follow-up information. Thus, despite the recognition that the use of free text fields can lead to the use of nonstandardized terminology, clinicians felt that a general free text field to enter additional information was needed.

Clinicians highlighted the importance of knowing whether an ADE was treated, and if so, how (see "ADE Treatment Information" fields in Table 1). They were particularly interested in the previous provider's actions related to the culprit medication: Was it discontinued? Was the dosage changed? Was it replaced? For clinicians, this information was crucial to determining the patient's medication regimen going forward and avoiding dangerous re-exposure while seeking alternative treatment(s) for the culprit drug's indication. One design option that was advanced was to link treatment data fields to the physician order page in the electronic medical record to allow physicians to document the event and initiate treatment using the same process.

Clinicians pointed out that a chief concern surrounding ADE documentation was that causality is often uncertain. They needed to be able to indicate their level of certainty regarding the causality of an ADE (see "Level of certainty" field in Table 1). They suggested that it would likely be more accurate for the clinicians to record the certainty or uncertainty of their causality assessment when entering data compared with that completed retrospectively by a data analyst reviewing the report who would lack the immediate knowledge of the patient's condition and circumstances of the event, as is commonly done in many pharmacosurveillance organizations. Our observations demonstrated that the limited certainty of patients' medical and medication histories led clinicians to manage patients based on a working, rather than definitive, diagnosis and that ADEs were diagnosed over time and across care settings. Thus, the report, including the level of certainty, is to be a living document with the capacity to edit, update, and refute information by multiple clinicians as information becomes available or as a patient's condition evolves. We propose that the definitions of this category should be readily available within any electronic system that uses this category to ensure a consistent use.

Clinicians pointed out that a chief concern surrounding ADE documentation was that causality is often uncertain. They needed to be able to indicate their level of certainty regarding the causality of an ADE (see "Level of certainty" field in Table 1). They suggested that it would likely be more accurate for the clinicians to record the certainty or uncertainty of their causality assessment when entering data compared with that completed retrospectively by a data analyst reviewing the report who would lack the immediate knowledge of the patient's condition and circumstances of the event, as is commonly done in many pharmacosurveillance organizations. Our observations demonstrated that the limited certainty of patients' medical and medication histories led clinicians to manage patients based on a working, rather than definitive, diagnosis and that ADEs were diagnosed over time and across care settings. Thus, the report, including the level of certainty, is to be a living document with the capacity to edit, update, and refute information by multiple clinicians as information becomes available or as a patient's condition evolves. We propose that the definitions of this category should be readily available within any electronic system that uses this category to ensure a consistent use.

Table 2. Data fields from the existing adverse drug event reporting forms that clinicians felt should be omitted.

Data field	Justification for excluding
Patient Information	
Height or weight	Future providers can obtain this information from patients or their records. Height and weight may be relevant to dosing, but are not essential for assessing most ADEs[a] and patients' medication regimen.
Medical history or concomitant disease states	Burdensome to enter, especially for complex patients. Future providers can often obtain this information from patients or their records.
Reporter Information	
Phone or mailing address or email	Burdensome to enter. If future providers have the reporter's name, role, and institution, they will likely be able to find the contact information online.
ADE Information	
Reaction start or end date, duration	Can be difficult to pinpoint (eg, delirium). Free text description of timelines is more accurate and in line with clinician charting practices.
Severity or seriousness	Even with standardized definitions, severity or seriousness assessments are often subjective, differ across contexts, and are prone to error. This information may be better communicated via other fields such as the patient's outcome (eg, was the patient hospitalized?), their treatment (did the ADE require treatment? Was the drug discontinued?), symptom or diagnosis (eg, anaphylactic reaction or upset stomach), lab data (eg, low sodium of 115 or 125), and dechallenge results (resolved or worsening).
Rechallenge information	Often unavailable at the point-of-care, or impractical, unethical, or harmful to re-expose the patient intentionally.
Health Product(s)	
Prescribed dose or frequency	The dose prescribed is less relevant than the dose that the patient actually took or received in relation to the ADE. Prescription information can be accessed elsewhere if needed.
Product strength	The dosage taken by the patient is more important. Given the product name or DIN[b] or NPN[c], product strength can usually be obtained.
Source (eg, pharmacy, grocery store, internet, other)	Generally not essential for assessing the ADE and the patient's medication regimen.
Product start or end date, duration	Can be difficult to accurately collect (must rely on patient memory or prescription records that may be unavailable or inaccurate). Free text description of timelines is more accurate and in line with clinician charting practices.
Manufacturer	Not essential for assessing the ADE and the patient's medication regimen.
Batch or lot #	Burdensome to gather; will often require tracing to pharmacy. Very rarely essential for assessing the ADE and the patient's medication regimen.
Concomitant health products	Providers should be able to enter multiple suspect drugs, but a complete account of the patient's medication regimen is burdensome to enter, especially for complex patients. Future providers can usually obtain other medication information from the patient, their records, or by linkage to electronic medication dispensing information depending on the jurisdiction where care is provided.

[a]ADE: adverse drug event.

[b]DIN: Drug Identification Number.

[c]NPN: Natural Product Number.

Table 2 provides an overview of some of the fields that were regularly included in ADE documentation as well as reporting forms encountered by us in our systematic review that clinicians felt could be excluded from our recommended data field set. Many of the fields in Table 2 exist for pharmacosurveillance purposes, including retrospective causality assessments. Clinicians rejected many of these fields, in part because they were skeptical about the accuracy of such retrospective assessments compared with the immediate assessment of the treating clinician. For the purposes of information sharing about ADEs, an indication of the treating clinician's certainty was seen as more important and required less data entry. Clinicians also rejected data fields such as the manufacturer, batch or lot number, and source, noting that these fields were infrequently available at the point-of-care and clinically irrelevant. Additionally, these fields exist to enable pharmacosurveillance agencies to detect deficiencies in manufacturer quality control that lead to patient harm, which contributed to none of the ADEs we observed. While clinicians noted that some of the fields in Table 2 might be relevant to specific ADEs, they felt that these fields would be less commonly used, would dissuade from reporting because their inclusion would render the form longer, and would have a lower utility for clinical care. They also noted that for cases where the excluded fields were relevant, the

reporter could supply this information in free text in a comment field.

Throughout our work, clinicians stressed that duplicate documentation of work was a problem with existing ADE reporting forms, which took time away from patient care activities. They argued in favor of a reporting form that was integrated into the local electronic medical record and could be prepopulated with reporter information (associated with their user account), patient information (associated with the patient's file), and possibly drug and dosing information (associated with the patient's medication history).

Discussions with clinicians emphasized striking a balance between too little and too much information. Clinicians felt that ADE documentation should be comprehensive enough to be clinically useful and not require future providers to seek out further information (eg, a documented allergy without enough information can complicate clinical decision making). At the same time, clinicians noted that in complex cases, they might be overwhelmed with the amount of information needed to keep track of a suspect ADE, until such time as a definitive ADE diagnosis could be made. Clinical utility, simplicity, convenience, and, to a lesser degree, signal generation were central considerations for clinicians when refining the set of data fields.

When observing clinical pharmacists pilot-tested the preliminary ADE reporting form containing the data field set developed by us, they felt that its length and level of detail were appropriate. They provided few important suggestions to abbreviate the dataset further. For example, they noted that the "Date of Last Dispense" field was irrelevant to clinical care and could be omitted and that "Follow-up Items" could be noted under "Additional Comments" [18]. Both of these fields were, therefore, removed.

Discussion

Principal Findings

Our objective was to describe a set of data fields for clinicians to document and communicate ADEs from the point-of-care to support clinical decision making and improve patient safety. We were able to take a large number of nonstandardized data fields currently in use by ADE reporting systems internationally and condense them to one standardized dataset, while mapping some required fields to internationally recognized standards. In this process, we had to make exclusions and tradeoffs. While not all participating clinicians agreed on every field, our iterative process engaged different types of end users and was far more inclusive than is customary in information technology design in health care.

We recognize that the omission of regulatory fields may be controversial. We have taken this approach from the perspective that clinical tools need to be designed foremost to enhance the immediate delivery of care. Incomplete and nonstandardized information sharing about ADEs across health care sectors and between care providers puts patients at risk [11,12]. However, there are other important reasons why the exclusions of regulatory fields may be justified and beneficial.

First, given the complexity of the ADEs observed by our team, the immediate clinician's assessment is likely more reliable than a retrospective, at-a-distance assessment. In addition, clinicians preferred to provide causality data from the point-of-care as this assessment was felt to be crucial for informing subsequent clinical decisions.

Second, it was clear that clinicians regarded data fields used to support retrospective information gathering for regulatory agencies as burdensome. To obtain information related to manufacturer quality control issues, such as batch and lot number, clinicians often must attempt to trace the drug back to the pharmacy, a time-consuming activity that is irrelevant for most ADEs. Similarly, fields gathering information already contained in the electronic format prior to the ADE assessment, such as concomitant therapies or product start and end dates, require clinicians to duplicate the entry of information that exists elsewhere. If regulatory assessments require this data, it may be more effective to establish links to complementary datasets (such as prescribing information in a jurisdictional drug information system) than to request that clinicians provide it. We may, simply by easing documentation burden, see an increase in ADE reporting, which would contribute to improved data compared with conventional systems that most clinicians reported never having accessed.

As health systems internationally struggle to motivate providers to report ADEs and new electronic infrastructures are established to improve health information sharing across settings through e-prescribing or drug information systems, our results are timely. New electronic systems offer the potential to streamline information gathering and data entry and consolidate the multiple forms, platforms, information sources, and medication ordering tools that are necessary for clinical work. However, in practice, these systems have the potential to increase documentation burden on clinicians, cause unexpected errors, and desensitize clinicians to alerts due to overflagging and alert fatigue [1,21,27-30]. We see an opportunity to capitalize on new technology by integrating ADE documentation into the systems that clinicians already use, incorporating reporting into clinical workflow, and avoiding duplicate data entry. Clinicians who prescribe and dispense medications expressed interest in using patient-specific ADE data to create patient-specific, medication-level alerts to help them avoid unintentionally re-exposing a patient to the same drug that previously caused harm.

While the selection of standardized data fields alone cannot guarantee the generation of high-quality ADE reports, a clinician-informed design is more likely to result in relevant data. Implementation strategies for this dataset should continue to seek input from clinicians to facilitate uptake and adoption and ensure end-user engagement and adaptation to local contexts. If implemented with attention to clinical workflow, standardized and clinically relevant data fields may yield more accurate and complete information that can inform clinical care and improve patient safety while providing higher-quality representative data for surveillance and research activities.

At the time of publication, our team has programmed this set of data fields into an electronic app, called *ActionADE*, which

is being pilot implemented on iPads in a teaching hospital in Vancouver, BC. Plans are underway for its integration with the provincial drug information system so that standardized ADE data can be communicated between providers and across health settings. The piloting and implementation phases of *ActionADE* will follow similar methodological rigor as undertaken in the development phase of the data fields. Throughout our work, our team has maintained contact with key national organizations such as Health Canada, Canada Health Infoway, the Institute for Safe Medication Practices, and Accreditation Canada in an effort to increase the likelihood that the data fields identified here will be adopted nationally. If successfully adopted and implemented, researchers and drug regulators may benefit from the data that would be generated as a by-product of safer care.

Conclusions

Existing electronic systems allow clinicians to document ADEs, but are nonstandardized and provide limited information that can be shared across health settings and between providers. The structured and standardized set of data fields presented by us are intended to meet the needs of frontline clinicians while enabling a standardized, unambiguous communication between care providers and electronic systems to increase care quality and safety. If implemented, the minimum required data fields have the potential to address the informational discontinuity and reduce ADEs while improving the available health data for pharmacosurveillance and research purposes as a by-product of safer care.

Acknowledgments

This study would not have been possible without the support of many frontline care providers who allowed us to observe their work and who provided us with insightful comments to enrich our understanding of the issues under study. This research was sponsored by the Canadian Institutes of Health Research, Partnership for Health System Improvement Grant (No. 293546), the Michael Smith Foundation for Health Research (No. PJ HSP 00002), Vancouver Coastal Health, the BC Patient Safety & Quality Council, and Health Canada. During the time of this study, CMH was funded through a CIHR New Investigator grant (No. 201109ND1-261895-157349).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Excerpt of adverse drug event (ADE) concepts and data fields found during our systematic review using Inspiration 9.2.

[PNG File, 181KB - medinform_v6i2e10248_app1.png]

References

1. Balka E, Whitehouse S, Coates ST, Andrusiek D. Ski hill injuries and ghost charts: Socio-technical issues in achieving e-Health interoperability across jurisdictions. Inf Syst Front 2011 Apr 20;14(1):19-42. [doi: 10.1007/s10796-011-9302-4]
2. Accreditation Canada. Required Organizational Practices Handbook 2017. 2016. URL: http://www.gov.pe.ca/photos/original/src_rophandbook.pdf [accessed 2018-01-05] [WebCite Cache ID 6wOzNu50F]
3. The Joint Commission. Hot Topics in Healthcare. 2012 Jun. Transitions of Care: The need for a more effective approach to continuing patient care URL: https://www.jointcommission.org/assets/1/18/Hot_Topics_Transitions_of_Care.pdf [accessed 2018-01-07] [WebCite Cache ID 6wOwVPOkb]
4. The Joint Commission. 2016 Hospital National Patient Safety Goals. 2016. URL: https://www.jointcommission.org/assets/1/6/2016_NPSG_HAP_ER.pdf [accessed 2018-01-08] [WebCite Cache ID 6wP0GTGR0]
5. National Institute for Health and Care Excellence. Medicines optimisation. 2016 Mar 24. Quality Standard (QS120) URL: https://www.nice.org.uk/guidance/qs120/resources/medicines-optimisation-pdf-75545351857861 [accessed 2018-01-09] [WebCite Cache ID 6wP15k50O]
6. World Health Organization. The High 5s Project: Standard Operating Protocol. Assuring medication accuracy at transitions in care: Medication reconciliation. Geneva: World Health Organization; 2014 Sep. URL: http://www.who.int/patientsafety/implementation/solutions/high5s/h5s-sop.pdf [accessed 2018-01-10] [WebCite Cache ID 6wP1JFo7r]
7. World Health Organization. Transitions of care: Technical series on safer primary care. Geneva: World Health Organization; 2016. URL: http://apps.who.int/iris/bitstream/10665/252272/1/9789241511599-eng.pdf [accessed 2018-01-11] [WebCite Cache ID 6wP1RtT4M]
8. Zed PJ, Abu-Laban RB, Balen RM, Loewen PS, Hohl CM, Brubacher JR, et al. Incidence, severity and preventability of medication-related visits to the emergency department: a prospective study. CMAJ 2008 Jun 3;178(12):1563-1569 [FREE Full text] [doi: 10.1503/cmaj.071594] [Medline: 18519904]
9. Budnitz DS, Lovegrove MC, Shehab N, Richards CL. Emergency hospitalizations for adverse drug events in older Americans. N Engl J Med 2011 Nov 24;365(21):2002-2012. [doi: 10.1056/NEJMsa1103053] [Medline: 22111719]

10. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998 Apr 15;279(15):1200-1205. [Medline: 9555760]
11. van der Linden CMJ, Kerskes MCH, Bijl AMH, Maas HAAM, Egberts ACG, Jansen PAF. Represcription after adverse drug reaction in the elderly: a descriptive study. Arch Intern Med 2006;166(15):1666-1667. [doi: 10.1001/archinte.166.15.1666] [Medline: 16908802]
12. Winick-Ng J, Richards L, Clemens K, Shariff S, Welk B. ImagineNation Challenges. 2016. Data Impact Challenge II Answer Submission URL: http://imaginenationchallenge.ca/wp-content/uploads/2016/04/ICES-Western-Question-1.pdf [accessed 2017-12-18] [WebCite Cache ID 6voCRiF0y]
13. van der Linden CM, Jansen PA, Grouls RJ, van Marum RJ, Verberne MA, Aussems LM, et al. Systems that prevent unwanted represcription of drugs withdrawn because of adverse drug events: a systematic review. Ther Adv Drug Saf 2013 Apr;4(2):73-90 [FREE Full text] [doi: 10.1177/2042098613477125] [Medline: 25083253]
14. van der Linden CM, Jansen PA, van Marum RJ, Grouls RJ, Korsten EH, Egberts AC. Recurrence of adverse drug reactions following inappropriate re-prescription: better documentation, availability of information and monitoring are needed. Drug Saf 2010 Jul 1;33(7):535-538. [doi: 10.2165/11532350-000000000-00000] [Medline: 20553055]
15. Peddie D, Small SS, Badke K, Wickham ME, Bailey C, Chruscicki A, et al. Designing an adverse drug event reporting system to prevent unintentional reexposures to harmful drugs: study protocol for a multiple methods design. JMIR Res Protoc 2016 Aug 18;5(3):e169 [FREE Full text] [doi: 10.2196/resprot.5967] [Medline: 27538362]
16. Bailey C, Peddie D, Wickham ME, Badke K, Small SS, Doyle-Waters MM, et al. Adverse drug event reporting systems-a systematic review. Br J Clin Pharmacol 2016 Jul;82(1):17-29. [doi: 10.1111/bcp.12944] [Medline: 27016266]
17. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions : a systematic review. Drug Saf 2006;29(5):385-396. [Medline: 16689555]
18. Chruscicki A, Badke K, Peddie D, Small S, Balka E, Hohl CM. Pilot-testing an adverse drug event reporting form prior to its implementation in an electronic health record. Springerplus 2016;5(1):1764 [FREE Full text] [doi: 10.1186/s40064-016-3382-z] [Medline: 27795906]
19. Peddie D, Hohl C, Bailey C, Wickham M, Badke K, Small S, et al. Quality Forum 2018. Vancouver, BC: BC Patient Safety & Quality Council–Quality Forum; 2015. Challenges care providers face documenting adverse drug events: an observational study URL: https://qualityforum.ca/qf2015/wp-content/uploads/2015/02/David-Peddie-et-al.-Challenges-Care-Providers-Face-Documenting-Adverse-Drug-Events-An-Observational-Study.pdf [WebCite Cache ID 6voCRiF0y]
20. Peddie D, Small S, Wickham M, Bailey C, Hohl C, Balka E. Designing novel health ICTs to support work, not generate it: five principles. Stud Health Technol Inform 2017;234:262-268. [Medline: 28186052]
21. Hohl CM, Small SS, Peddie D, Badke K, Bailey C, Balka E. Why clinicians don't report adverse drug events: qualitative study. JMIR Public Health Surveill 2018 Feb 27;4(1):e21 [FREE Full text] [doi: 10.2196/publichealth.9282] [Medline: 29487041]
22. Randall D, Rouncefield M, Harper R. Fieldwork for design: theory and practice. London: Springer-Verlag; 2007.
23. Hohl CM, Badke K, Zhao A, Wickham ME, Woo SA, Sivilotti MLA, et al. Prospective validation of clinical criteria to identify emergency department patients at high risk for adverse drug events. Acad Emerg Med 2018 Mar 08. [doi: 10.1111/acem.13407] [Medline: 29517818]
24. Hohl CM, Yu E, Hunte GS, Brubacher JR, Hosseini F, Argent CP, et al. Clinical decision rules to improve the detection of adverse drug events in emergency department patients. Acad Emerg Med 2012 Jun;19(6):640-649 [FREE Full text] [doi: 10.1111/j.1553-2712.2012.01379.x] [Medline: 22687179]
25. Hohl CM, Partovi N, Ghement I, Wickham ME, McGrail K, Reddekopp LN, et al. Impact of early in-hospital medication review by clinical pharmacists on health services utilization. PLoS One 2017;12(2):e0170495 [FREE Full text] [doi: 10.1371/journal.pone.0170495] [Medline: 28192477]
26. Canada Health Infoway. Terminology Gateway: prescription indication for use subset. 2017. URL: https://tgateway.infoway-inforoute.ca/ [accessed 2018-01-12] [WebCite Cache ID 6wQDR5Jpa]
27. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004 Apr;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
28. Hartswood M, Procter R, Rouncefield M, Slack R. Making a case in medical work: implications for the electronic medical record. Computer Supported Cooperative Work (CSCW) 2003 Sep;12(3):241-266. [doi: 10.1023/A:1025055829026]
29. Toll E. A piece of my mind. The cost of technology. JAMA 2012 Jun 20;307(23):2497-2498. [doi: 10.1001/jama.2012.4946] [Medline: 22797449]
30. Berg M. Implementing information systems in health care organizations: myths and challenges. Int J Med Inform 2001 Dec;64(2-3):143-156. [Medline: 11734382]

Abbreviations

ADE: adverse drug event

BC: British Columbia
DIN: Drug Identification Number
EMR: electronic medical record
FHIR: Fast Healthcare Interoperability Resources
IM: intramuscular
IU: International Unit
IV: intravenous
MedDRA: Medical Dictionary for Regulatory Activities
NPN: Natural Product Number
SC: subcutaneous

Edited by G Eysenbach; submitted 28.02.18; peer-reviewed by J Klopotowska, KNB Nor Aripin; comments to author 29.03.18; revised version received 08.05.18; accepted 24.05.18; published 28.06.18.

Please cite as:
Peddie D, Small SS, Badke K, Bailey C, Balka E, Hohl CM
Adverse Drug Event Reporting From Clinical Care: Mixed-Methods Analysis for a Minimum Required Dataset
JMIR Med Inform 2018;6(2):e10248
URL: http://medinform.jmir.org/2018/2/e10248/
doi:10.2196/10248
PMID:29954724

Original Paper

The Importance of Nonlinear Transformations Use in Medical Data Analysis

Netta Shachar[1*], MSc; Alexis Mitelpunkt[1,2,3*], MD; Tal Kozlovski[1], MSc; Tal Galili[1], PhD; Tzviel Frostig[1], MSc; Barak Brill[1], MSc; Mira Marcus-Kalish[4], PhD; Yoav Benjamini[1,4], PhD

[1]Department of Statistics and and Operations Research, Tel Aviv University, Tel Aviv, Israel

[2]Pediatric Neurology, Dana-Dwek Children's Hospital, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

[3]School of Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[4]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

[*]these authors contributed equally

Corresponding Author:
Alexis Mitelpunkt, MD
Department of Statistics and and Operations Research
Tel Aviv University
Ramat Aviv
Tel Aviv, 6997801
Israel
Phone: 972 3 6974115
Email: alexism@post.tau.ac.il

Abstract

Background: The accumulation of data and its accessibility through easier-to-use platforms will allow data scientists and practitioners who are less sophisticated data analysts to get answers by using big data for many purposes in multiple ways. Data scientists working with medical data are aware of the importance of preprocessing, yet in many cases, the potential benefits of using nonlinear transformations is overlooked.

Objective: Our aim is to present a semi-automated approach of symmetry-aiming transformations tailored for medical data analysis and its advantages.

Methods: We describe 10 commonly encountered data types used in the medical field and the relevant transformations for each data type. Data from the Alzheimer's Disease Neuroimaging Initiative study, Parkinson's disease hospital cohort, and disease-simulating data were used to demonstrate the approach and its benefits.

Results: Symmetry-targeted monotone transformations were applied, and the advantages gained in variance, stability, linearity, and clustering are demonstrated. An open source application implementing the described methods was developed. Both linearity of relationships and increase of stability of variability improved after applying proper nonlinear transformation. Clustering simulated nonsymmetric data gave low agreement to the generating clusters (Rand value=0.681), while capturing the original structure after applying nonlinear transformation to symmetry (Rand value=0.986).

Conclusions: This work presents the use of nonlinear transformations for medical data and the importance of their semi-automated choice. Using the described approach, the data analyst increases the ability to create simpler, more robust and translational models, thereby facilitating the interpretation and implementation of the analysis by medical practitioners. Applying nonlinear transformations as part of the preprocessing is essential to the quality and interpretability of results.

(JMIR Med Inform 2018;6(2):e27) doi:10.2196/medinform.7992

KEYWORDS

data mining; statistics; preprocessing; medical informatics; health informatics; big data; transformations

Introduction

Medical Data Analysis

The volume of data collected these days is constantly growing and is expected to reach 44 zettabytes by 2020 [1], and medical data are rapidly catching on to this trend. Informed use of data collected from the entire population, in a way that can lead to providing better treatment to each patient, is a challenging goal. Unlike the traditional way of collecting data for a specific purpose, big data and its relevant subsets are analyzed for multiple purposes, in multiple ways by means of statistical models, data mining algorithms, machine learning methods, and others. Moreover, the accessibility of big data through easier-to-use platforms, such as the open source KNIME Analytics Platform or older commercial software such as SPSS Modeler, will allow practitioners who are not expert data analysts to get answers by analyzing big data. Systems designed to specifically cater to a wider range of less-sophisticated data analysts in the field of health informatics are becoming more common. For example, the Medical Informatics Platform of the European Human Brain Project is an innovative data analysis system that "provides an interface through which clinicians, neuroscientists, epidemiologists, researchers but also health managers and even the general public can access and analyze imaging and clinical data currently locked in hospital and research archives and public databases." We demonstrate that transformation-based preprocessing is an enhancer tool important for even simple questions using simple tools. Incorporating transformations into the preprocessing stage is essential in order to enable less-sophisticated data analysts to obtain informative and relevant results by simple means.

With the ever-increasing numbers of variables and subjects, there is a notion in the general public that simply the amount of data will reveal all there is to understand from it. This is not necessarily so. Data analysis can be greatly simplified when (1) the distribution of a variable (feature) is symmetric across subjects, (2) variability is stable across different conditions, (3) relationships are linear between variables of interest, and (4) models are additive rather than having a more complex structure. The need for these desirable properties can be circumvented by very complex modeling, which is feasible in data-rich research, as discussed below. However, such complex modeling leads to results that are harder to interpret and less appropriate for generalizing, and the ability to extrapolate beyond the given data or to set thresholds with diagnostic values is reduced. In addition, complex modeling requires a high level of expertise in data analysis—a level most physicians and health specialists cannot devote enough time to achieve, even if they are interested in acquiring quantitative answers to their questions.

In practice, data as extracted and measured are not usually characterized by the above four desirable properties. Moreover, the widely used essentially linear transformations such as normalization or Z score are practically linear transformations and therefore cannot change the symmetry or linearity. It is well known that income information is better analyzed after applying a logarithmic transformation, and the same is true for concentrations of substances in body fluids or the use of log odds for risk modeling. It is similarly recognized that gene expression data should be log-transformed and its variability stabilized before it can be compared across subjects and conditions. This remains true no matter what role it plays in the modeling effort: whether it is the variable of interest to be analyzed or used for prediction and explanation of another quantity of interest. The principles that the data need not remain in their original scale of measurement and that monotone transformations of variables can be used to facilitate the interpretation and generalization are well accepted in specific research situations. It is a cornerstone of the Exploratory Data Analysis approach to analyzing small and medium datasets [2].

The above principle is undervalued and rarely used when addressing bigger and more complex datasets. A search in the Web of Science database for the terms "classification OR prediction OR clustering" published in one of the three leading medical informatics journals (*Journal of the American Medical Informatics Association*, *International Journal of Medical Informatics*, *BMC Medical Informatics and Decision Making*) between January 2013 and April 2015 yielded 226 articles. This list was further screened manually, and 43 articles that specifically dealt with clinical data analysis were selected. Of these, 16 used data mining and machine-learning algorithms for classification of clinical data. Only one article mentioned performing a preprocessing stage, which consisted only of correcting misalignments, missing data, and the selection of the most predictive variables (see Multimedia Appendix 1). This does not necessarily imply that all should have, but rather reflects the common practice.

The Importance of Symmetry-Improving Nonlinear Transformations in Preprocessing

We demonstrate that transforming a variable so that its distribution across individuals is approximately symmetric goes a long way towards achieving the other goals for which transformations are useful: stabilizing the variance across conditions, assuring linearity in the relationship between variables, and the additivity of the response of interest [2,3]. The symmetric distribution need not be Gaussian, but symmetry does imply that summaries of such transformed distribution will be close to Gaussian, allowing the call of outliers and the setting of confidence statements of regression coefficients and other inferences to be more justified. Moreover, symmetry is highly desirable when computing distances on many features reflecting different underlying entities, where skewed distributions may hide important differences that are not in the tails.

Sometimes it is not feasible to transform to symmetry because the variable has a more complex structure, such as multimodality ("bumps" or "humps"). Even then, transforming in a way that causes the main body of data to be in a symmetric hump is advantageous.

Limiting the search for appropriate transformation at the preprocessing stage towards symmetry only becomes essential when confronted with big data. The detailed inspection of the relationship between every pair of variables, inspecting the residuals from every model under consideration, for lack of linearity or homogeneity of variability is impossible. Nor is it known ahead of time what the problem of interest to a future

user will be. In contrast, when searching for symmetry-inducing transformations at the preprocessing stage, the number of manual inspections is at most linear in the number of variables and thus becomes feasible.

We therefore propose a semi-automated practice for symmetry-targeting preprocessing that enables the data scientist to select an appropriate transformation for each variable in the dataset, thereby allowing the analysis of thousands of variables in an efficient and reproducible manner.

It is important to acknowledge that for every specific analysis, more complex procedures may be used to overcome the limitations discussed above, including nonlinear methods based on splines or loess, robust methods that ignore outliers, and hierarchical methods that capture interactions [4-6]. These and other methods are comprehensively described in Hastie et al [7]. Even such high-powered users will enjoy a better starting point post transformation. The more casual ones will be saved from disaster.

Methods

Monotone Transformations

The monotone transformations we use belong to the power transformations family, where x is transformed to x^p. We recommend Tukey's ladder of re-expression, where squares, square root, their reciprocals, and their likes are used [2]. Formally, use x^p, where $p=m/n$, $n=1,2,3$ and $m=-2,-1,0,1,2$. In all formulations, $p=0$ means using the logarithmic transformation (the base is not important, so base 2, the natural base, and base 10 are usually used). We prefer this ladder, where a discrete set of powers is used, over allowing any p, as in the Box-Cox formulation [8]. In the latter, p is tailored to optimize a specific goal, while at the preprocessing stage, no single goal is clear at the outset. Tukey's ladder includes all transformations that are known to be particularly useful such as log, square root and the cubic roots, and their inverses. Note that when p is negative, the transformation monotonically decreases rather than increases. In many cases, the meaning remains intuitive. For example, when the time to pass 1 meter is measured, once transformed by $(\text{time})^{-1}$ it is merely the speed that is recorded. The longer the time, the lower the speed.

The above transformations are further combined with other transformations that are tailored to the type of variable encountered. Tukey's taxonomy of measured variables [2] as is applicable to medical variables is presented, as well as the adequate transformation to be used for each type in conjunction with the power family. This taxonomy is helpful when dealing with a dataset that encompasses a large and varied number of variables. Further explanations, the mathematical and statistical reasoning, and the transformation equations can be found in Multimedia Appendix 2.

Tukey's Taxonomy Applied to Medical Variables and Their Appropriate Transformations

Amounts are measurements taking any positive value $y\geq 0$ (eg, serum protein level). Some measurements are inherently bounded from below, that is, $y\geq C$, in which case y C is a valid amount. In order to arrive at a symmetric distribution for amounts, we can use any power transformation directly (see Multimedia Appendix 2 for handling amounts that include zeros).

Counts are counts of units and therefore can take only integer values 0,1,2,… (eg, number of adverse events, days under antibiotic treatment). Counts tend to be right-skewed and have variance increasing with their mean, stemming from the Poisson nature of their distribution. The square root is a variance stabilizing transformation for Poisson counts and often achieves symmetry.

Ratio is one amount x divided by another amount y, that is, $r=x/y$ (eg, protein/creatinine ratio in the urine). Ratios often require log transformations. This then takes the form of $\log(r)=\log(x)-\log(y)$. Other power transformations can replace log, in the form of x^p y^p, if the two elements are available.

Fraction is a ratio of one amount x to some other amount y, which is always larger, that is, $0\leq r=x/y\leq 1$ (eg, fraction of a blood vessel that is clotted). Usually deciding on the property measured by the ratio r is equivalent to deciding to measure its complement 1 r (instead of the fraction of a blood vessel that is clotted, one could define the fraction that is open). Log (r/1 r), also known as the "logit" transformation, is the most useful (the widely used log odds for comparing morbidity).

Counted fraction is the fraction of counts, counting how many out of how many have some property (n out of m). It is therefore bound between 0 and 1 (eg, the number of patients admitted on a day due to a specific symptom out of the total number of admitted patients that day). Here too, the most useful transformation is the logit transformation: $\log(n+\frac{1}{3}/m-n+\frac{1}{3})$. We add a constant $c=\frac{1}{3}$ to avoid dividing by zero [2].

Amounts that are inherently bounded $a\leq y\leq b$, are essentially fractions, given by $0\leq r=(y\ a)/(b\ a)\leq 1$ (eg, the length of time a child sleeps per day). We first write them as fractions and then transform them as discussed above.

Counts that are inherently bounded $l\leq n\leq m$, are essentially counted fractions, given by $0\leq r=n\ l)/(m\ l)\leq 1$ (eg, the number of correctly answered questions in a questionnaire), and as such, they should be handled in the same manner explained for counted fractions, $\log(n\ l+\frac{1}{3})/(m\ n+\frac{1}{3})$.

Difference in amounts, counts, or fractions may take positive or negative values (eg, the difference between the number of words forgotten in an immediate recall assignment and number of words forgotten in a delayed recall assignment). Differences are best handled by transforming the subtracted variables separately, be they amounts or counts. If the two variables that are differenced are not available, the difference variable should rarely be transformed. Instead, it can practically be only positive where it should be handled as amount.

Ranks are bounded counts of how many are below and equal the observation out of the total number ranked (eg, the rank of a specific symptom-describing word out of all descriptive words used). They will be treated exactly as bounded counts.

Ordinal variables are variables whose values are named categories that can be naturally ordered (eg, everyday cognition assessment taking the values 1-4). Each category can be assigned a numeric value (similar to ranks with ties) depending on the proportion of cases in the category and below it, compared to the background of reference distribution. An explicit formula for a logistic reference distribution is given in Multimedia Appendix 2.

The medical information dataset should, therefore, be accompanied by a metafile, specifying for each variable its type (according to the above 10-item list), its context-related lower and upper bounds (when these exist). Finally, an indicator of whether the variable should be reversed, in order to ease interpretation should be added, so that variables carrying similar meaning, say "healthy," are presented in the same direction.

A Semi-Automated Approach to Choosing Symmetry-Targeted Transformations

With the above information at hand, we recommend that the choice of transformations be performed in a semi-automated manner. The automated part can be guided by the measure of skewness that directed the analyst towards a few alternative transformations around the power p that was calculated. Yule's measure of skewness [9] (also known as Bowley's) usually serves this purpose well, with a value of 0 for symmetric data:

$$sk=0.5(m_3+m_1)- m_2/0.5*(m_3-m_1)$$

where m_1, m_2, and m_3 are the lower quartile (ie, the 25thpercent quantile), the median, and the upper quartile (ie, the 75thpercent quantile), respectively. The interquartile range $0.5*(m_3-m_1)$ that is a measure of the spread is equal to the median absolute deviation for exactly symmetrical distributions.

A less-resistant version of the Yule index can serve well, even when the data consist of bounded counts over a small range (see Multimedia Appendix 2).

The automatic search is combined with subjective assessment in the following way. For a given variable x_i, if $|sk_i|$ c, no transformation was needed, and the next variable is inspected. As a rule of thumb, a recommended setting is c=0.1 when the number of cases is bigger than 180, and larger when the number is smaller: 0.12 for 120 and 0.15 for 80. For all skewed variables, according to the type of variable as listed in the metafile, histograms are computed and displayed for the original values and the power transformation over Tukey's ladder that makes the variable most symmetric, as well as of closely related transformations. Unlike the Box-Cox transformations that are flexibly chosen, those on Tukey's ladder are all easily interpretable and more robust. Moreover, we use the chosen transformation as the starting point: the plots of the distribution following these potentially useful transformations are shown with the corresponding skewness measure. Finally, the researcher should choose which of the transformations (if any) best suits the specific need. The type of transformation performed on each of the variables is registered in order to enhance reproducibility.

Data

The above approach was implemented in R and used to preprocess data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimer's disease (AD) is the most common form of dementia. There is currently no known treatment nor one that slows the progression of this disorder [10]. Therefore, finding pathways for treatment is a major research effort. ADNI was conceived at the beginning of the millennium as a North American multicenter collaborative effort funded by public and private bodies, in order to facilitate a progression in the understanding, assessment, and treatment of AD. The initiative obtains data on patients of normal cognitive state, early and late mild cognitive impairment, significant memory concern, and AD. Clinical, neuropsychological, biological markers, imaging, and genetic data are collected on the patients [10]. The clinical measurements were obtained from multiple data tables, extracted from the ADNIMERGE R package (version 0.0.1 downloaded June 2014). A no-missing-values-subset of the ADNI data with 185 potentially explanatory variables and a target variable—the assigned diagnosis on 658 participants—was used for the demonstration.

A second dataset involved 1575 Parkinson's disease (PD) patients and their first degree relatives assembled in the Tel Aviv Medical Center: 1185 idiopathic PD patients, 164 PD patients who are carriers of the G2019S mutation in the *LRRK2* gene, and 226 PD patients who are carriers of mutations in the *GBA* gene. Patients were followed between 2008 and 2015. Each subject underwent a battery of medical exams and questionnaires ranging from demographic to physical, cognitive, and performance-based assessments. These included physical and neurological examinations as well as clinical assessment questionnaires, for example, the Unified Parkinson's Disease Rating Scale (UPDRS). Other symptoms were evaluated by the Beck Depression Inventory, Geriatric Depression Scale, Scopa-Aut, and more. Overall, 772 different measures were made available to us.

In addition to the clinical datasets described above, the benefits of transformations for cluster analysis were demonstrated using simulated, rather than real research data. First, most clustering algorithms require the unknown number of clusters as input; this number is known in the simulated data scenario. Second, assessing the accuracy of the clusters is not straightforward, as the real separation into subgroups is absent. When the data are generated according to a known clustered structure, the accuracy of the results can be easily measured using, for example, using the adjusted Rand index [11].

We constructed a simulated dataset of the clinical situation described as originally three assigned diseases: A, B, and C. The "real" situation is that disease A has two subtypes A1, A2; Diseases B and C are actually two subtypes of disease BC. The data contained 100 continuous, discrete, and binary features, of which 7 features define 6 disease subgroups (A1, A2, B, C, BC, and Normal). Each feature was created symmetrically and then transformed in order to exhibit a skewed distribution. The clustering analysis was performed on both the original and the transformed data.

In order to facilitate the use of the study methodology, an implementation in an R shiny application was developed [12]. The software allows graphic selection of variable transformations. The application workflow is as follows: (1) the user uploads a dataset and metadata file that defines the variables' properties, (2) the software output is a set of the most suitable transformations for each variable, presented in the form of a histogram/bar plot (with a density curve when appropriate), (3) the user selects the desired transformation and can download the transformed dataset, and (4) variables are displayed in a descending order according to their Yule index and relevance.

The application allows data transformation regardless of modeling procedure. Visualization of the choices and selection made allows the user to transform hundreds of variables in a matter of minutes, without the burden of using scripts, remembering the available transformations, and tuning parameters.

Results

Findings

The improvement in the results of statistical analysis methods such as analysis of variance and linear regression, after the preprocessing methods were applied, are described below. Additional benefits gained by symmetry-targeting transformations, such as reduction of variance variability and linearity, are presented next. A dissimilarity matrix is usually based on some distance metric and may therefore be highly deformed due to the absence of symmetry. In a highly skewed variable, the distances between the elements in the bulk of the data will be substantially smaller than the distances between the very few large ones, the latter marring the importance of differences for most of the observations. The improvement in clustering is demonstrated using simulated data in the concluding section of the Results.

Transforming Medical Data for Symmetry

The first example is the semi-automated output for the variable NPITOTAL (neuropsychiatric inventory total score) in the ADNI cohort represents the total score of the psychiatric inventory exam. Examinations are shown in Figure 1. Since the original distribution was right skewed, the application offers performing an inverse transformation, a log transformation, and a power of 0.5 transformation (namely square root transformation). In this case, it appears that the inverse transformation is the most suitable.

The variable UPDRS Part 3 measures the score for the integrated motor-related condition of PD patients. Figure 2 displays the right-skewed frequency histogram (density plot) for the raw values, and the appropiate square root transformation.

Even after transformation, the distributions of the variables are not necessarily perfectly symmetrical, but their skewness decreased substantially.

The Gains From Striving for Symmetry

Although the selected transformations relied only on symmetry as the desirable target, they demonstrate how helpful these can be in achieving other goals as well.

Analysis of variance is most frequently used in order to check for differences in the means between three or more groups. It is therefore a useful tool when screening for variables, which have some marginal association with categorical variables such as the diagnosis groups. Figure 3 shows the variable of the total score of Everyday Cognition as self-stated by the participant compared to 10 years before (EcogPtTotal [everyday cognition participant total score]) before and after the above-mentioned symmetry transformation. The raw values over the five diagnosis categories assigned by the ADNI research using boxplots are compared in Figure 3. The variability (as reflected by the box lengths and the ranges) changes between the groups: the higher the median (reflecting the center), the higher the variability. The same groups are displayed for the transformed variable on the right side of Figure 3. Not only is the distribution within each group symmetric, but the dependency of variability on the median disappears and is quite homogeneous across the diagnosis groups.

Figure 1. Screenshot of semi-automated transformation workflow application of Alzheimer's Disease Neuroimaging Initiative (ADNI) patients' data variable neuropsychiatric inventory total score distribution.

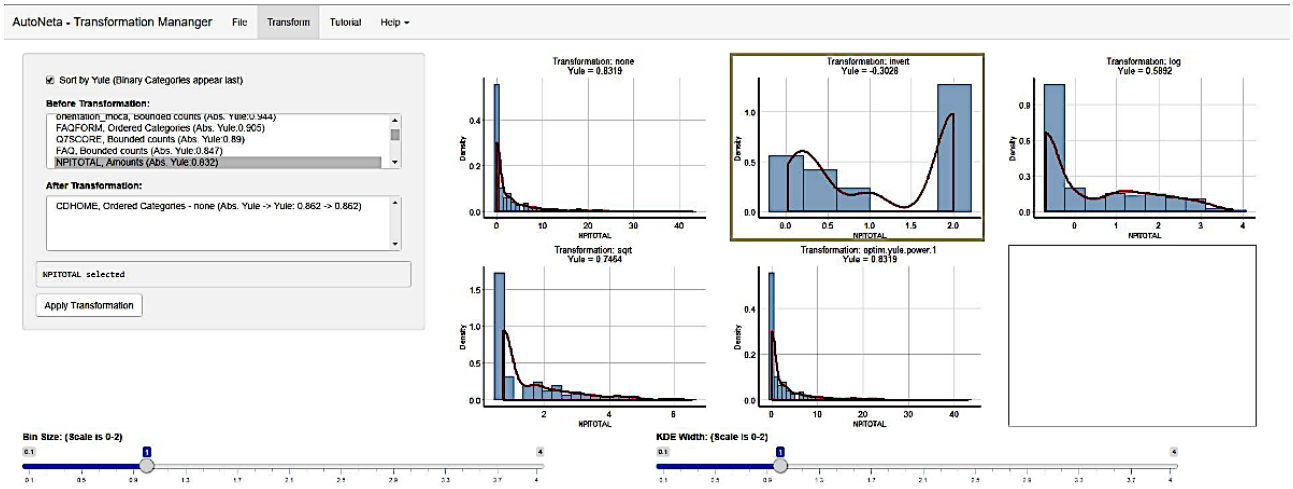


Figure 2. Unified Parkinson's Disease Rating Scale (UPDRS) Part 3 (motor part): variable raw data distribution, variable after sqrt transformation.

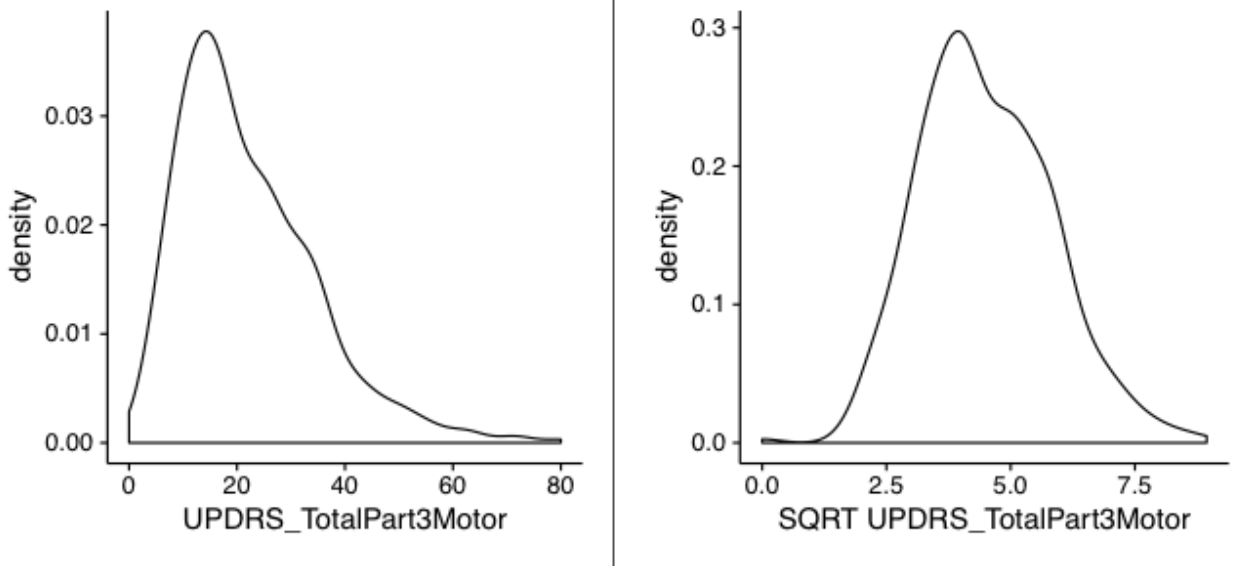
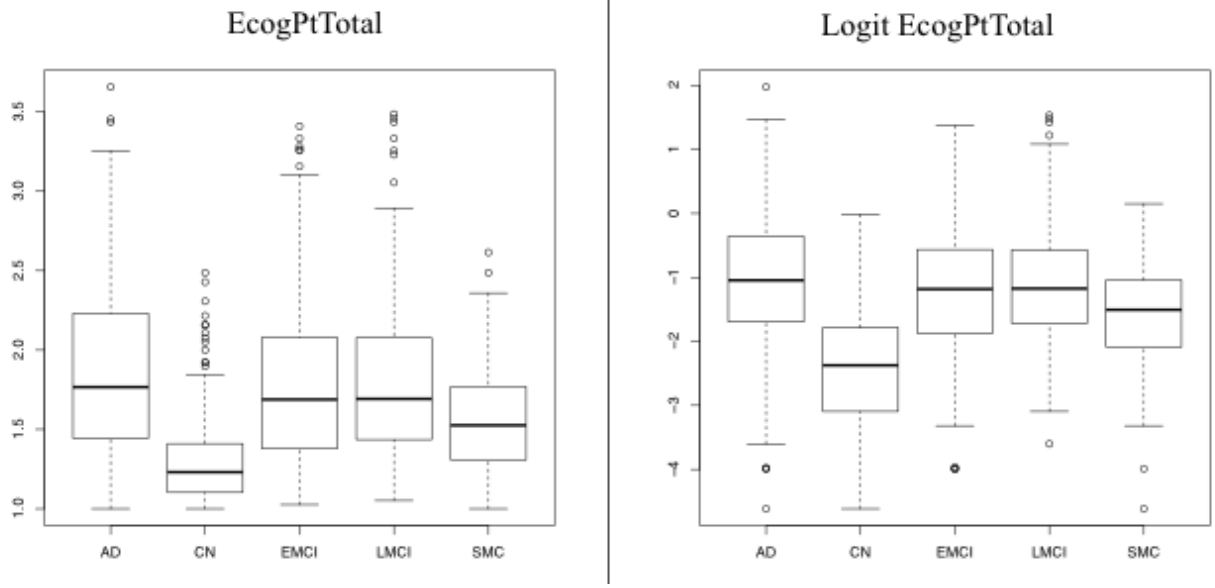


Figure 3. EcogPtTotal (everyday cognition participant total score) variable raw data over Alzheimer's Disease Neuroimaging Initiative (ADNI)'s five diagnosis categories before and after Logit transformation.



The linearity of relationships is a desirable property, allowing simpler models and better predictions outside the range of available data. These can later be extrapolated and easily explained. The relationship between EcogPtTotal and geriatric depression total score (GDTotal; score for the integrated geriatric depression scale in the ADNI data) is explored and shown in Figure 4 for the raw variables. The relationship is shown after the variables were transformed, using symmetry of each variable separately as our only criterion. On the raw scales, the monotone relationship (solid line) is far from linear (dashed line) and is rather concave. On the transformed scales, the relationship is quite linear.

Increase of the stability of variability following the transformation is beneficial as well. The distance between the two quartile lines (grayed area) increases with the progression in GDTotal value (Figure 4). In contrast, after transformation, the distance is stable all across the relationships (Figure 4). It should be emphasized that even though the transformation was chosen to approximately symmetrize the original variable, it is the distribution of the residuals that gained symmetry and homogeneous variability—the property we need. This fortunate outcome is likely to happen across many modeling efforts.

Both linearity of relationships and increase of stability of variability can be used to demonstrate the practical importance of simplicity of a relationship. Suppose we want to identify unusually low self-assessed deterioration, not explained merely by depression, which may indicate a more serious cognitive deterioration. The physician should insert the result of GDTotal

into a computer program (that expresses the nonlinear relationship), get the typical EcogPtTotal (standard deviation 2), calculated for that specific GDTotal value, and check whether the patient's value is above it. Instead, if the measurement is reported in the transformed way, the physician can immediately see whether the patient is above this threshold and by how much. The stability of the calculation and the possibility to extend the relationship into the region where only a few measurements exist are substantially improved.

The Gains in Cluster Analysis by Aiming for Symmetry

Recall that a simulated dataset of the clinical situation described above was generated. First, we performed a 2-step analysis of the original symmetric data: (1) select a subset of the clinical measurements using stepwise selection with false discovery rate control for q=0.05 [13], and (2) perform cluster analysis using the partition around medoids (K-medoids) algorithm [14] with Manhattan distance matrix. Then, we "damaged" some of the meaningful features by monotone transformation so that their distribution became asymmetric and repeated the analysis (see Multimedia Appendix 2 for further details on the simulation). The lack of symmetry following the process is clearly observed and is further emphasized by the value of the Yule index reported below the histograms. We evaluated the quality of the clustering using the confusion matrices and the adjusted Rand index. The results were conclusive: while symmetric variable distribution has little to no effect on the selection results, it is crucial for the clustering step. The confusion matrices and the adjusted Rand values that demonstrate the results are presented in Table 1. The use of different subset selection methods, different clustering algorithms, or different distance matrices had no effect on the direction of the results and had only a minor, if any, effect on its magnitude.

Figure 4. Relationship between EcogPtTotal (everyday cognition participant total score) and GDTotal (score for the integrated geriatric depression scale) raw values and after transformations.

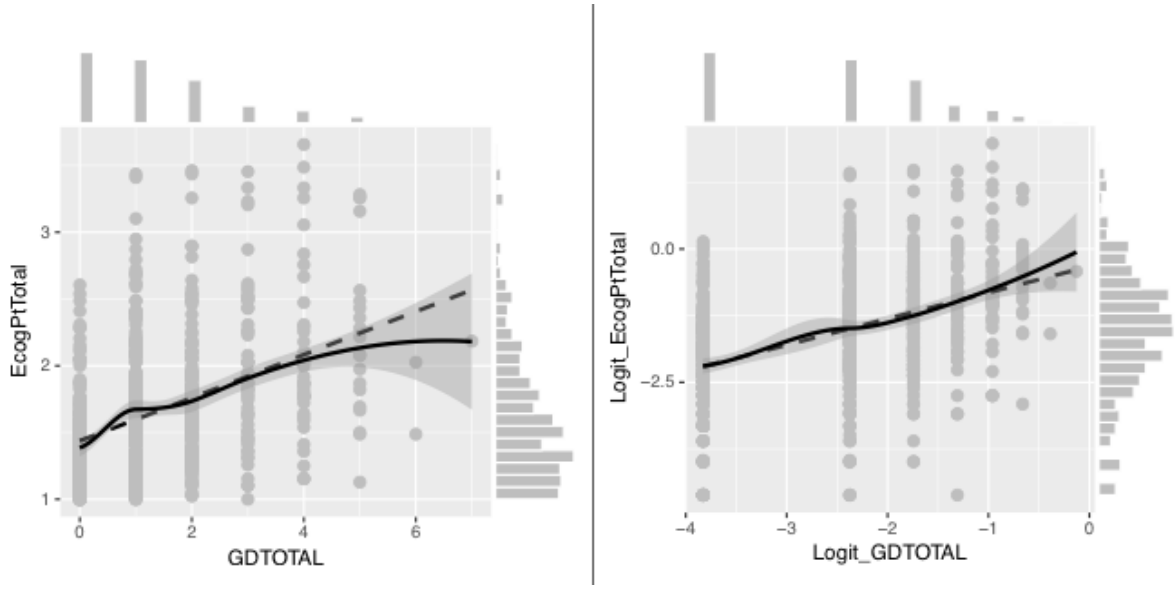


Table 1. Comparison of cluster analysis results[a].

True cluster	Assigned membership using symmetric data[b]						Assigned membership using skewed data[c]					
	1	2	3	4	5	6	1	3	2	4	6	5
0	200	0	0	0	0	0	194	0	6	0	0	0
1.1	0	200	0	0	0	0	0	200	0	0	0	0
1.2	0	0	200	0	0	0	0	63	137	0	0	0
2	0	0	0	200	0	0	0	0	0	140	1	59
3	0	0	0	0	199	1	2	0	0	2	196	0
23	0	0	0	0	6	194	0	1	0	91	7	101

[a]Confusion matrix of real group membership versus assigned cluster using K-medoids algorithm and Manhattan distance matrix used on symmetric (left) and asymmetric (right) features.

[b]Adjusted Rand=0.986.

[c]Adjusted Rand=0.681.

Discussion

Principal Results

The importance of data transformation in the preprocessing stage for any big data analysis, especially in medicine, is presented in this paper. We focused on the use of data transformation as the lesser known part of data preparation. Reviewing articles in medical informatics, we noticed that the preprocessing methods are not always mentioned. Even when they are performed, in many cases the process does not include variable transformations (Multimedia Appendix 1). It is our strong belief that the preprocessing steps taken have to be detailed as this may be crucial to both achievement of valuable results and to their reproducibility.

As seen in the examples from the large and robust databases of clinical and genetics information of AD and PD patients, the use of transformations results in simpler, linear, additive models with homogeneous variability. This enhances the ability to extrapolate and produce more accurate predictions, with better distance calculations and reduced complexity facilitating the integration into systems and devices. As shown, performing data analysis without transformations is possible but may require complex nonlinear models to explain the data.

The purpose of data analysis is to discover new insights. In this sense, the interpretation is an essential factor in the success of a data analysis process. The described methods increase the translational ability of the results for clinicians. They provide the possibility to apply simpler and more explainable models. The variables can be reverted to the original values before presenting and communicating the results. Methods that serve as a "black-box" have difficulties gaining the trust of physicians. This, in part, is due to the lack of transparent explanations of the process leading to the results.

Moreover, there are patterns in the data that only a human can notice. An example is a case where one transformation had the smallest skewness value, but the proper transformation was actually another (Figure 1). This situation occurred due to the "double hump" of the distribution of the data, where many subjects had a value of 0 while the rest were distributed symmetrically. This example is not rare in medical data and can easily be recognized by the human eye. Therefore, the methodology we propose aims to benefit from automation of most of the procedures but leaves the final decision to the human team of experts.

Limitations

This paper is based on our experience in using the methodology and the application. As use of the application will grow, we expect to improve with the feedback of users. The use of the methodology requires understanding of the data and metadata characteristics—this might be time consuming but intimate familiarity with the data is key to successful data analysis.

The benefit of nonlinear transformations was demonstrated by simulation only for the purpose of clustering, as it is difficult to find a medical example of clusters of patients, well defined by a known set of variables, which can serve as a vehicle for such a demonstration.

Conclusions

The use of nonlinear transformations as part of the preprocessing is important and affects the quality of the results. Symmetry-targeted transformations contribute significantly to other aspects of data analysis, enabling simpler and more translational models.

Acknowledgments

Data collection and sharing for this project was funded by ADNI (National Institutes of Health Grant U01 AG024904) and Department of Defense ADNI (DOD award #W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from many private pharmaceutical companies and nonprofit organizations. Data used in preparation of this article were obtained from the ADNI database. As such, the investigators within ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found on the ADNI website. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement #604102 (Human Brain Project). We thank Shiri Diskin, PhD, for her editorial assistance in manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Articles of data mining and machine-learning algorithms used for classification of clinical data.

[PDF File (Adobe PDF File), 60KB - medinform_v6i2e27_app1.pdf]

Multimedia Appendix 2

Transformations extended details.

[PDF File (Adobe PDF File), 31KB - medinform_v6i2e27_app2.pdf]

References

1. Wang W, Krishnan E. Big data and clinicians: a review on the state of the science. JMIR Med Inform 2014 Jan 17;2(1):e1 [FREE Full text] [doi: 10.2196/medinform.2913] [Medline: 25600256]
2. Tukey JW. Exploratory data analysis. New York, NY: Pearson; 1977:i-688.
3. Emerson J, Stoto M. Exploratory Methods for Choosing Power Transformations. J Am Stat Assoc 1982 Mar;77(377):103. [doi: 10.2307/2287775]
4. Andrews D. A Robust Method for Multiple Linear Regression. Technometrics 1974 Nov;16(4):523. [doi: 10.2307/1267603]
5. Durrleman S, Simon R. Flexible regression models with cubic splines. Stat Med 1989 May;8(5):551-561. [Medline: 2657958]
6. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. J Am Stat Assoc 1979 Dec;74(368):829. [doi: 10.2307/2286407]
7. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, NY: Springer; 2001.
8. Box G, Cox D. An Analysis of Transformations. J R Stat Soc 1964;26(ii):221-252.
9. Benjamini Y, Krieger A. Concepts and measures for skewness with data-analytic implications. Can J Statistics 1996 Mar;24(1):131-140. [doi: 10.2307/3315695]
10. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement 2013 Sep;9(5):e111-e194 [FREE Full text] [doi: 10.1016/j.jalz.2013.05.1769] [Medline: 23932184]
11. Rand W. Objective Criteria for the Evaluation of Clustering Methods. J Am Stat Assoc 1971 Dec;66(336):846-850. [doi: 10.2307/2284239]
12. Frostig T, Brill B. AutoTransMan. URL: https://tzviel.shinyapps.io/AutoTransMan/ [accessed 2018-04-16] [WebCite Cache ID 6yjDaabV7]
13. Benjamini Y, Gavrilov Y. A simple forward selection procedure based on false discovery rate control. Ann Appl Stat 2009 Mar;3(1):179-198. [doi: 10.1214/08-AOAS194]
14. Kaufman L, Rousseeuw P. Clustering by means of medoids. Netherlands: Faculty of Mathematics and Informatics, Delft University of Technology; Jan 01, 1987.

Abbreviations

AD: Alzheimer’s disease
ADNI: Alzheimer’s Disease Neuroimaging Initiative
EcogPtTotal: everyday cognition participant total score
GDTotal: geriatric depression total score
PD: Parkinson’s disease
UPDRS: Unified Parkinson’s Disease Rating Scale

Edited by G Eysenbach; submitted 13.05.17; peer-reviewed by D De Veaux, J Bian, A Benis; comments to author 15.08.17; revised version received 10.01.18; accepted 03.02.18; published 11.05.18.

Please cite as:
Shachar N, Mitelpunkt A, Kozlovski T, Galili T, Frostig T, Brill B, Marcus-Kalish M, Benjamini Y
The Importance of Nonlinear Transformations Use in Medical Data Analysis
JMIR Med Inform 2018;6(2):e27
URL: http://medinform.jmir.org/2018/2/e27/
doi:10.2196/medinform.7992
PMID:29752251

Original Paper

Agile Acceptance Test–Driven Development of Clinical Decision Support Advisories: Feasibility of Using Open Source Software

Mujeeb A Basit[1], MD, MMSc; Krystal L Baldwin[1], BS; Vaishnavi Kannan[1], MS; Emily L Flahaven[1], MSN, RN; Cassandra J Parks[1], BS; Jason M Ott[1], BS; Duwayne L Willett[1], MD, MS

University of Texas Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
Mujeeb A Basit, MD, MMSc
University of Texas Southwestern Medical Center
5323 Harry Hines Boulevard
Dallas, TX, 75390
United States
Phone: 1 214 648 1303
Email: mujeeb.basit@utsouthwestern.edu

Abstract

Background: Moving to electronic health records (EHRs) confers substantial benefits but risks unintended consequences. Modern EHRs consist of complex software code with extensive local configurability options, which can introduce defects. Defects in clinical decision support (CDS) tools are surprisingly common. Feasible approaches to prevent and detect defects in EHR configuration, including CDS tools, are needed. In complex software systems, use of test–driven development and automated regression testing promotes reliability. Test–driven development encourages modular, testable design and expanding regression test coverage. Automated regression test suites improve software quality, providing a "safety net" for future software modifications. Each automated acceptance test serves multiple purposes, as requirements (prior to build), acceptance testing (on completion of build), regression testing (once live), and "living" design documentation. Rapid-cycle development or "agile" methods are being successfully applied to CDS development. The agile practice of automated test–driven development is not widely adopted, perhaps because most EHR software code is vendor-developed. However, key CDS advisory configuration design decisions and rules stored in the EHR may prove amenable to automated testing as "executable requirements."

Objective: We aimed to establish feasibility of acceptance test–driven development of clinical decision support advisories in a commonly used EHR, using an open source automated acceptance testing framework (FitNesse).

Methods: Acceptance tests were initially constructed as spreadsheet tables to facilitate clinical review. Each table specified one aspect of the CDS advisory's expected behavior. Table contents were then imported into a test suite in FitNesse, which queried the EHR database to automate testing. Tests and corresponding CDS configuration were migrated together from the development environment to production, with tests becoming part of the production regression test suite.

Results: We used test–driven development to construct a new CDS tool advising Emergency Department nurses to perform a swallowing assessment prior to administering oral medication to a patient with suspected stroke. Test tables specified desired behavior for (1) applicable clinical settings, (2) triggering action, (3) rule logic, (4) user interface, and (5) system actions in response to user input. Automated test suite results for the "executable requirements" are shown prior to building the CDS alert, during build, and after successful build.

Conclusions: Automated acceptance test–driven development and continuous regression testing of CDS configuration in a commercial EHR proves feasible with open source software. Automated test–driven development offers one potential contribution to achieving high-reliability EHR configuration. Vetting acceptance tests with clinicians elicits their input on crucial configuration details early during initial CDS design and iteratively during rapid-cycle optimization.

(JMIR Med Inform 2018;6(2):e23) doi:10.2196/medinform.9679

KEYWORDS

clinical decision support systems; electronic health records; software validation; software verification; agile methods; test driven development

Introduction

Defects in Clinical Decision Support Tools

"Making the right thing the easy thing to do" for clinicians using an electronic health record (EHR) drives many current efforts to promote delivery of reliable, high-quality care. Clinical decision support (CDS) within the EHR provides one mechanism, by supplying advisories suggesting best practice care for a patient's specific conditions [1,2].

With the move to EHRs came recognition that unintended consequences can ensue [3,4], even jeopardizing patient safety [5,6]. Modern EHRs comprise complex software code with extensive local configurability options, affording opportunities for defects to be introduced. Feasible approaches to prevent and detect defects in EHR configuration are needed.

Defects in CDS tools are surprisingly common and can cause either over-expression or under-expression of alerts [7-9]. The latter can go undetected for long periods. Common causes of CDS defects include changes to data codes, terminologies, or modules external to the CDS itself [7].

Test–Driven Development

In complex software systems, use of test–driven development (TDD) and automated regression testing promotes reliability [10,11]. In TDD, a new requirement is specified as a test before code is written, following a "red-green-refactor" pattern: the test fails initially ("red") then passes once the software meets all test-specified requirements ("green"). Subsequent refinements to the underlying code (refactoring) can occur, following the same cycle (Figure 1).

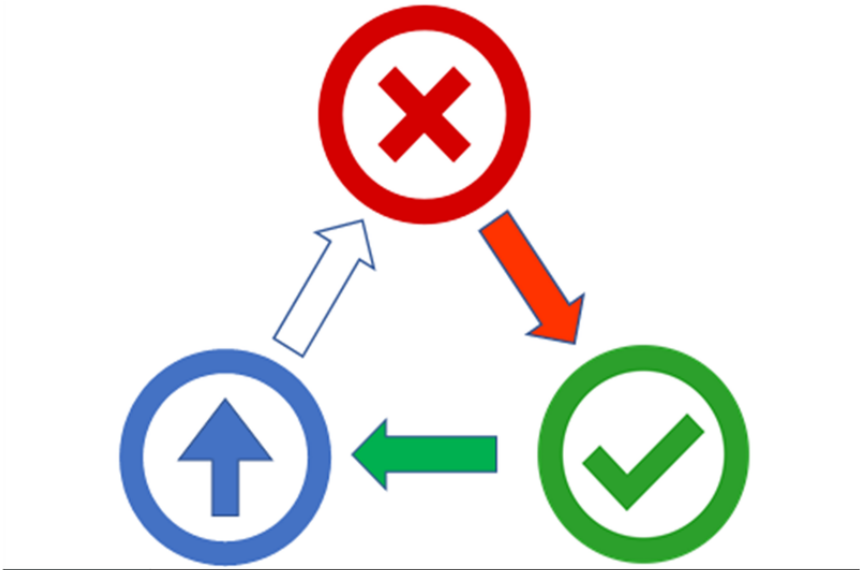
Benefits of TDD include (1) encouragement of modular design and (2) growth of regression test suites. Automated regression test suites improve software quality and provide a "safety net" for later modification without fear of undetected breakage [12]. TDD can be done at the micro (unit test) and macro (acceptance test) levels. Each automated acceptance test serves multiple purposes, as requirements definition (prior to build), acceptance testing (on completion of build), regression testing (after go-live), and documentation of design (for long-term reference) [13].

Potential Applications of Acceptance Test–Driven Development for Clinical Decision Support Advisories

Rapid-cycle development or "agile" methods are being successfully applied to CDS development [14-16]. The agile practice of automated TDD is not widely adopted, perhaps because most EHR software is vendor-developed. However, key CDS advisory configuration design decisions amenable to automated testing as "executable requirements" include [17]:

- any *restrictions* on where the CDS alert logic should be evaluated (ie, restricted to only certain practice locations, encounter types, provider types) to help target the most appropriate situations and limit "alert fatigue" [18,19]
- *triggering action(s)* that prompt evaluation of the CDS advisory logic at the right time in the workflow (eg, opening the chart, placing an order, entering a diagnosis, and other options)
- *rule logic* for evaluating whether the advisory should display ("fire"), decided by evaluating discrete data in the EHR
- the *user interface* (UI) displayed after the rule logic passes, including instructions and contextual information presented, and the range of action options provided
- *system actions and state changes* that should occur in the EHR following any clinician interactions with the UI.

Figure 1. Test–driven development cycle.



In this paper, we use examples of each of the above (in a widely used EHR) to demonstrate how TDD of CDS advisories can work in practice during development of a CDS tool.

Clinical Background for an Example Clinical Decision Support Request

Patients who present to the emergency room with an acute stroke may have impaired swallowing mechanisms. Attempting to give medications orally creates a risk of the patient aspirating medication into the lungs. Accordingly, patients with known or suspected stroke are screened for swallowing difficulties prior to attempting administration of oral medication. In a busy emergency room setting, keeping track of whether the needed screening has been done can be challenging. Accordingly, interruptive CDS was requested if the intended swallow screening had not yet occurred.

Methods

Location

All activities in this report took place at the University of Texas Southwestern Medical Center in Dallas, Texas. This work was judged not to be human subjects research and thus did not require presentation to our Institutional Review Board.

Software

Automated testing employed the open source testing software FitNesse, based on the Framework for Integrated Testing, along with the dbFit extension for querying databases [20-22]. Time and personnel requirement estimates for initial configuration of FitNesse and dbFit testing framework are given in Table 1. Electronic health record software at UT Southwestern is from Epic, and the incident management software is ServiceNow.

Procedures

High-Level Requirements With User Stories

Initial high-level requirements for new CDS advisories were gathered as user stories [23], written from the perspective of the clinician receiving the alert: “As a <clinician role>, I want <to be advised about something>, so that <a benefit can be achieved>”.

Through clinical conversations, user stories were elaborated with more specific acceptance criteria describing what would constitute successful CDS advisory behavior, often initially as a simple bulleted list. In this project, certain acceptance criteria were further detailed unambiguously as automatable acceptance tests.

Automated Acceptance Test–Driven Development

Acceptance tests were initially constructed as tables in a spreadsheet (Microsoft Excel workbook), to facilitate clinical vetting and shared review. Each table specified one configuration aspect of the CDS advisory. Table contents were then imported into an automated test suite in FitNesse (see the earlier section, Software). For each table-based test, a structured query language (SQL) query retrieved the corresponding CDS advisory configuration information from the EHR development environment’s database. A FitNesse test suite template was created containing the most frequently used specification tables and corresponding SQL queries, streamlining test generation for each new CDS advisory.

For configuration management, both the CDS advisory and its associated test were migrated from the development environment to the test environment at the same time, for integrated testing. Similarly, when migrating to production, the corresponding FitNesse test(s) were added to the automated regression test suite for the production environment.

Table 1. Configuration of FitNesse and dbFit: time and personnel requirements. EHR: electronic health record; IT: information technology; SQL: structured query language.

Task category	Task	Frequency	Time (range)	Type of personnel
Initial set-up of FitNesse + dbFit testing framework	Download and install FitNesse to point of functioning FitNesse wiki	Once	30 minutes	IT analyst
	Configure FitNesse to use Active Directory login permissions (if desired)	Once	2 hours to 1 day	IT analyst knowledgeable about one’s local Active Directory
	Configure dbFit	Once	Few minutes to 2 hours	IT analyst
	Set up database connection for FitNesse/dbFit to query an EHR (or other) database	Once per database	1 hour (if first time doing); a few minutes per connection once experienced	IT analyst
Create a test “template” for a given type of test	Write SQL to serve as template for given type of test	Once per new type of test	1 to 2 hours	EHR analyst; SQL writer (can be same person)
Configure an individual test instance	Create Microsoft Excel copy of test template and populate for given test instance, ready for vetting with clinician or other customer	Once per test instance	15 to 60 minutes	EHR analyst
	Import Microsoft Excel test to FitNesse Test page, and test	Once per test instance	10 to 15 minutes	EHR analyst or test team analyst

Any subsequent failures of regression tests in production would initiate a new entry in the incident management system for investigation and resolution.

Requirements Elicitation

A nurse informaticist (EF) and an EHR analyst (JO) met with the front-line nurses and nurse manager from the Emergency Department (ED) to define the problem and frame the user story for the alert in a way these nurse clinicians believed would be beneficial within their workflow. The same EHR analyst also had standing meetings with the ED nursing and medical staff at least weekly; those sessions were used to further elaborate more detailed acceptance criteria for the user story.

Results

User Story for a Clinical Decision Support Best Practice Advisory

"As an emergency room nurse, I want to be alerted before I administer an oral medication to a patient with known or suspected stroke if they've not yet had their Swallow Screen performed, so that my patient can receive their medications by the most safe and effective route."

Automated Acceptance Tests for the Clinical Decision Support Advisory

Restrictions

Restrictions help focus the advisory to the right practice setting and clinician type, reducing alert fatigue for clinicians where the advisory would not be relevant (Figure 2). This test specified that this alert should apply only in Emergency Medicine departments and only to nurses.

Triggering Action

Triggering actions further focus when the advisory's logic should be evaluated to the most relevant point(s) in clinicians' workflow—for example, only when entering or signing an order, entering a diagnosis, administering a medication, or (most invasively) on every entry into the patient's chart. Our stroke swallowing advisory was to trigger logic evaluation when the nurse prepares a medication for administration to a patient—specifically, at the time of barcode scanning the medication due (Figure 3).

Rule Logic

The rule logic for deciding whether a CDS advisory should appear to a clinician was first modeled as a decision tree (Figure 4), then specified as test tables (Figure 5). The specified logic checks for any of three potential indications that the patient has a known or suspected stroke diagnosis, then for a planned oral route of the barcode scanned medication, and finally whether the Stroke Swallow (dysphagia) Screen has been performed.

User Interface

In addition to specifying the wording on the advisory (not shown and which includes instructional diagrams and text for performing the Swallow Screen), acceptance tests can also specify what follow-up actions the clinician may be prompted to perform (Figure 6).

This requirement test specified that the nurse should be able to indicate directly from the alert's UI whether the patient passed or failed the Swallow Screen, without having to leave the alert and navigate to the swallow screening flowsheet in another part of the chart. This follow-up action still populated the same flowsheet behind the scenes, however, for data consistency. Neither option was to be defaulted as pre-selected—both were specified to initially appear unselected.

System Actions

As an alternative to the prompted action, the clinician may select an "acknowledge reason" exception why the primary action was not taken, resulting in the system setting a "lock out" time to avoid repetitive firing, and optionally file a specific data element for data capture (Figure 7).

For instance, Line 1 of this test specifies that once a clinician determines oral medications are allowed for this patient, the alert should not fire on subsequent medication administrations during the current ED encounter for a lockout period of 24 hours, limiting alert fatigue.

Figure 2. Screenshot of FitNesse test specifying Department Specialty and Provider Type restrictions. n/a: not applicable.

Specify restrictions on when this alert should apply:

Line	Encounter Type?	Department Specialty?	Department?	Provider Type?
1	n/a	Emergency Medicine	n/a	Registered Nurse
2	n/a	Emergency Medicine	n/a	Licensed Nurse

Figure 3. Screenshot of test specifying triggering action for this advisory.

Specify what triggering actions prompt alert logic:

Triggering Action?
Medication Administration

Figure 4. Decision tree for the advisory. NIH: National Institutes of Health.

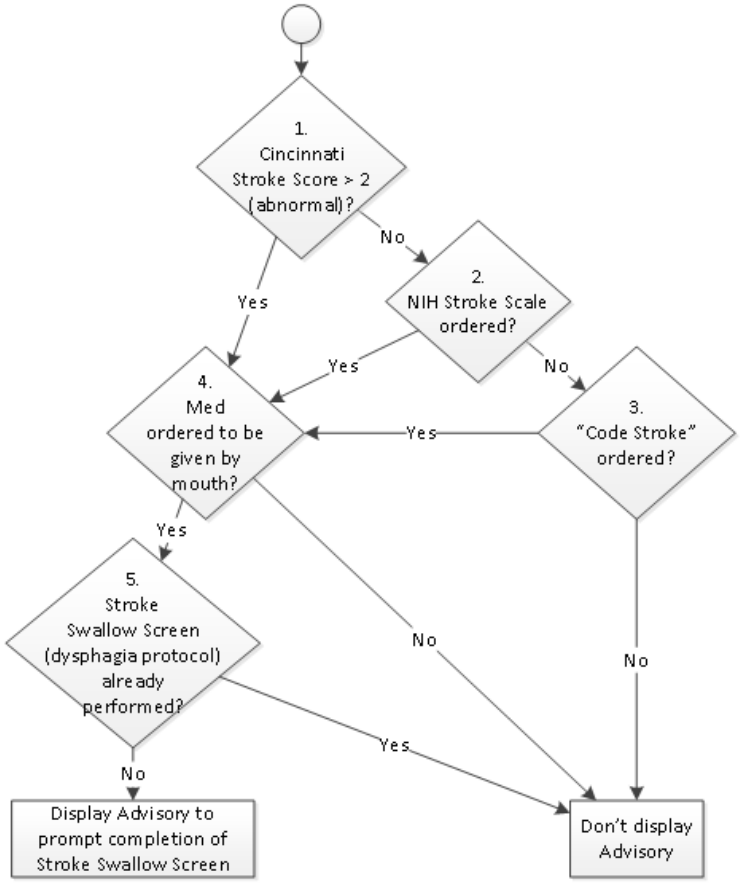


Figure 5. Screenshot of test specifying clinical decision support rule logic. CINN: Cincinnati; NIH: National Institutes of Health.

Specify linked criteria used to evaluate this alert:

Line	Linked Criteria Name?
1	UTSW ED STROKE ABNORMAL CINN STROKE 2 OR GREATER CRITERIA
2	UTSW ED STROKE IS NIH STROKE SCALE ORDERED
3	UTSW ED STROKE IS CODE STROKE ORDERED CRITERIA
4	UTSW ED STROKE RX MEDS WITH ROUTE ORAL CRITERIA
5	UTSW ED STROKE DYSPHAGIA PROTOCOL PERFORMED CRITERIA

Specify Boolean logic to combine above linked criteria:

Boolean Logic
(1 OR 2 OR 3) AND 4 AND NOT 5

Figure 6. Screenshot of test specifying user interface actions for the advisory. BPA: Best Practice Advisory

Specify Follow-up Actions available from the alert's user interface:

Line	Followup Action Button Label?	Preselected Or Unselected?	Followup Extension Description?
1	Swallow Screening Passed	Unselected	This extension is used to file a numeric or string flowsheet value by attaching to the Follow-Up Actions of a BPA base record
2	Swallow Screening Failed	Unselected	This extension is used to file a numeric or string flowsheet value by attaching to the Follow-Up Actions of a BPA base record

Figure 7. Screenshot of test specifying system actions following clinician response. BPA: Best Practice Advisory; PO: per os.

Specify Acknowledge Reason options on the alert's user interface:

Line	Button Caption?	Acknowledge Reason?	Lockout Hours?	Lockout Context?	SDE ID?
1	PO Meds are allowed for this patient	Level 1	24	All users, current encounter only	n/a
2	Cinn Stroke Scale is now higher than 12	Level 2	6	All users, current encounter only	n/a
3	BPA Inappropriate	Inappropriate	24	All users, current encounter only	n/a

Test–Driven Development Cycle

Before Development

Before development has begun, all test assertions should fail and do (Figure 8).

During Development

During development, some tests begin to pass. When construction of the CDS advisory is complete, the test suite can indicate if any requirements are not yet met (Figure 9).

FitNesse automatically displays any discrepancies between expected and actual advisory design. On Line 1 in Figure 9, the Lockout Hours setting was specified as 24 hours but initially configured to 2 hours, which if unchanged would cause significant over-firing of the alert to busy nurses.

After Successful Development

Following completion of build and resolution of any discrepancies from specified requirements, the test page for the "base" alert record passes completely (Figure 10).

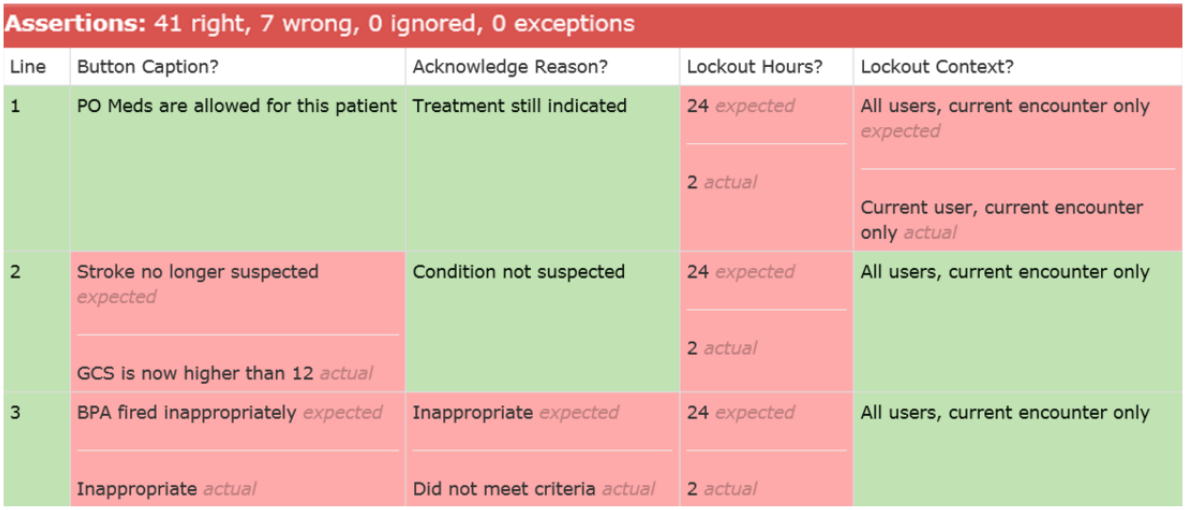
Similar test pages were developed to specify acceptance criteria for the 5 "criteria" records referenced by the base alert record (see Multimedia Appendices 1 and 2). The full test suite thus consisted of 6 test pages, encompassing 24 individual tests making 133 individual assertions. The total time to execute each test page and the full test suite are given in Table 2 (times are the average of 5 test suite executions). The full suite averages 0.933 seconds to run, most of which is suite set-up and wrap-up time. Each test page execution takes only 2-4 ms (0.002-0.004 s).

For reference, our current FitNesse regression test suite in production currently has 85 Test Pages, 6126 individual test Assertions, and runs in 165 seconds (2 min, 45 sec). Once the automated acceptance tests are fully passing, the CDS advisory then can be migrated from the Development environment to the Integrated Testing environment, and then to Production. The automated acceptance test suite is also added to the regression test suites in the latter two environments contemporaneously with migrating the CDS code, to ensure continued proper behavior in all environments.

Figure 8. Screenshot of acceptance test: all assertions fail as expected prior to build.

Assertions: 0 right, 15 wrong, 0 ignored, 0 exceptions (0.644 seconds)

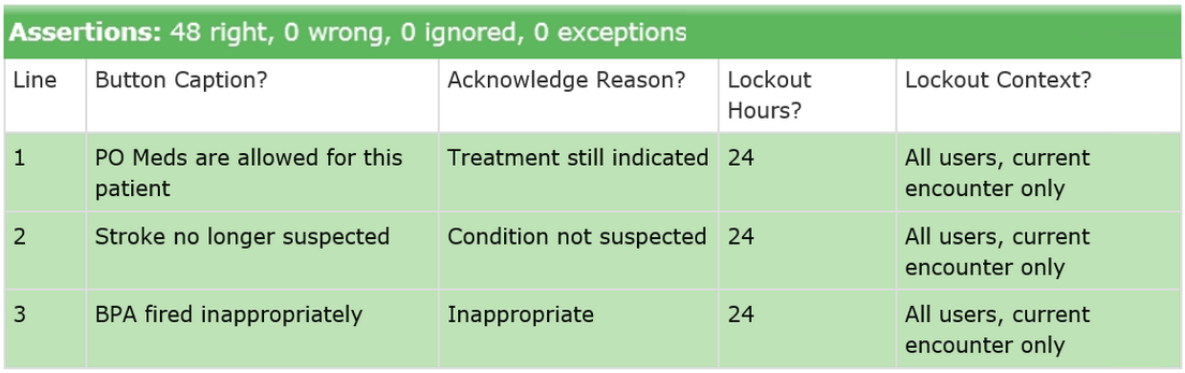
Figure 9. Screenshot of a test table included in the acceptance test suite: acceptance test partially passes following initial build. GCS: Glasgow Coma Scale; PO: per os.



Assertions: 41 right, 7 wrong, 0 ignored, 0 exceptions

Line	Button Caption?	Acknowledge Reason?	Lockout Hours?	Lockout Context?
1	PO Meds are allowed for this patient	Treatment still indicated	24 *expected* 2 *actual*	All users, current encounter only *expected* Current user, current encounter only *actual*
2	Stroke no longer suspected *expected* GCS is now higher than 12 *actual*	Condition not suspected	24 *expected* 2 *actual*	All users, current encounter only
3	BPA fired inappropriately *expected* Inappropriate *actual*	Inappropriate *expected* Did not meet criteria *actual*	24 *expected* 2 *actual*	All users, current encounter only

Figure 10. Screenshot of acceptance test assertions for "base" alert record, all passing following successful build. BPA: Best Practice Advisory; PO: per os.



Assertions: 48 right, 0 wrong, 0 ignored, 0 exceptions

Line	Button Caption?	Acknowledge Reason?	Lockout Hours?	Lockout Context?
1	PO Meds are allowed for this patient	Treatment still indicated	24	All users, current encounter only
2	Stroke no longer suspected	Condition not suspected	24	All users, current encounter only
3	BPA fired inappropriately	Inappropriate	24	All users, current encounter only

Table 2. Test suite: number of tests and individual assertions, with execution times. NIH: National Institutes of Health.

Type	Test page name	Tests	Assertions	Time (s)
Base	Alert Stroke Suspected But No Swallow Screen	8	48	0.003
Criteria	Criteria Abnormal Cincinnati Stroke Scale	3	14	0.002
Criteria	Criteria NIH Stroke Scale Ordered	3	19	0.001
Criteria	Criteria Code Stroke Ordered	4	24	0.002
Criteria	Criteria Med With Oral Route	3	19	0.002
Criteria	Criteria Stroke Dysphagia Screen Performed	3	14	0.002
Suite	Suite Story Stroke Swallow Screen	24	138	0.869

Iterative Development

Number of Iterations Required

Three 2-week development iterations were required for full implementation of this advisory, following requirements gathering with a user story and initial acceptance criteria.

- During a first 2-week iteration, automated acceptance tests were written and a first working version of the best practice advisory created and demonstrated. During testing, we discovered that the initial follow-up action specified by the test (a hyperlink to jump the nurse to the Swallow Screen documentation flowsheet) was not compatible with the trigger action desired (beginning medication administration).
- Accordingly, during a follow-on 2-week iteration, we pivoted to a different follow-up action to be taken from the advisory's UI, which enabled the nurse to document the Stroke Swallow screen results directly from the advisory UI. This filed the nurse's response to the identical Stroke Swallow documentation flowsheet row, while avoiding the need for the nurse to leave the advisory and jump to the flowsheet itself. Since the advisory's UI also includes graphical instructions for performing the Stroke Swallow

screen, this approach was well received by nursing representatives.
- During a third 2-week iteration, the alert was turned on in Production silently (not visible to end-users) to observe what situations triggered its firing. No over-firing in unwanted situations was detected. Under-firing was observed, due to frequent use in the ED of as-needed oral medication orders rather than scheduled medication orders. The criteria record's rule determining whether oral meds were ordered originally used a property evaluating for oral scheduled medications. This rule was re-specified to include an additional property evaluating for as-needed oral medications as well. After development to pass the revised test, the modified rule was re-migrated to Production.

Go-Live in Production

The alert was re-observed silently in Production for approximately 24 hours prior to enabling its display to end-users. Investigation of the alert's criteria evaluation for both real ED patients and test patients confirmed that the alert was behaving as expected in Production. Following "go-live" of the visible alert, no customer-logged "tickets" for aberrant alert behavior (eg, firing in unintended locations or situations) were received.

Discussion

Principal Results

Defects and unintended consequences occur too commonly in CDS advisories present in modern complex EHRs. Test–driven development offers one approach to help achieve higher reliability. In this study, we used open source software (FitNesse) to create "executable requirements" covering multiple important structural and behavioral dimensions of CDS advisory design: restrictions to applicable clinical settings, trigger(s) to invoke rule evaluation, rule logic, UI design, and system responses to clinician selections. This work demonstrates that acceptance TDD can feasibly be applied to configuring CDS advisories in a commercial EHR, generating suites of automated acceptance and regression tests.

Comparison With Prior Work

User-centered design methods now being applied in health care seek to optimize clinician and patient experience with software and include the equivalent of iterative manual acceptance testing [24-27]. We consider the use of automated acceptance and regression testing complementary, and a means of capturing insights from user-centered design in explicitly testable ways to ensure accurate implementation. Sophisticated automated generation of test cases for complex CDS tool logic has been previously described, to identify and test all possible guideline-permitted decision paths [28,29]. In those studies, clinician and patient user acceptance testing of interactions with the CDS tool itself remained manual, though testing of the CDS logic was fully automated.

Limitations

In this study, we demonstrated the feasibility of using TDD for CDS configuration in a commercial EHR: investigation over a longer period of adoption will be needed to measure the effect of TDD on CDS tools' quality in production.

The example chosen shows application of TDD to only one type of CDS (best practice advisories), in an advisory executing simple logic. However, the FitNesse framework in our experience can be readily applied to specifying more complex CDS rule logic assessing a wide variety of patient-specific data in the EHR and to testing many other aspects of EHR and non-EHR system configuration. For instance, we have applied FitNesse automated testing to:

- ensuring conformance with data business rules not enforced directly in software (eg, "If a provider is marked as participating in the EHR Incentive Program, they should also have their e-Prescribing flag set to Yes")
- specifying expected contents of tables with potential for major downstream impact if unexpectedly changed (eg, exact contents of the Provider Type and Encounter Type look-up tables, used extensively in CDS targeting, in reporting, and in a variety of operational uses)
- cross-system testing of mutually consistent configuration (eg, for the exact operating room location of vital sign monitoring equipment used by anesthesiologists, validate 100% consistency between middleware software and the EHR, to ensure vital signs are always interfaced to the correct surgical patient's record)

Given this versatility, we expect automated acceptance TDD to prove readily applicable to other types of CDS (such as order sets, cascading order questions, and rule-driven banners).

Another potential limitation is that FitNesse by design tests software "under the hood"; that is, under the UI level. FitNesse purposefully tests the business logic and data storage layers driving important application behavior, ideally insulated by modular design from minor modifications to the UI. Automated testing through the UI generally requires more maintenance and is more time-consuming and expensive to configure [13]. Nonetheless, testing through the UI can be necessary in some circumstances, for instance if the EHR software embeds certain business logic completely within the UI layer (without reference to business logic modules or configuration tables). To test those aspects, FitNesse would need to be augmented with an automated testing tool operating through the UI. We use such a tool (ie, TestComplete, SmartBear Software) for automated "journey" or scenario testing by a simulated user, complementary to automated TDD and regression testing of EHR configuration using FitNesse.

Conclusions

Automated acceptance testing and continuous regression testing of CDS configuration in a commercial EHR proves feasible with open source software. The problem of EHR safety is multifaceted, and multiple safety-enhancing approaches will almost certainly be needed [30]. Automated TDD offers one potential contribution towards achieving high-reliability EHR systems.

As another benefit, clinician frustration with the EHR can be reduced by judiciously limiting interruptive alerts to truly relevant circumstances where pop-up advice is seen as helpful,

not extraneous [31]. Vetting acceptance tests with clinicians elicits their input on crucial configuration details early during initial CDS design, as well as iteratively during rapid-cycle evolutionary development.

Acknowledgments

We thank our colleagues in UT Southwestern's Health System Information Resources and Enterprise Data Services departments for configuring the FitNesse automated testing framework, with a special thanks to Preston Park. We appreciate the leadership support of Marc Milstein, Mark Rauschuber, Kathryn Flores, Dennis Pfeifer, and Ki Lai making this work possible. Research was supported in part by NIH Grant: 5UL1TR001105-05 UT Southwestern Center for Translational Medicine.

Conflicts of Interest

None declared.

Multimedia Appendix 1

FitNesse test tables for Stroke Swallow Advisory - main alert record.

[XLSX File (Microsoft Excel File), 21KB - medinform_v6i2e23_app1.xlsx]

Multimedia Appendix 2

FitNesse test tables for Stroke Swallow Advisory - criteria records.

[XLSX File (Microsoft Excel File), 33KB - medinform_v6i2e23_app2.xlsx]

References

1. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 03;157(1):29-43. [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
2. Middleton B, Sittig DF, Wright A. Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. Yearb Med Inform 2016 Aug 02;Suppl 1:S103-S116 [FREE Full text] [doi: 10.15265/IYS-2016-s034] [Medline: 27488402]
3. Coiera E, Ash J, Berg M. The Unintended Consequences of Health Information Technology Revisited. Yearb Med Inform 2016 Nov 10(1):163-169 [FREE Full text] [doi: 10.15265/IY-2016-014] [Medline: 27830246]
4. Weiner JP, Kfuri T, Chan K, Fowles JB. "e-Iatrogenesis": the most critical unintended consequence of CPOE and other HIT. J Am Med Inform Assoc 2007;14(3):387-388; discussion 389 [FREE Full text] [doi: 10.1197/jamia.M2338] [Medline: 17329719]
5. Sittig DF, Classen DC. Safe electronic health record use requires a comprehensive monitoring and evaluation framework. JAMA 2010 Feb 03;303(5):450-451 [FREE Full text] [doi: 10.1001/jama.2010.61] [Medline: 20124542]
6. Sittig DF, Singh H. Toward More Proactive Approaches to Safety in the Electronic Health Record Era. Jt Comm J Qual Patient Saf 2017 Oct;43(10):540-547. [doi: 10.1016/j.jcjq.2017.06.005] [Medline: 28942779]
7. Wright A, Hickman TT, McEvoy D, Aaron S, Ai A, Andersen JM, et al. Analysis of clinical decision support system malfunctions: a case series and survey. J Am Med Inform Assoc 2016 Nov;23(6):1068-1076 [FREE Full text] [doi: 10.1093/jamia/ocw005] [Medline: 27026616]
8. Kassakian SZ, Yackel TR, Gorman PN, Dorr DA. Clinical decisions support malfunctions in a commercial electronic health record. Appl Clin Inform 2017 Sep 06;8(3):910-923. [doi: 10.4338/ACI-2017-01-RA-0006] [Medline: 28880046]
9. Liu S, Wright A, Hauskrecht M. Change-Point Detection Method for Clinical Decision Support System Rule Monitoring. Artif Intell Med (2017) 2017 Jun;10259:126-135 [FREE Full text] [doi: 10.1007/978-3-319-59758-4_14] [Medline: 28795172]
10. Beck K. Test-Driven Development: By Example. Boston, MA: Addison-Wesley Longman Publishing; Nov 18, 2002.
11. Pugh K. Lean-Agile Acceptance Test-Driven Development: Better Software Through Collaboration. Boston, MA: Addison-Wesley; Jan 05, 2011.
12. Freeman S, Pryce N. Growing Object-Oriented Software, Guided by Tests. Boston, MA: Addison-Wesley; Oct 22, 2009.
13. Adzic G. Specification by Example: How Successful Teams Deliver the Right Software. Greenwich, CT: Manning Publications Co; Jun 03, 2011.
14. Kannry J, McCullagh L, Kushniruk A, Mann D, Edonyabo D, McGinn T. A Framework for Usable and Effective Clinical Decision Support: Experience from the iCPR Randomized Clinical Trial. EGEMS (Wash DC) 2015;3(2):1150 [FREE Full text] [doi: 10.13063/2327-9214.1150] [Medline: 26290888]

15. Kannan V, Willett D. Agile Alliance Experience Reports. 2016. Agile clinical decision support development URL: https://www.agilealliance.org/resources/experience-reports/agile-clinical-decision-support-developments/ [WebCite Cache ID 6vuNlyvcv]
16. Kannan V, Fish JS, Mutz JM, Carrington AR, Lai K, Davis LS, et al. Rapid Development of Specialty Population Registries and Quality Measures from Electronic Health Record Data*. An Agile Framework. Methods Inf Med 2017 Jun 14;56(99):e74-e83 [FREE Full text] [doi: 10.3414/ME16-02-0031] [Medline: 28930362]
17. Wright A, Goldberg H, Hongsermeier T, Middleton B. A description and functional taxonomy of rule-based decision support content at a large integrated delivery network. J Am Med Inform Assoc 2007;14(4):489-496 [FREE Full text] [doi: 10.1197/jamia.M2364] [Medline: 17460131]
18. Topaz M, Seger DL, Slight SP, Goss F, Lai K, Wickner PG, et al. Rising drug allergy alert overrides in electronic health records: an observational retrospective study of a decade of experience. J Am Med Inform Assoc 2016 May;23(3):601-608. [doi: 10.1093/jamia/ocv143] [Medline: 26578227]
19. Press A, Khan S, McCullagh L, Schachter A, Pardo S, Kohn N, et al. Avoiding alert fatigue in pulmonary embolism decision support: a new method to examine 'trigger rates'. Evid Based Med 2016 Dec;21(6):203-207. [doi: 10.1136/ebmed-2016-110440] [Medline: 27664174]
20. Adzic G. Test Driven .NET Development with FitNesse. London: Neuri Limited; Feb 28, 2008.
21. Nikolov Y. DbFit - Automated open source database testing. 2016. URL: http://www.methodsandtools.com/tools/dbfit.php [WebCite Cache ID 6vvAJRzlC]
22. Benilov J. DbFit: Test-driven database development. 2015. URL: http://dbfit.github.io/dbfit/ [WebCite Cache ID 6vv9lwBI8]
23. Cohn M. User Stories Applied: For Agile Software Development. Boston, MA: Addison Wesley Longman Publishing; Mar 01, 2004.
24. Press A, McCullagh L, Khan S, Schachter A, Pardo S, McGinn T. Usability Testing of a Complex Clinical Decision Support Tool in the Emergency Department: Lessons Learned. JMIR Hum Factors 2015 Sep 10;2(2):e14 [FREE Full text] [doi: 10.2196/humanfactors.4537] [Medline: 27025540]
25. Witteman HO, Dansokho SC, Colquhoun H, Coulter A, Dugas M, Fagerlin A, et al. User-centered design and the development of patient decision aids: protocol for a systematic review. Syst Rev 2015 Jan 26;4:11 [FREE Full text] [doi: 10.1186/2046-4053-4-11] [Medline: 25623074]
26. Norman DA, Draper SW. User Centered System Design; New Perspectives on Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.
27. Melnick ER, Hess EP, Guo G, Breslin M, Lopez K, Pavlo AJ, et al. Patient-Centered Decision Support: Formative Usability Evaluation of Integrated Clinical Decision Support With a Patient Decision Aid for Minor Head Injury in the Emergency Department. J Med Internet Res 2017 May 19;19(5):e174 [FREE Full text] [doi: 10.2196/jmir.7846] [Medline: 28526667]
28. Lobach DF, Johns EB, Halpenny B, Saunders T, Brzozowski J, Del FG, et al. Increasing Complexity in Rule-Based Clinical Decision Support: The Symptom Assessment and Management Intervention. JMIR Med Inform 2016 Nov 08;4(4):e36 [FREE Full text] [doi: 10.2196/medinform.5728] [Medline: 27826132]
29. Furberg RD, Williams P, Bagwell J, LaBresh K. A Mobile Clinical Decision Support Tool for Pediatric Cardiovascular Risk-Reduction Clinical Practice Guidelines: Development and Description. JMIR Mhealth Uhealth 2017 Mar 07;5(3):e29 [FREE Full text] [doi: 10.2196/mhealth.6291] [Medline: 28270384]
30. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
31. McCoy AB, Waitman LR, Lewis JB, Wright JA, Choma DP, Miller RA, et al. A framework for evaluating the appropriateness of clinical decision support alerts and responses. J Am Med Inform Assoc 2012;19(3):346-352 [FREE Full text] [doi: 10.1136/amiajnl-2011-000185] [Medline: 21849334]

Abbreviations

BPA: Best Practice Advisory
CDS: clinical decision support
EHR: electronic health record
GCS: Glasgow Coma Scale
IT: information technology
NIH: National Institutes of Health
PO: per os
SQL: structured query language
TDD: test–driven development

Edited by G Eysenbach; submitted 26.12.17; peer-reviewed by T Colicchio, D Sittig; comments to author 25.01.18; revised version received 22.03.18; accepted 23.03.18; published 13.04.18.

Please cite as:
Basit MA, Baldwin KL, Kannan V, Flahaven EL, Parks CJ, Ott JM, Willett DL
Agile Acceptance Test–Driven Development of Clinical Decision Support Advisories: Feasibility of Using Open Source Software
JMIR Med Inform 2018;6(2):e23
URL: http://medinform.jmir.org/2018/2/e23/
doi:10.2196/medinform.9679
PMID:29653922

Original Paper

Diabetes-Related Behavior Change Knowledge Transfer to Primary Care Practitioners and Patients: Implementation and Evaluation of a Digital Health Platform

Samina Abidi[1], MD, PhD; Michael Vallis[2], PhD; Helena Piccinini-Vallis[2], MD; Syed Ali Imran[3], MD; Syed Sibte Raza Abidi[4], MSc, PhD

[1]Medical Informatics Program, Department of Community Health and Epidemiology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada
[2]Department of Family Medicine, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada
[3]Division of Endocrinology and Metabolism, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada
[4]Knowledge Intensive Computing for Healthcare Enterprises Research Group, Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada

Corresponding Author:
Samina Abidi, MD, PhD
Medical Informatics Program
Department of Community Health and Epidemiology, Faculty of Medicine
Dalhousie University
Centre for Clinical Research
5790 University Avenue
Halifax, NS, B3H 1V7
Canada
Phone: 1 902 448 2447
Fax: 1 902 494 1597
Email: samina.abidi@dal.ca

Abstract

Background: Behavioral science is now being integrated into diabetes self-management interventions. However, the challenge that presents itself is how to translate these knowledge resources during care so that primary care practitioners can use them to offer evidence-informed behavior change support and diabetes management recommendations to patients with diabetes.

Objective: The aim of this study was to develop and evaluate a computerized decision support platform called "Diabetes Web-Centric Information and Support Environment" (DWISE) that assists primary care practitioners in applying standardized behavior change strategies and clinical practice guidelines–based recommendations to an individual patient and empower the patient with the skills and knowledge required to self-manage their diabetes through planned, personalized, and pervasive behavior change strategies.

Methods: A health care knowledge management approach is used to implement DWISE so that it features the following functionalities: (1) assessment of primary care practitioners' readiness to administer validated behavior change interventions to patients with diabetes; (2) educational support for primary care practitioners to help them offer behavior change interventions to patients; (3) access to evidence-based material, such as the Canadian Diabetes Association's (CDA) clinical practice guidelines, to primary care practitioners; (4) development of personalized patient self-management programs to help patients with diabetes achieve healthy behaviors to meet CDA targets for managing type 2 diabetes; (5) educational support for patients to help them achieve behavior change; and (6) monitoring of the patients' progress to assess their adherence to the behavior change program and motivating them to ensure compliance with their program. DWISE offers these functionalities through an interactive Web-based interface to primary care practitioners, whereas the patient's self-management program and associated behavior interventions are delivered through a mobile patient diary via mobile phones and tablets. DWISE has been tested for its usability, functionality, usefulness, and acceptance through a series of qualitative studies.

Results: For the primary care practitioner tool, most usability problems were associated with the navigation of the tool and the presentation, formatting, understandability, and suitability of the content. For the patient tool, most issues were related to the tool's screen layout, design features, understandability of the content, clarity of the labels used, and navigation across the tool. Facilitators and barriers to DWISE use in a shared decision-making environment have also been identified.

Conclusions: This work has provided a unique electronic health solution to translate complex health care knowledge in terms of easy-to-use, evidence-informed, point-of-care decision aids for primary care practitioners. Patients' feedback is now being used to make necessary modification to DWISE.

(JMIR Med Inform 2018;6(2):e25) doi:10.2196/medinform.9629

KEYWORDS

type 2 diabetes mellitus; self-management; health behavior; knowledge management; clinical decision support system

Introduction

Background

An estimated 9 million Canadians are living with diabetes, prediabetes, or undiagnosed diabetes [1]. Effective diabetes self-management, which relies on the behavior of an individual [2,3], shows great potential, as it can improve outcomes, decrease the risk of complications, and reduce diabetes-related hospitalizations and costs. Optimal diabetes control requires ongoing adherence to medication and diabetes care, self-monitoring of blood glucose, achieving a healthy weight, eating healthy, abstinence from smoking, moderate alcohol consumption, being physically active, and managing stress. The goal of diabetes self-management intervention is to support the individual to achieve positive behavior changes for achieving optimal diabetes control. Behavioral science is now being integrated into diabetes self-management interventions [2-4] to better educate and engage individuals in the self-management of their condition. In this regard, theory-driven, evidence-based behavior change approaches have been applied to (1) increase motivation to change when it is low, using the stages of change model [5,6], theory of planned behavior [7], social cognitive theory [8,9], and motivational interviewing techniques[10]; (2) support effective behavior change when motivation is present [11]; and (3) address emotional and relational barriers to behavior change [12,13]. Behavior change interventions have been developed and applied to chronic disease management [14-17], and more specifically to achieve diabetes control [18-20]. Diabetes management requires an interprofessional team effort, with the family physician as an initial and long-term health care provider, diabetes specialists providing therapeutic support, diabetes educators providing assistance to achieve diabetes control, behavior change experts influencing positive self-management behavior, and finally, with the patient as the most integral member of this team. Canadian Diabetes Association's (CDA) Clinical Practice Guideline (CPG) [21] also suggests an interdisciplinary team approach toward diabetes management. Generally, diabetes in Canada is managed by PCPs, with auxiliary support provided by nurses and dietitians qualified by the CDA as certified diabetes educators (CDEs) in diabetes management centers (DMCs). Patients are referred to a DMC by their PCP. At a DMC, patients are provided with self-management education and tools to help them self-manage their condition and associated risk factors. The CDEs work closely with PCPs.

Although PCPs are heavily involved in the long-term care of their patients with diabetes, studies have shown suboptimal and nonstandardized diabetes care at the primary care level [22-24]. Despite the availability of specialized behavior change interventions and evidence-based CPG on diabetes management, the challenge is how to translate these knowledge resources during care so that PCPs can easily use them to offer evidence-informed behavior change support and diabetes management recommendations to patients with diabetes. The Behavior Change Institute (BCI) [25] at Nova Scotia Health Authority in Halifax offers PCPs and CDEs with behavior change training and support to help them educate patients who require assistance in modifying unhealthy behaviors and need guidance to self-manage their chronic condition. However, it is also noted that because of resource constraints, there are limited competency-based behavioral support training opportunities available for PCPs [26-28], and likewise, there are limited opportunities for patients to access the services provided by DMCs [29]. Given the challenges faced by both PCPs and patients to access behavior change programs, we argue that it is prudent to leverage digital health technologies to (1) deliver to PCPs CDA's CPG-based diabetes care decision support as well as behavior change interventions planning support to help them manage both the clinical and behavioral aspects of diabetes control and (2) empower patients to better self-manage their diabetes through personalized behavior change interventions accessible to them by mobile technologies. To achieve these functional objectives, working in collaboration with BCI, we have developed computerized behavior change training modules for PCPs as well as patients with diabetes to improve diabetes control outcomes.

Objectives

In this paper, we present an innovative computerized decision support platform to (1) assist PCPs in administering evidence-based behavior change strategies and CPG-based recommendations for diabetes management and (2) empower patients with the skills and knowledge required to self-manage and monitor their diabetes through planned, personalized, and pervasive behavior change strategies. The key research tasks pursued in this project include (1) the development of a behavior change strategy based on evidence-based theories to better engage, empower, and inform the PCPs and their patients about behavior change strategies pertaining to diabetes control; (2) formulation of a comprehensive and validated knowledge base (in terms of a high-level behavior change ontology) that encapsulates semantic associations between multiple elements, that is, patient profile, CPG-derived diabetes management recommendations, and behavioral theory constructs that are coupled with behavior change strategies; (3) implementation of an integrated clinical decision support and behavior change intervention planning framework called Diabetes Web-centric Information and Support Environment (DWISE) that leverages semantic Web technologies to computerize behavioral and

clinical knowledge in terms of an ontological knowledge model and generate personalized behavior change strategies by reasoning over the computerized knowledge using the patient profile. DWISE can be accessed by PCPs via a secure Web interface, and patients can access it via the DWISE mobile app; (4) evaluation of DWISE in terms of its usefulness, usability, and functionality through qualitative studies involving PCPs and patients. In the subsequent sections, we discuss in detail the design, development, and evaluation of DWISE.

Problem Description and Solution Rationale

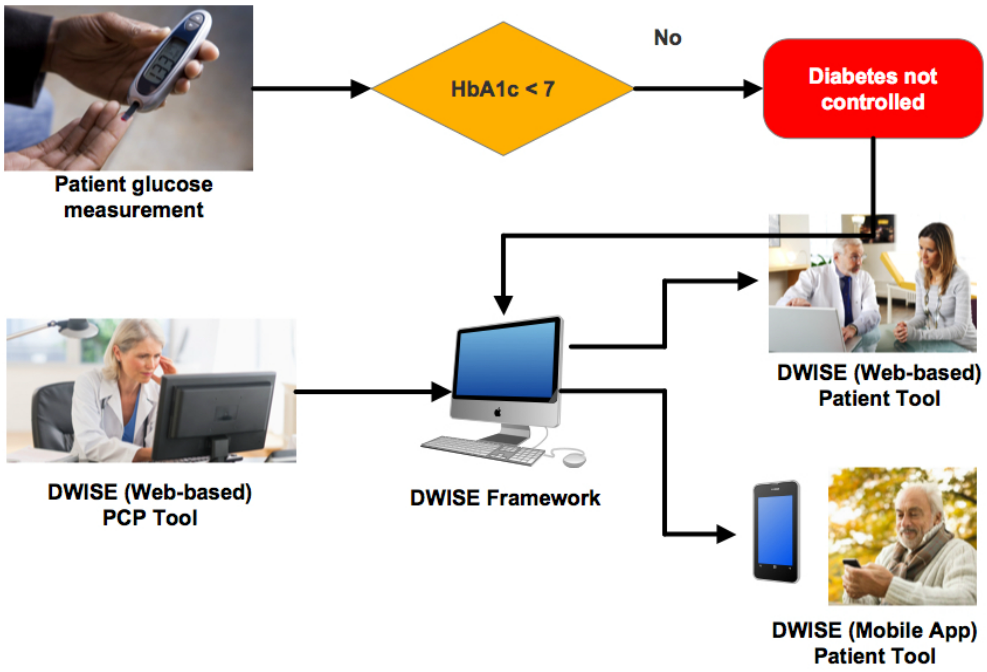
CDA's CPG [21] recommends that individuals with diabetes manage their disease with the help of an integrated diabetes health team while using a self-management model that incorporates knowledge and skills development coupled with cognitive behavioral interventions. Furthermore, it is recommended that individuals (and their families) with diabetes should be regularly screened for symptoms of psychological distress, and preventive interventions such as participative decision making, feedback, and psychological support should be incorporated within diabetes self-management interventions. Although DMCs exist across Canada, approximately 70% of individuals living with diabetes are unable to benefit from a DMC because of access limitations; rather, they may receive care from their PCP [22]. Several Canadian [22-24] and international studies [30] have found suboptimal management of type 2 diabetes in primary care settings, including suboptimal glycemic control [23,24] and failure to achieve CPG-recommended targets for glycated hemoglobin (HbA_{1c}) and low-density lipoprotein cholesterol in patients with type 2 diabetes [22,24,30]. On observation, the recommendations proposed by the CDA are currently not being implemented in the diabetes care process because of (1) lack of access to psychosocial resources within diabetes medical services where PCPs are not well equipped to manage behavior change in individuals with low motivation or who face psychosocial barriers to change and (2) lack of access to diabetes management CPG during care. Given the limited training opportunities to empower PCPs to administer behavior change interventions [26-28] and also limited patient access to the DMC and behavior change support [29], we argue that new knowledge transfer approaches need to be implemented to overcome this prevailing knowledge gap, and digital health technologies should be leveraged to administer behavior change programs that can potentially help to improve diabetes control outcomes.

To address the gaps in diabetes-related behavior change knowledge transfer, in this project, we demonstrate the applicability of digital health technologies to (1) provide decision support for PCPs to design and administer personalized behavior change strategies and (2) simultaneously provide motivational and educational support for patients with diabetes to self-manage their condition for improving diabetes control outcomes. In this regard, we present a knowledge management–based approach, together with its implementation and deployment, in terms of a DWISE that features the following functionalities: (1) assessment of PCPs' readiness to administer validated behavior change interventions to patients with diabetes; (2) educational support for PCPs to help them offer behavior change interventions to patients with diabetes; (3) access to evidence-based material, such as the CDA's CPG, to the PCPs; (4) development of personalized patient self-management programs to help patients with diabetes achieve healthy behaviors to meet CDA targets for managing type 2 diabetes; (5) educational support for patients to help them achieve behavior change; (6) monitoring the patients' progress in adhering to their behavior change program and motivating them to be in compliance with their program. DWISE offers these functionalities to PCPs through an interactive Web-based interface, whereas the patient's self-management program and associated behavior interventions are delivered through a mobile patient diary via mobile phones and tablets (Figure 1). A key feature of our solution is the incorporation of semantic Web-based knowledge modeling and execution technologies [31,32] that are applied to translate diabetes CPG and behavior change knowledge resources in terms of point-of-care decision support and mobile self-management support resources for PCPs and patients, respectively.

Diabetes Web-Centric Information and Support Environment Solution Approach

We contend that health education and support for chronic disease self-management should not just focus on changing the patient's awareness of the disease but should also help to empower the patient to make the right choices to achieve effective disease management via self-management support mechanisms. In this regard, our solution approach is to incorporate validated behavior change theories—in our case social cognition theory (SCT) [33]—to address an individual's self-efficacy expectations and perceived capabilities to perform self-care actions. Self-efficacy attainment has been shown to influence an individual's motivation, accomplishments, self-regulation, and efforts to perform self-care actions [33]. Patient education programs grounded in self-efficacy theory have been shown to enhance patient's adherence to self-care behavior, which in turn has been shown to improve clinical outcomes [34-37]. On the basis of the principles of SCT, our approach is to develop a specialized behavior change strategy that first assesses the PCP and patient's readiness to undertake behavior change interventions, and then, in response to their readiness levels, stipulate a personalized behavior change program. Although there are several existing self-management programs that target patients' behaviors, a unique aspect of our solution approach is the assessment of PCPs' readiness and in turn enhancement of their self-efficacy to administer behavior change counseling to patients who are facing psychosocial barriers to change behaviors that are affecting their condition. We argue that to implement an effective behavior change program at the primary care level, it is important to initially assess PCPs' readiness and self-efficacy to administer behavior change counseling and then provide them necessary training and decision support services so that they can effectively administer personalized behavior change counseling to their patients. Therefore, a unique aspect of our digital health solution approach is the provision of PCP-focused behavior change readiness assessment and decision support services that help PCPs to achieve self-efficacy in behavior change counseling.

Figure 1. Diabetes Web-Centric Information and Support Environment (DWISE) framework overview. HbA_{1c}: glycated hemoglobin.



The DWISE solution has 2 main elements: (1) behavior change strategies that are grounded in theoretical behavior change models [38,39] and (2) knowledge translation or transfer methods that are guided by a pragmatic health care knowledge management approach [40].

Our solution involves modeling and computerization of diabetes management–related clinical and behavior change knowledge sources using semantic Web methods and then using digital health technologies to design personalized behavior change interventions and deliver behavior change support using Web and mobile interfaces [41-44]. For knowledge modeling and computerization, we have pursued an ontology-based knowledge modeling approach [45] to develop a behavior change ontology (BCO) [41] that semantically represents and integrates the BCI's behavior change strategies, CDA's diabetes management CPG recommendations, and the SCT behavior change model; in this way, the BCO both computerizes and translates these knowledge sources in terms of actionable behavior change programs. We used the Web Ontology Language (OWL) [46], a computational logic-based language to develop the BCO. Our ontology-based behavior change modeling approach is novel and, in practice, we achieved an integrated knowledge model that entails (1) sections of the CDA's CPG pertaining to the management of glycemic control and (2) elements of BCI's behavior change strategy, including readiness to change assessments, motivational enhancement interventions, and self-efficacy attainment and self-management. The BCO is used to personalize behavior change strategies at 2 broad levels: (1) clinical level, where CPG-derived clinical variables are used to tailor most relevant recommendations for the given patient and (2) behavioral level, where the behavioral variables derived from the relevant behavioral models are used to personalize behavior change interventions.

Methods

Overview

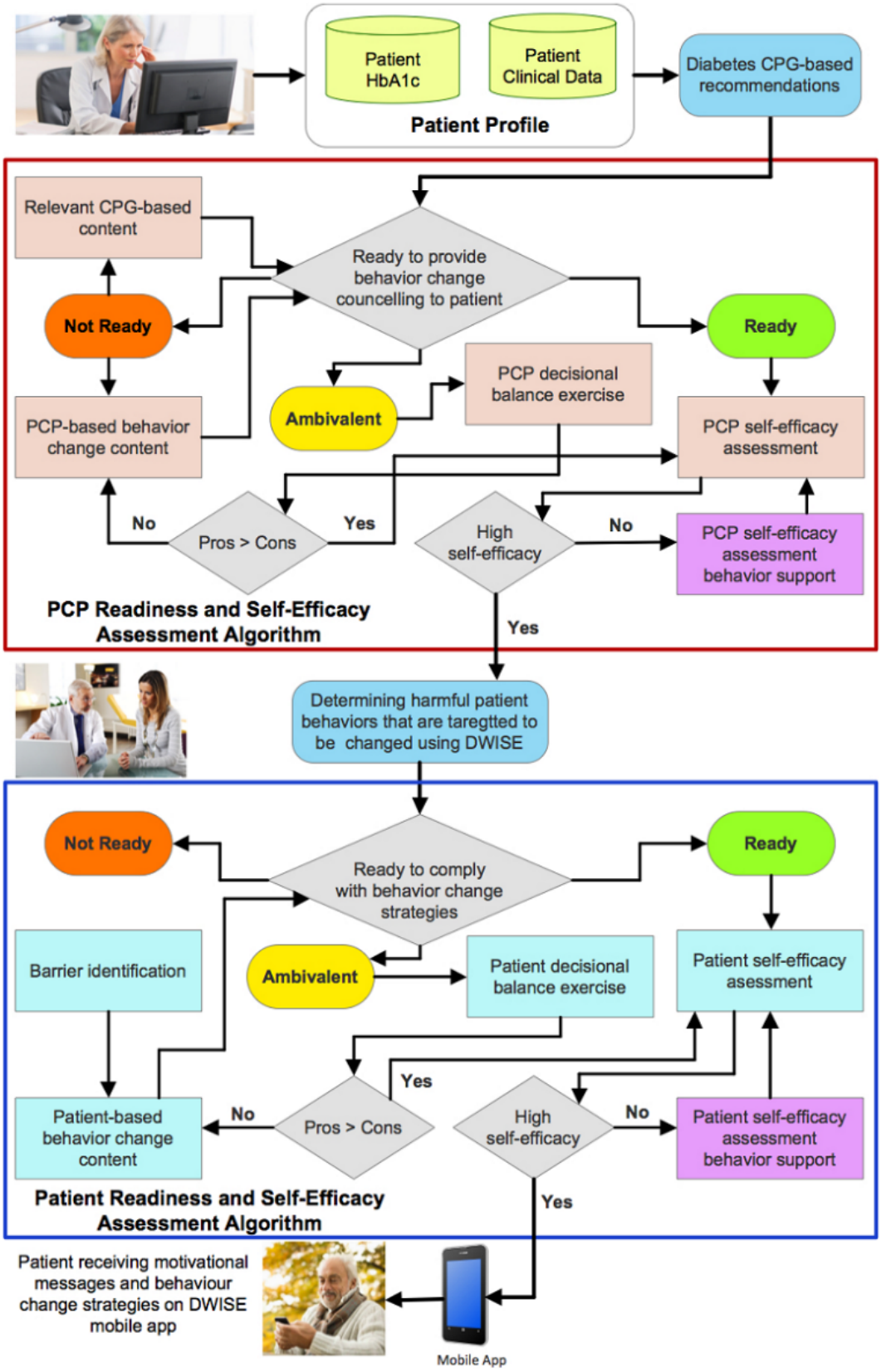
To develop DWISE and, in particular, its knowledge backbone (ie, the BCO), we used an ontology-based information system design framework called Methontology [47] that stipulates a cyclical ontology engineering approach, which involves experts to validate each iteration of the knowledge model. We discuss the different aspects of DWISE design and development in the following paragraphs.

Behavior Change Knowledge Modeling

The purpose of the behavior change knowledge modeling step is to develop specialized behavior change algorithms targeting the specific needs of both the PCPs and patients, which will be used to formulate theory-driven behavior change strategies. To develop behavior change algorithms based on our knowledge management approach, we abstracted and modeled key behavior change knowledge constructs from the available knowledge sources—the BCI strategies, diabetes CPG, and SCT model.

Knowledge modeling involved the abstraction of clinical and behavioral determinants from the paper-based knowledge sources that influence diabetes control outcomes. These determinants were then systematically linked and represented in terms of a rich multidimensional patient behavior profile. To enhance the ability of the patient's behavior profile to personalize a behavior change strategy, the behavioral determinants were represented using multiple levels or ranges to capture the nonlinear nature of behavior change among different individuals. We developed 2 high-level behavior change algorithms, one each for PCPs and patients (as shown in Figure 2).

Figure 2. High-level PCP and patient behavior change algorithm. CPG: clinical practice guidelines; ; HbA_{1c}: glycated hemoglobin; DWISE: Diabetes Web-Centric Information and Support Environment; PCP: primary care practitioner.



Each algorithm systematizes a variety of assessment tools based on behavioral determinants, range of PCP and patients' inputs (in terms of observations, goals, and preferences), behavior change strategy options, behavior change strategy elements, motivational messages, and education material. The algorithms are based on 3 behavior change models that are described in the following paragraphs.

A behavior change readiness assessment model was developed by our team at the BCI to assess the readiness levels of both PCPs and patients toward behavior change programs. When used in the PCP's behavior change strategy tool, our model assesses the readiness of a PCP to provide behavior counseling to help modify harmful behaviors in patients. When used in the patient's behavior change support tool, the model measures the readiness of a patient to comply with recommended self-management support strategies. The behavior change readiness assessment model uses a systematic questionnaire (with responses of "yes," "no," and "maybe") to categorize an individual into 3 stages of readiness, that is, ready, ambivalent, and not ready. As noted in other stage models, for instance, the transtheoretical model [48], the transition between readiness stages is nonlinear and is dependent on the levels of motivation and self-efficacy that can be improved by cognitive and behavioral therapies. To account for the nonlinear progression in behavior change, we included 2 additional behavior models—a decisional balance measure and a self-efficacy assessment model—to our behavior change readiness assessment model.

We used the decisional balance measure to determine an individual's perception about the expected benefits (pros) of modifying a behavior as opposed to the disadvantages or costs (cons) of this behavior change. An individual who is deemed ambivalent or not ready for behavior change programs is required to undergo a decisional balance assessment, which includes up to 5 pros and 5 cons that measure positive and negative perceptions of PCPs in administering self-management behavior change support, and that of patients in adopting self-management behaviors. Decisional balance assessment is repeated after educating the ambivalent or not ready individual about the benefits of behavior change programs.

After the decisional balance assessment where the individual has been assessed to be ready for behavior change programs, a self-efficacy assessment is performed to measure a "ready" PCP's degree of confidence that he or she can administer behavior change counseling, as well as a "ready" patient's degree of confidence in complying with behavioral change strategies.

In practice, the PCP behavior change algorithm initially performs behavior change readiness assessment, decisional balance assessment, and self-efficacy assessment, and in response to the assessment, a range of targeted educational material is provided to the PCP. During a patient encounter, the PCP behavior change algorithm supports the design of a personalized behavior change strategy by providing targeted CPG-based recommendations, most suitable behavior to be targeted for a patient, and the corresponding behavior change strategy that is tailored based on the Specific, Measurable, Action-Oriented, Relevant, Timely (SMART) goals jointly set by the PCP and the patient and the patient profile.

The patient's behavior change algorithm generates the patient support material based on the SMART behavior change goals to help the patient achieve the recommended HbA_{1c} level. The patient behavior change algorithm operationalizes the personalized behavior change strategy to generate and deliver motivational messages, educational material, and recommendations for overcoming barriers to change so that patients can achieve their SMART goals.

Ontology-Based Modeling of Behavior Change Knowledge

To formally represent the behavior change knowledge model (and the algorithms), we developed a high-level BCO [41] using ontology modularization principles [49,50] (see Figure 3), resulting in distinct yet interconnected knowledge modules representing the different aspects of behavior change programs. Ontology modularization approach reduces the complexity of developing a large multifaceted ontology, and we could handle the complexity of formally representing and integrating multiple behavior change models, assessment tools, behavior change strategies, and the CPG for diabetes management in a single, comprehensive behavior change knowledge model. We used the Methontology [47] methodology to develop the 2 distinct yet integrated knowledge modules within BCO, which are as follows (Figure 3):

- Information Personalization module that is used to create patient behavior change profiles, and it consists of 4 knowledge submodules:
  - Clinical Profile module represents clinical attributes derived from the CDA's CPG, pertinent to describing the patient's clinical medical profile with respect to diabetes control.
  - Readiness Assessment module represents the behavior change readiness assessment strategy for both PCPs and the patients, as developed by our team at BCI.
  - Decisional Balance Assessment module represents the positive and negative perceptions of "not ready" and "ambivalent" PCPs.
  - Self-Efficacy Assessment module represents the SCT-based self-efficacy assessments of PCPs and patients.

The patient behavior change profile is dynamically created when the information personalization module is executed using the patient's attributes.

- Domain Knowledge module that is used to represent the domain (ie, behavior change and self-management for diabetes control) knowledge, and it comprises 2 submodules:
  - Diabetes Domain module represents the evidence-based diabetes control recommendations as stipulated by CDA's CPG.
  - Self-Management Domain module represents SCT-based self-management knowledge, for example, barriers to diabetes self-management and behavior change, self-management and behavior change support materials and strategies, and SMART goal setting support.

Figure 3. Information personalization and domain knowledge elements of Behavior Change Ontology.

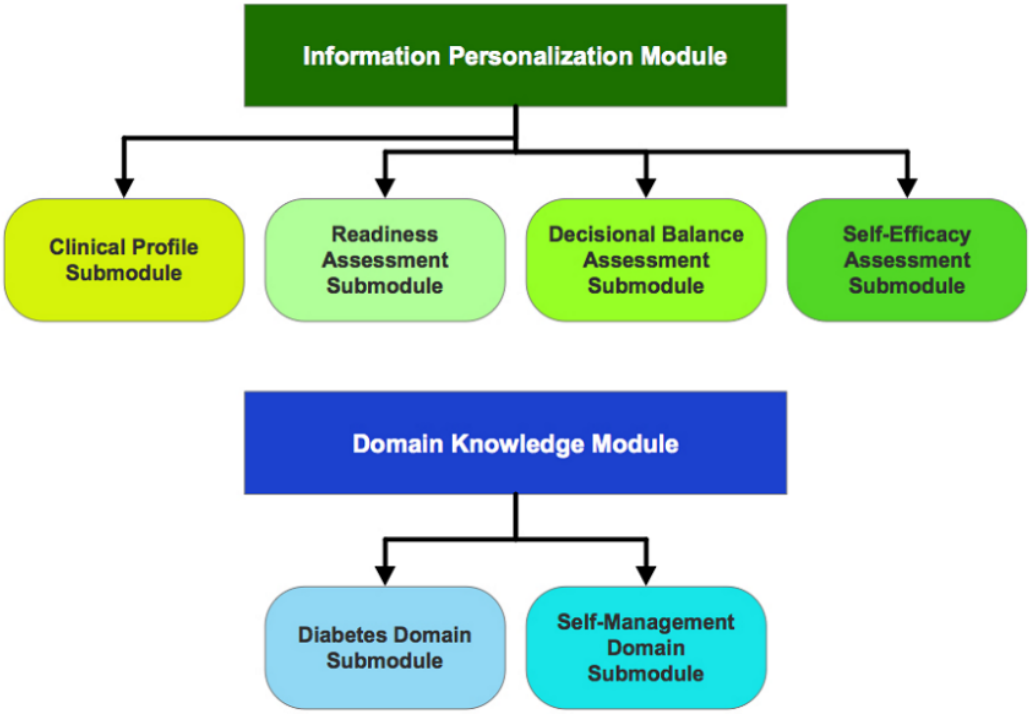
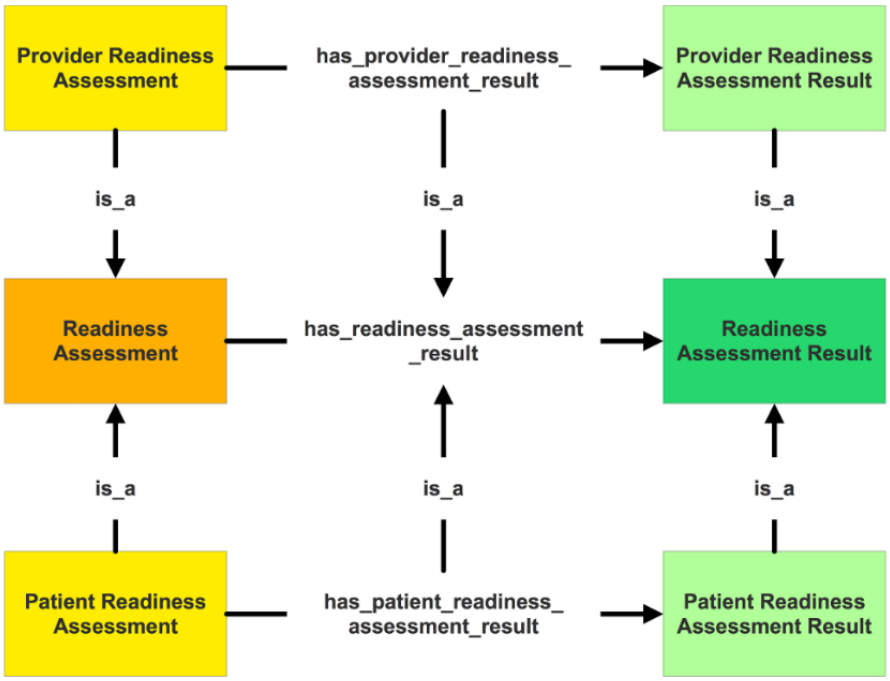


Figure 4. A subset of readiness assessment module in Behavior Change Ontology, depicting procedural relationships between classes "Readiness Assessment" and "Readiness Assessment Result.".



To develop the BCO, we used object properties to represent declarative and procedural relationships between classes. Axioms and rules were used to augment the procedural aspect of the BCO further. Figure 4 shows a subset of Readiness Assessment module in the ontology, depicting procedural relationships between the classes "Readiness Assessment" and "Readiness Assessment Result."

Information personalization submodules—readiness assessment, decisional balance assessment, and self-efficacy assessment—computerize assessment questionnaires to identify the patient's current behavioral predisposition. The questionnaires are represented as an object and the questions within them as their properties. The questions' responses such as values from 1 to 4 are represented as datatype properties, whereas questions that require a predefined statement as a response (eg, what is the highest level of education you have achieved?) are represented as object type properties. Property restrictions, such as cardinality restrictions, range (both for

object type and data type), and allowed values were used to ensure knowledge integrity. The Information Personalization module contains 10 classes (Barrier, Barrier to Change, Clinical Profile, Decisional Balance, Decisional Balance Result, Readiness Assessment, Readiness Assessment Result, Self-Efficacy Questionnaire, Self-Efficacy Result, and Prognostic Factors). In total, BCO represents 18 top-level classes, whereas the entire class hierarchy consists of 80 classes. There are 16 top-level object properties with 18 object subproperties and 7 top-level data type properties with 40 data type subproperties. Finally, BCO was instantiated using the content that was gathered and developed in the PCP and patient algorithms. We used Protégé 2000 knowledge acquisition tool (Stanford Center of Biomedical Informatics Research, Stanford University, Stanford, California) [51] as the ontology editor using the OWL [46].

BCO was evaluated for (1) knowledge accuracy and utility by 3 domain experts (psychologist, endocrinologist, and a family physician) and (2) semantic accuracy to ensure logical consistency. Although experts generally agreed with the representations, they nevertheless suggested a few improvements, for example, better definition of the class Clinical_Profile in terms of its properties and relationship with the class Clinical_Parameter. Changes were made to BCO in terms of its concept description, relationships, and constraints after each evaluation event in response to the experts' comments. The technical evaluation of BCO was carried out in accordance with the criteria suggested by Gomez Perez [52], which include the 3 Cs: Consistency, Completeness, and Conciseness. FaCT++ [53], an open-source OWL DL reasoner, was used to perform the subsumption tests on BCO to establish its concept satisfiability and consistency. Fact++ was also used to compute the inferred class hierarchy and identify redundant arcs between the classes. Our classification tests did not indicate any redundant arcs in BCO; therefore, it is concluded that the asserted hierarchy is similar to the inferred hierarchy. Finally, in terms of the evaluation of the necessary and sufficient conditions of a predicate, domain and range of relations, and generalization and specialization of classes, it was noted that BCO demonstrated representational capacity to adequately instantiate all the domain concepts and relationships captured in the behavior change algorithms and CPG. The complete BCO can be accessed at the Github website [54].

Implementation of Diabetes Web-Centric Information and Support Environment

A prototype of DWISE framework consisting of the Web-based PCP interface and a mobile app for patients has been implemented. The overall technical architecture of DWISE framework is illustrated in Figure 5. The Web application is written in Java and Vaadin, and information is stored in a relational database on a secure centralized server. Within DWISE, BCO serves as the main knowledge resource—an open-source Java library for OWL, and RDF is used to read and manipulate the domain knowledge contained in BCO's OWL files. The main BCO elements—resources, properties, and property values—are used to model the temporal relations inherent in DWISE's information flow. The decision rules are translated into JENA rule syntax, which are input into the JENA reasoning system; JENA uses the rules and temporal relations specified in the BCO to integrate ontological modules during execution of the BCO to infer knowledge-based decision support and behavior change strategy planning based on patient profile. We have leveraged our existing work in semantic Web technologies for knowledge representation [55-60], and we exploit our OWL-based reasoning methods [61] to infer a dynamically generated patient-specific behavior change strategy.

Diabetes Web-Centric Information and Support Environment Web-Based Decision Support System for Primary Care Practitioners

DWISE decision support framework consists of a PCP and a patient tool. DWISE PCP decision support tool (Figure 6) provides recommendations from the CDA'S CPG (eg, most appropriate target HbA_{1c} for a patient) that are tailored toward the clinical profile of the patient for whom behavior change support is being sought. In addition, the PCP decision support tool also assesses the readiness and self-efficacy of the PCP to administer behavior change counseling to help patients achieve the CPG-stipulated diabetes control targets by changing harmful behaviors. For PCPs who are deemed as "ready" to administer behavior change counseling, DWISE enables the PCP to engage their patients in a shared decision-making setting to develop a personalized behavior change strategy, akin to behavior change consultations performed at the BCI. The strategy is tailored toward a patient's behavioral profile that is formulated through a series of behavioral assessment exercises administered through DWISE in a shared decision-making setting.

Diabetes Web-Centric Information and Support Environment Patient Support Mobile App

DWISE patient support tool is implemented as both a Web-based system and a mobile app, with the functionality to deliver the following self-management support to patients: (1) behavior change strategies such as goal setting, behavior shaping, stimulus control, and reinforcement management; (2) context-aware motivational and behavior change educational messages; and (3) communication with care providers. The DWISE mobile app (Figure 7) is designed to support the patient to enact the self-management plan that he or she has formulated with the PCP. The DWISE mobile app features are (1) patient diaries for capturing vitals, diet, exercise, stress, and mood; (2) proactive alerts to underline stimulus control and reinforcement management; (3) context-aware motivational and behavior change educational messages and reminders to help the patient adhere to the self-management schedule; and (4) communication with the care providers. The DWISE mobile app has been developed for the Android platform using necessary data security and privacy regulations, with provisions for a future iOS-based app implementation.

Evaluation of Diabetes Web-Centric Information and Support Environment

Two qualitative studies that incorporated Think Aloud Protocol (TAP) method [62,63] were conducted to evaluate DWISE's usability by PCPs and patients. In addition, a focus group study was conducted to examine the perspectives of the PCPs and

patients about the usefulness of DWISE for diabetes self-management in shared decision-making environment.

Study Design

For the usability study, we used a cognitive and usability engineering framework [62] to establish whether DWISE meets the functional goals, content suitability or comprehensiveness, and usability needs of PCPs and patients. Furthermore, our intent was to identify the usability- and content-related issues that need to be resolved in the next version of DWISE and to receive the end-user feedback.

After ethics approval, we randomly recruited 10 PCPs (4 family physicians and 6 CDEs) for the PCP study and 11 patients for the patient study. The sample size estimate is based on the evidence that 70% of severe usability problems can be uncovered within the first 5 users, and up to 85% by the eighth user [62]. The PCP study included PCPs, family medicine residents, and CDEs licensed in Canada and practicing in Halifax. Adult patients with diabetes who were residing in Halifax were included in the patient study. Exclusion criteria included being on the DWISE research team and lack of proficiency in English (because DWISE is currently developed in the English language). Both groups were asked to complete a demographic and background questionnaire. Each PCP/CDE was provided with 3 standard case scenarios that simulated 3 different patients to interact with DWISE, thus yielding 30 TAPs from 10 PCPs.

Figure 5. Diabetes Web-Centric Information and Support Environment (DWISE) technical architecture. BCI: Behavior Change Institute; CDA: Canadian Diabetes Association; CPG: clinical practice guidelines; PCP: primary care practitioner.

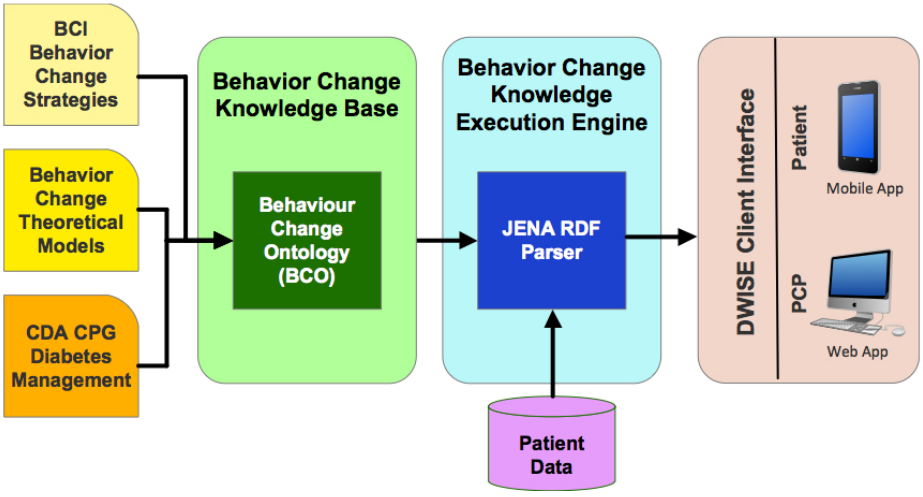


Figure 6. Tool for primary care practitioners assessing self-efficacy of the family physicians/certified diabetes educators.

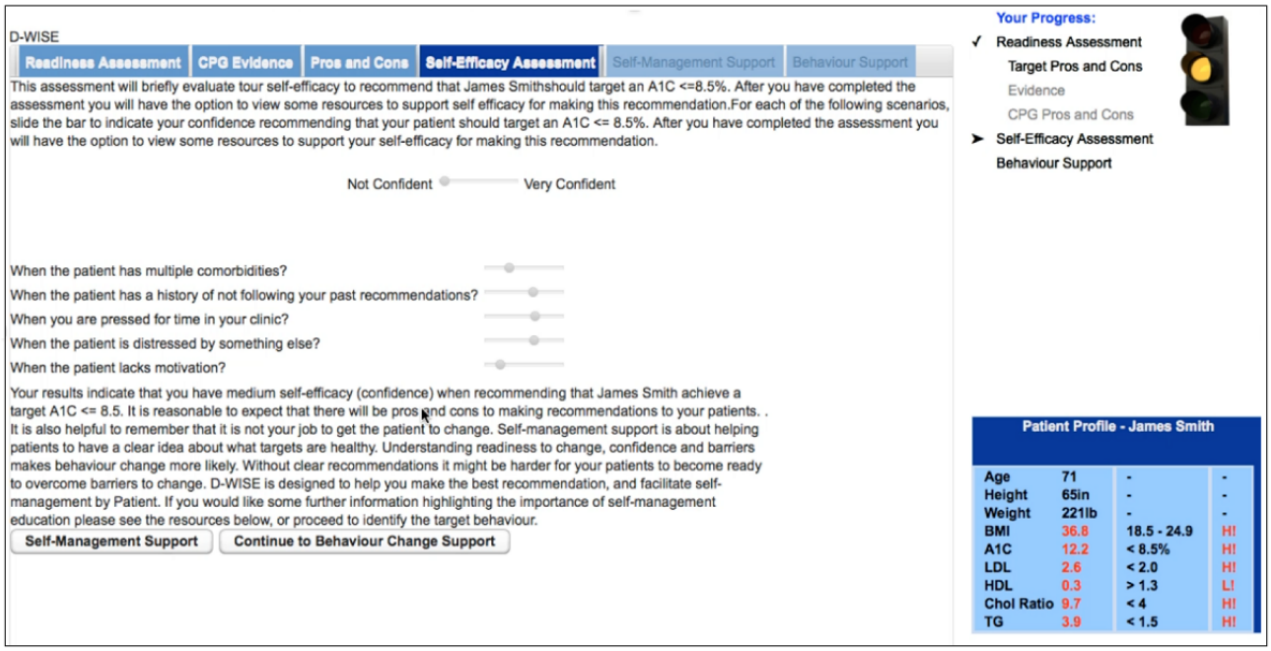
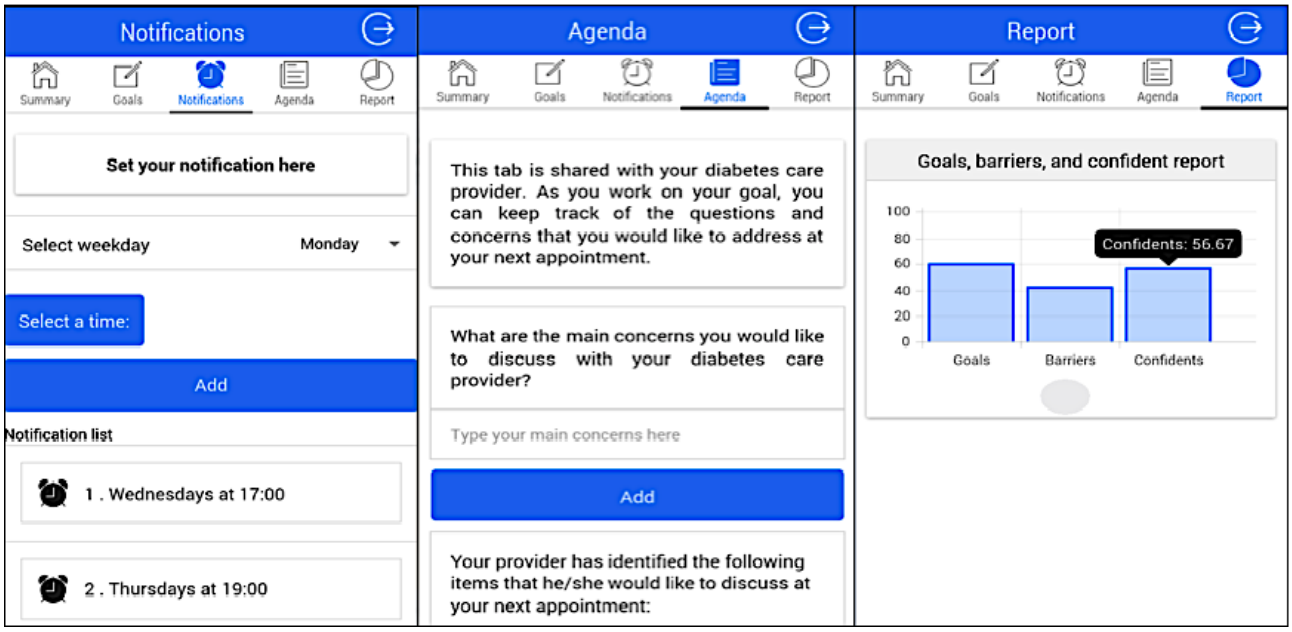


Figure 7. Screenshots from Diabetes Web-Centric Information and Support Environment (DWISE) mobile apps depicting functionalities such as scheduling, notifications, barrier identification, and feedback.



Textbox 1. A priori categories for focus group data analysis.

1. Barriers to Diabetes Web-Centric Information and Support Environment (DWISE) usage
2. Facilitators to DWISE usage
3. Patient behavior change
4. Patient empowerment and education
5. Patient autonomy and preference
6. Patient-provider communication
7. Encounter-related issues
8. Professional roles and responsibilities
9. Usability- or acceptability-related issues

Patients were presented with a standard behavioral recommendation, that is, physical activity that they have hypothetically agreed to pursue in concert with their PCP/CDE. The patients then defined a specific target behavior, assessed their readiness, and received support in setting a goal for behavior change. This yielded 11 TAPs from 11 patients. During the interactions with DWISE, each participant was encouraged to think aloud. Participants' screen activity and audio were recorded using the QuickTime player. The usability study design is presented in the study by Abidi et al [64] in detail.

For the focus group study, a purposive sampling strategy was used to recruit 4 patients and 3 PCPs, after acquiring the ethics approval. The purpose of this study was to engage patients and PCPs in shared decision-making environment to elicit (1) initial impression about the DWISE system; (2) advantages and disadvantages of DWISE in providing CPG-based recommendation and behavior change strategies to the PCPs and patients; (3) potential impact of DWISE on patient-provider communication and relationship around diabetes-related behavior change; and (4) suggestions to improve DWISE. The session was audio-recorded and transcribed verbatim and was supplemented by field notes, sketches, and observation logs. The experts on the team prepared a semistructured moderator's guide based on their clinical and research experience and the review of the related literature. The guide included open-ended questions and problem-based representative scenarios related to various self-management processes to stimulate conversations in case of unresponsive participants. Content validity of the guide was established by review of the literature on diabetes self-management in populations that are culturally and socioeconomically similar to the population of interest. Further validity was established through critique, change, and consensus of the expert research team members.

Data Analysis

Computer screen activity, audio files transcribed verbatim, field notes, and observation logs were analyzed in the ATLAS.Ti software (ATLAS.ti Scientific Software Development GmbH) using inductive thematic coding [65] method. Ten PCPs with 3 case scenarios yielded 30 TAPs, and 11 patients with 1 case scenario generated 11 TAPs. Segments of interesting audio and

screen activity data were selected as quotations, which were professed as unit of analysis. Open codes were directly applied to the quotations. ATLAS.Ti also calculated code frequency, that is, number of quotations to which a particular code is applied. Large numbers of quotations associated with a code indicate strong evidence found for this code, which in turn endorses the "groundness" of that code in the data. Axial coding was then applied to draw categories from the open codes based on commonality between them. To perform thematic analysis on the focus group data, a priori categories (Textbox 1) based on the open-ended questions in the moderator's guide were established. During data analysis, the open codes assigned to the quotations were classified as axial codes based on their commonality. The axial codes were constantly compared against a priori categories listed in Textbox 1 and assigned to one or more of these categories. Validation of the identified open and axial codes was performed by continual referral back to the original computer screen activity, audio files, transcripts, and observational notes. Furthermore, another researcher on the team reviewed the data, so that any conflicts or discrepancies were resolved through discussion and consensus before the codes were finalized.

Results

Participants' Demographics and Background

The background and demographic information of PCPs and patients, which was collected for the TAP study, is presented in detail in the study by Abidi et al [64]. In general, both the PCP and patient samples were biased toward women and those comfortable with computers. The spread of experience for PCPs was uniform. Although all of them used electronic medical records, they were not particularly experienced with respect to the use of decision support systems. Their comfort level with administering behavior change strategies in terms of experience and confidence was found to be variable. Median age of patients was 52 years. The patient sample was slightly more educated than the general population and was predominantly urban. All patients had previous behavior change experience and generally were somewhat confident in using behavior change strategies.

The 7 participants for the focus group study included 3 CDEs and 4 patients. All 3 CDEs were females, and of 4 patients, 3 were females, and 1 was male. Ages for CDEs ranged from 29 to 55 years, and patients were aged between 49 and 64 years. All CDEs worked at DMCs. CDEs had a median of 11 years of experience (range 3-19 years). Patients had diabetes for a median of 13.5 years (range 2-25 years).

Qualitative Results From Primary Practitioner Tool Study

In total, 31 independent open codes based on usability issues were identified in TAPs by 30 PCPs. The detailed results are presented in the study by Abidi et al [64]. Several of these codes occurred more frequently in the data, for example, "Need more patient information for pros and cons" occurred in 19 quotations and "Need more information for readiness assessment" occurred in 11 quotations. In total, 17 themes of axial categories of usability issues emerged based on the commonalities noted in the open codes (Figure 8). For example, the axial category "Navigation Problems and Lack of Flexibility" (Figure 8) had 8 codes, and one of the codes in this category, that is, "Navigation Problem-Unsure when trying to get back to pros and cons after looking at CPG evidence" occurred in 2 quotations. Example of one such quotation is "Is readiness assessment the one where I came from...so should I go there."

Qualitative Results From Patient Tool Study

Patients' TAPs yielded 17 open codes and 9 axial categories of usability issues. The detailed results are presented in study by Abidi et al [64]. Figure 9 shows the axial categories of usability problems with the number of codes in each axial category. Of 17 open codes, most critical codes include "Unsure of goal setting data entry field," which occurred in 11 quotations; "Sliding bar problems,"which occurred in 7 quotations; and "Problems with scrolling," which occurred in 6 quotations. An exemplar quotation that contains the code "unsure of goal setting data entry field" is as follows: "I have entered my goal and now it is asking me to be specific...so it should already be somewhere..."

Figure 8. Axial codes in a primary care practitioner;'s Think Aloud Protocol.

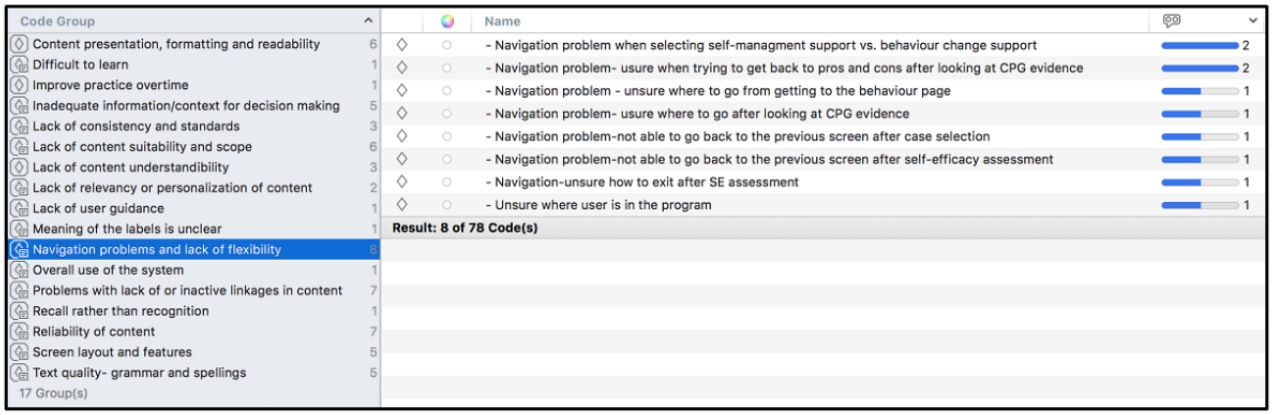
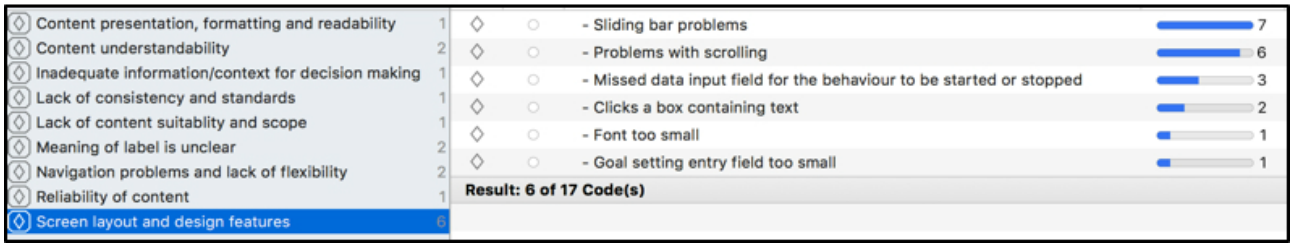


Figure 9. Axial category "Screen layout and design features" contain 6 open codes.



Each axial code represents areas for possible enhancement needed in DWISE content, screen layout and features, design, and other usability features. Sorting by theme and related codes and their frequency gives a detailed picture of the specific improvements and modifications that are needed before DWISE can be tested for its efficacy. The criticality of usability problems is based on the number of open codes in each category and the frequency of these codes in the quotations.

We considered the criticality of a code and the axial category to improve the design of DWISE. Any open code that occurred in a single quotation was discarded. In general, for the PCP tool, most problems were associated with the navigation of the tool and the presentation, formatting, understandability, and suitability of the content in the tool. For the patient support tool, most issues were related to the tool's screen layout and design features, understandability of the content, clarity of the labels used, and navigation across the tool. We used this feedback from the qualitative study to modify DWISE in terms of its user interface design and its content.

Salient Findings From the Focus Group Study

In all, 73 open codes were discovered in data that were classified into 27 independent axial codes. Table 1 shows the a priori categories and axial codes assigned to these categories. The number in parenthesis next to each axial code indicates the number of open codes contained in each axial code, thereby indicating groundness of each axial code in the data. These categories are explained in an integrated manner, that is, from both the patients' and PCPs' perspective.

Facilitators to Diabetes Web-Centric Information and Support Environment Usage

The PCPs appreciated that DWISE could be used as a teaching tool to teach diabetes-related self-management skills to their patients:

> *I try to guide my patients towards diabetes related self-management resources available...but really, there are silos of teaching that affects one's ability to learn. DWISE is good...it is more comprehensive and is more relevant to a patient's needs...this can help deal with these problems...I believe this should be accessible to most patients.*

The participants felt that although there are many diabetes educational resources, they want DWISE type apps that consider a patient's personal preferences and psychosocial concerns, when designing self-management strategies. Both patients and PCPs felt that there is more need for information about the psychological issues for patients with diabetes:

> *There should be apps that talk more about things like...distress, depression and psychology...I mean diabetes is hard, sometimes we are distressed and we need information and we rely a lot on Internet...*

Finally, patients who were more engaged in their self-management felt that DWISE is an ideal tool for them, and they believed that using DWISE will be easier for them. Participants overwhelmingly stated that they would like some kind of technical support, such as online or face-to-face sessions on DWISE, to teach them how to use DWISE. They suggested that there should be a helpdesk or other resources to help them troubleshoot to facilitate use of DWISE.

Barriers to Diabetes Web-Centric Information and Support Environment Usage

From the PCPs' perspective, time constraints were deemed to be a key barrier in the use of DWISE. PCPs felt that inclusion of DWISE-based intervention might not be feasible during the patient encounter because of the limited time that they have with their patients:

> *I love it...but can I do a good job with it? How can I incorporate this within the time restrictions?...at times it may not be conducive to my schedule...a patient gets just 15 minutes with the provider.*

PCPs were also worried about liability related issues:

> *Ok, suppose I am using this app with my patients, what if I missed something or I fail to do what is expected of me? Would I be liable...what will be the impact?*

PCPs also stated that including DWISE in their practice might result in additional work for them. Patients were worried that if they failed to achieve the goals that they have set through DWISE, they might lose respect in the eyes of the PCP, or disappoint them, and might feel burdened or stressed:

> *I mean respect is a two-way street...what if I don't meet that goal...what would my doctor think about me?*

Patients also stressed on preference for direct patient-provider contact:

> *Sometimes I just want to talk during an appointment with my doctors...maybe I don't want to talk through an app during this time.*

Table 1. Axial codes in each category. DWISE: Diabetes Web-Centric Information and Support Environment; PCP: primary care practitioner.

A priori category	Axial codes (number of open codes contained in each axial category)
Facilitators to DWISE usage	Technical support to facilitate DWISE usage (10)
	Teaching tool for patients (5)
	Need for personalized diabetes self-management apps (4)
	Need for information about psychological issues (2)
	Compliance with DWISE easier for engaged patients (1)
Barriers to DWISE usage	Practicality of DWISE due to PCP time constraints (4)
	Impact on patients who fail to achieve DWISE set goals (4)
	Age-related suitability (4)
	Practicality of DWISE because of technically challenged users (3)
	Preference for direct patient-PCP contact (3)
	Additional work (1)
	Liability-related issues (1)
Patient self-management	Potential to improve self-management and monitoring (3)
	Potential to modify behavior (1)
Patient education	Teaching tool for patients (5)
	Potential to improve patient awareness of disease (3)
Patient autonomy and preference	Patient autonomy in choosing self-management support delivery method (4)
	Power dynamic between patient and provider (2)
	Potential to improve patient empowerment (3)
Patient-provider communication	Insight into patient's self-management practices (2)
	Potential to improve patient-provider communication (3)
	Preference for direct patient-provider contact (3)
Encounter-related issues	Impact on patient provider encounter (3)
	Practicality of DWISE due to PCP time constraints (4)
Professional roles and responsibilities	Professional roles and responsibilities around DWISE usage (2)
	Additional work (1)
Usability- or acceptability-related issues	Reminders to improve usability (1)
	System feedback (2)
	Information presentation in DWISE (2)
	Integration with other devices (2)
	Need for personal features in DWISE (3)

Both PCPs and patients felt that technology ineptness might be a deterrent to their use of DWISE:

> *One of my colleague is not tech savvy...there might be other providers like her. How can these people benefit from DWISE?...would they be interested?* [PCP]

One of the patients said:

> *I am not technologically adept, these are new and exciting...I like help with managing my diabetes...but there might be big learning curve for me.* [Patient]

Patient Self-Management

Participants felt that DWISE has the potential to improve diabetes self-management, especially given that mobile phones are ubiquitous and self-management plans formulated through DWISE can easily be integrated into the patients' lives. Participants indicated that DWISE has the potential to improve diabetes-related monitoring:

> *Phone is ubiquitous, so more opportunities. I love apps for recording and monitoring...this can help me monitor my sugar.* [Patient]

One PCP remarked:

> *It helps me gain more information about diabetes-related behavior change and about my patient and both my patient and I can see if my patient is on the right track...we will have something to talk about next time we meet.* [PCP]

Patient Education

Participants felt that DWISE may help improve patient's awareness of the disease and can be used as a teaching tool for patients:

> *DWISE makes me more aware...more informed...I feel like I want to know more so that I can better take care of myself.*

Patient Autonomy and Preference

Patients felt that they should have autonomy in choosing self-management support delivery method. One patient said:

> *I don't believe that one size fits all...it is good to have platforms like apps...DWISE is easily available...it should not be made mandatory for every patient...I mean it has to be my choice.*

A patient also stated that DWISE has potential to improve power dynamics between the patient and PCP and help patients gain more control over their diabetes management:

> *I feel balance of power is always in favour of my doctor...it's not bad...but I like to be more involved...make decisions that fits my life...DWISE can give me more control.*

In general, participants felt that using a tool such as DWISE might make them feel more empowered to self-manage their condition.

Patient-Provider Communication

Although appreciating that DWISE has the potential to provide better insights into their patients' self-management practices, PCPs felt that DWISE could also improve communication between patients and PCPs around diabetes-related self-management. One of the PCPs said:

> *When a patient is first diagnosed with diabetes...DWISE can be a good avenue for discussion...about how a patient is feeling, what is it they want...how can they fit the self-management in their lives.* [PCP]

Patients also felt that DWISE could potentially help them to communicate personal issues that might affect their self-management practices and that otherwise would not come up during an appointment:

> *Doctors don't live with diabetes...I live with diabetes...I have lived with diabetes for so long...this type of technology and apps can support me to better communicate with my doctor...what I am going through...why I am not able to follow proper diet or...not exercising...*

Encounter-Related Issues

Although some PCPs expressed that it might not be practical to use DWISE during the encounter because of PCPs' time constraints, other PCPs and patients expressed that DWISE can have a positive impact on the patient-PCP encounter in terms of shared decision making around the setting of SMART goals. PCPs underscored that a patient might be more prepared during the encounter:

> *Every patient is different...and self-management requirements vary so much...so patients coming prepared will be so good for the appointment...I think appointment time will be better spent.*

Patients expressed that they will be more motivated to comply with plan set through DWISE to have a meaningful encounter:

> *There are higher problems that are not in my control...that might mix the schedule...but I will still try to do this or change it to have a better appointment...I'll go to the appointment with something...*

Professional Roles and Responsibilities

PCPs were unsure about the professional roles and responsibilities pertaining to the usage of DWISE in a clinical setting. They wondered how doctors, CDEs, and nurses would coordinate and collaborate to ensure that a tool such as DWISE can be used effectively:

> *How would this work...I mean how do we collaborate...should this be administered through a doctor or a nurse educator...who would monitor.*

Usability- or Acceptability-Related Issues

Finally, participants offered some feedback regarding issues related to the usability and acceptability of DWISE. Although some participants suggested that there should be more reminders to help them comply with self-management plans and upcoming activities, others suggested that a user should be able to disable the reminder when he or she feels like. Participants commented on the information presentation in DWISE and suggested that there should be a better layout, too much text should be avoided, and it should be replaced by user-friendly features such as pictures and figures. They suggested that a good feature would be to have a space for users to type their notes in free text. Participants also expressed their desire to have some personal features included in the app, such as provision to include their "profile picture," "personal profile information," and "personal diabetes story." Participants further suggested that DWISE should be integrated with other data collection devices, such as Fitbit, smart watch, and so on.

Discussion

Strengths

Digital health technologies have been effectively used for health information collection, information utilization, and sharing solutions. In this study, we have demonstrated that digital health applications can effectively and efficiently incorporate evidence-based health care knowledge to provide

evidence-informed decisions or recommendations to support both health care professionals and patients. This is an ongoing work and DWISE is a proof of concept. Nevertheless, this work has provided a unique digital health solution to translate complex health care knowledge, that is, guidelines, clinical workflows, behavior models, educational content, and long-term care plans, in terms of easy-to-use, evidence-informed, point-of-care decision aids for both PCPs and patients. From a clinical perspective, the contribution of this research is the translation of specialized behavior change knowledge to family physicians and diabetes educators, thus enabling them to offer behavior change interventions to a larger population of patients with diabetes—at present, only one-third of Canadians with diabetes receive diabetes educational programs [66]. From the patients' perspective, the contribution is a self-management program that engages and empowers them to manage their condition in a home-based and primary care setting as opposed to relying on visits to specialist clinics. From a health professional's perspective, this work is the first attempt to empower and engage PCPs to administer behavior change counseling, and it provides a comprehensive readiness assessment and educational framework to educate PCPs about how to perform standardized behavior change counseling. The provision of PCPs facing behavior change support tools is a novel initiative and can potentially improve patients' access to professional behavior change counseling.

A unique aspect of this research is the integration of paper-based medical knowledge, behavior change models, health care knowledge management methods, and mobile technologies to develop "intelligent and adaptive" mobile patient-centered solutions that are customizable to specific care contexts, users' knowledge, and interests. The project has contributed a generic digital health strategy and technology, based on theoretical models that can be applied to a range of medical conditions, to deliver intelligent and ubiquitous health educational and decision aids. In the long term, we plan to extend the research to other chronic diseases where we will account for different disease-specific factors pertaining to the personalization of behavior change strategies. In the medium term, we would augment the research scope to incorporate other related metabolic conditions that are characterized by hyperglycemia such as prediabetes.

Limitations

The 3 qualitative studies included small sample sizes. However, for usability studies, a small number of participants are deemed sufficient for determining the major usability issues [61,62]. Through our focus group study, our aim was to better understand the underlying barriers and facilitators to the use of DWISE for diabetes self-management in a shared decision-making setting, its impact on patient encounter and experiences, and its role in patient-provider communication. We do not know to what extent the sample we tested is representative; it is likely that we had a sample that is more favorable toward a digital health solution, as the participants volunteered to take part in this study. We sought the perspectives of a small number of participants after a 15-min demonstration of the system—we realize that maybe participants might have needed more time to adjust to DWISE because it was a new technology for them. As such, it is possible that participants might have missed or misinterpreted some of the features and functionalities offered by DWISE, which may have resulted in some of the responses being biased. Nevertheless, we draw confidence from the study methodology that ensured 2 researchers were present throughout the focus group study to provide clarifications and to alleviate any misconceptions participants may have had about the functionality of DWISE. We realize that this is a pilot study with a small sample size. Moreover, the PCP participants are mainly represented by the CDEs, which limits the generalizability of the study. Nevertheless, we are taking the feedback generated through these studies to improve the design, content, usability, and usefulness of DWISE framework. In the next stage, we plan to clinically evaluate DWISE for its effectiveness and safety in primary care settings, with the intent to disseminate it across the province of Nova Scotia.

Conclusions

In this paper, we have presented a digital health solution to translate complex health care knowledge, that is, guidelines, clinical workflows, behavior change models, educational content, and long-term care plans, in terms of easy-to-use, evidence-informed, point-of-care decision aids for both PCPs and patients. The knowledge modeling methods and decision support technologies being developed are both scalable and generic in nature, such that they can be readily applied to computerize CPG for other chronic diseases to develop low-cost decision support aids that can standardize the care of chronic diseases and comorbidities at the primary care level.

DWISE has been evaluated for usability, functionality, usefulness, and acceptance in a shared decision-making environment through a series of qualitative studies. In general, for the PCP decision support tool, most usability problems were associated with the navigation of the tool and the presentation, formatting, understandability, and suitability of the content in the tool. For the patient support tool, most usability issues that were raised were related to the tool's screen layout and design features, understandability of the content, clarity of the labels used, and navigation across the tool. With regard to the usefulness of DWISE in a shared decision-making environment, the most significant barrier from the PCPs' perspective is the limited time PCPs have during an encounter and, and from patients' perspective, the concern is the fear of failure to accomplish their goals to achieve diabetes control through behavior change interventions. In terms of facilitators, PCPs identified the potential of DWISE as a teaching tool for their patients, and the patients appreciated that DWISE provides personalized information especially on psychological issues that could be very useful to them. In general, participants felt that provision of technical support, especially to the elderly users and those who are not proficient in technology will facilitate the use of DWISE. Patients preferred that DWISE should not be made mandatory and should not completely replace the direct interactions with the PCPs, rather should be regarded as an additional support mechanism. Patients felt that DWISE may help them gain more control over their diabetes management, whereas PCPs suggested that it could assist them to gain more insight into a patient's self-management practices. PCPs seemed unsure about their respective roles and

responsibilities around DWISE usage. The results of these studies were used to guide the modification of DWISE in terms of its functionalities, screen layout and navigation, and content.

In conclusion, we contend that digital health technology, such as DWISE, that integrates a patient's (clinical and behavioral) profile with CPG-based best evidence and SCT-based behavior change theories, when used in a shared decision-making environment, has the potential to improve self-management and increase sense of collaboration and trust in the care process. Our finding suggests a dynamic interplay between patients, PCPs, as well as systemic and technology factors in terms of the operationalization of the DWISE framework for diabetes management. However, we also believe that the implementation of an integrated framework such as DWISE, in a shared decision-making clinical environment, requires additional time for the technology to mature, technical innovation, organizational support for technology uptake, and a clear definition of professional roles and responsibilities.

Acknowledgments

The authors acknowledge the support of Canadian Institute of Health Research. The project was supported through an eHealth Innovation grant.

References

1. Diabetes Canada. Diabetes. About diabetes URL: http://www.diabetes.ca/about-diabetes [WebCite Cache ID 6vdUtimEzb]
2. Fisher EB, Fitzgibbon ML, Glasgow RE, Haire-Joshu D, Hayman LL, Kaplan RM, et al. Behavior matters. Am J Prev Med 2011 May;40(5):e15-e30 [FREE Full text] [doi: 10.1016/j.amepre.2010.12.031] [Medline: 21496745]
3. Glasgow RE, Fisher EB, Anderson BJ, LaGreca A, Marrero D, Johnson SB, et al. Behavioral science in diabetes. Contributions and opportunities. Diabetes Care 1999 May;22(5):832-843. [Medline: 10332691]
4. Packer TL, Boldy D, Ghahari S, Melling L, Parsons R, Osborne RH. Self-management programs conducted within a practice setting: who participates, who benefits and what can be learned? Patient Educ Couns 2012 Apr;87(1):93-100. [doi: 10.1016/j.pec.2011.09.007] [Medline: 21992799]
5. Jones H, Edwards L, Vallis TM, Ruggiero L, Rossi SR, Rossi JS, Diabetes Stages of Change (DiSC) Study. Changes in diabetes self-care behaviors make a difference in glycemic control: the Diabetes Stages of Change (DiSC) study. Diabetes Care 2003 Mar;26(3):732-737. [Medline: 12610030]
6. Vallis M, Ruggiero L, Greene G, Jones H, Zinman B, Rossi S, et al. Stages of change for healthy eating in diabetes: relation to demographic, eating-related, health care utilization, and psychosocial factors. Diabetes Care 2003 May;26(5):1468-1474. [Medline: 12716806]
7. Nguyen MN, Potvin L, Otis J. Regular exercise in 30- to 60-year-old men: combining the stages-of-change model and the theory of planned behavior to identify determinants for targeting heart health interventions. J Community Health 1997 Aug;22(4):233-246. [Medline: 9247847]
8. King DK, Glasgow RE, Toobert DJ, Strycker LA, Estabrooks PA, Osuna D, et al. Self-efficacy, problem solving, and social-environmental support are associated with diabetes self-management behaviors. Diabetes Care 2010 Apr;33(4):751-753 [FREE Full text] [doi: 10.2337/dc09-1746] [Medline: 20150299]
9. Mishali M, Omer H, Heymann AD. The importance of measuring self-efficacy in patients with diabetes. Fam Pract 2011 Feb;28(1):82-87 [FREE Full text] [doi: 10.1093/fampra/cmq086] [Medline: 21047940]
10. Miller WR, Rose GS. Toward a theory of motivational interviewing. Am Psychol 2009 Sep;64(6):527-537 [FREE Full text] [doi: 10.1037/a0016830] [Medline: 19739882]
11. Wing RR, Goldstein MG, Acton KJ, Birch LL, Jakicic JM, Sallis JF, et al. Behavioral science research in diabetes: lifestyle changes related to obesity, eating behavior, and physical activity. Diabetes Care 2001 Jan;24(1):117-123. [Medline: 11194216]
12. Soo H, Lam S. Stress management training in diabetes mellitus. J Health Psychol 2009 Oct;14(7):933-943. [doi: 10.1177/1359105309341146] [Medline: 19786520]
13. Fisher EB, Thorpe CT, Devellis BM, Devellis RF. Healthy coping, negative emotions, and diabetes management: a systematic review and appraisal. Diabetes Educ 2007;33(6):1080-1103. [doi: 10.1177/0145721707309808] [Medline: 18057276]
14. Oka RK, DeMarco T, Haskell WL. Effect of treadmill testing and exercise training on self-efficacy in patients with heart failure. Eur J Cardiovasc Nurs 2005 Sep;4(3):215-219. [doi: 10.1016/j.ejcnurse.2005.04.004] [Medline: 15908277]
15. Izawa KP, Watanabe S, Omiya K, Hirano Y, Oka K, Osada N, et al. Effect of the self-monitoring approach on exercise maintenance during cardiac rehabilitation: a randomized, controlled trial. Am J Phys Med Rehabil 2005 May;84(5):313-321. [Medline: 15829777]
16. Clark NM, Dodge JA. Exploring self-efficacy as a predictor of disease management. Health Educ Behav 1999 Feb;26(1):72-89. [doi: 10.1177/109019819902600107] [Medline: 9952053]
17. Gardner JK, McConnell TR, Klinger TA, Herman CP, Hauck CA, Laubach Jr CA. Quality of life and self-efficacy: gender and diagnoses considerations for management during cardiac rehabilitation. J Cardiopulm Rehabil 2003;23(4):299-306. [Medline: 12894004]

18. Maddigan SL, Majumdar SR, Guirguis LM, Lewanczuk RZ, Lee TK, Toth EL, et al. Improvements in patient-reported outcomes associated with an intervention to enhance quality of care for rural patients with type 2 diabetes: results of a controlled trial. Diabetes Care 2004 Jun;27(6):1306-1312. [Medline: 15161780]
19. Aljasem LI, Peyrot M, Wissow L, Rubin RR. The impact of barriers and self-efficacy on self-care behaviors in type 2 diabetes. Diabetes Educ 2001;27(3):393-404. [doi: 10.1177/014572170102700309] [Medline: 11912800]
20. Kavanagh DJ, Gooley S, Wilson PH. Prediction of adherence and control in diabetes. J Behav Med 1993 Oct;16(5):509-522. [Medline: 8254654]
21. Canadian Diabetes Association. 2017. URL: http://guidelines.diabetes.ca/App_Themes/CDACPG/resources/cpg_2013_full_en.pdf [accessed 2012-11-17] [WebCite Cache ID 6vdVTKgJb]
22. Kirkman MS, Williams SR, Caffrey HH, Marrero DG. Impact of a program to improve adherence to diabetes guidelines by primary care physicians. Diabetes Care 2002 Nov;25(11):1946-1951. [Medline: 12401737]
23. Harris SB, Ekoé J, Zdanowicz Y, Webster-Bogaert S. Glycemic control and morbidity in the Canadian primary care setting (results of the diabetes in Canada evaluation study). Diabetes Res Clin Pract 2005 Oct;70(1):90-97. [doi: 10.1016/j.diabres.2005.03.024] [Medline: 15890428]
24. Braga MF, Casanova A, Teoh H, Gerstein HC, Fitchett DH, Honos G, Diabetes Registry to Improve Vascular Events [DRIVE] Investigators. Poor achievement of guidelines-recommended targets in type 2 diabetes: findings from a contemporary prospective cohort study. Int J Clin Pract 2012 May;66(5):457-464. [doi: 10.1111/j.1742-1241.2012.02894.x] [Medline: 22452524]
25. Nova Scotia Health Authority. Primary Health Care URL: http://www.nshealth.ca/primary-health-care [accessed 2017-12-11] [WebCite Cache ID 6vdVpSgdW]
26. Hunter CL, Goodie JL, Oordt MS, Dobmeyer AC. Integrated Behavioral Health in Primary Care: Step-by-Step Guidance for Assessment and Intervention. Washington: American Psychological Association; 2009.
27. Vogel ME, Kirkpatrick HA, Collins AS, Cederna-Meko CL, Grey MJ. Integrated care: maturing the relationship between psychology and primary care. Prof Psychol Res Pr 2012;43(4):271-280. [doi: 10.1037/a0029204]
28. Beeney LJ, Bakry AA, Dunn SM. Patient psychological and information needs when the diagnosis is diabetes. Patient Educ Couns 1996;29(1):109-116. [doi: 10.1016/0738-3991(96)00939]
29. Cauch-Dudek K, Victor JC, Sigmond M, Shah BR. Disparities in attendance at diabetes self-management education programs after diagnosis in Ontario, Canada: a cohort study. BMC Public Health 2013;13:85. [Medline: 23360373]
30. Stein R, Aubart-Schuller M, Vandenberghe H. [Evaluation and care of patients with type Ii diabetes regarding recommendations concerning cardiovascular risk factors and adherence to values in general medicine in the Grand Duchy of Luxembourg]. Bull Soc Sci Med Grand Duche Luxemb 2010(2):289-301. [Medline: 20882748]
31. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
32. Robu I, Robu V, Thirion B. An introduction to the Semantic Web for health sciences librarians. J Med Libr Assoc 2006 Apr;94(2):198-205 [FREE Full text] [Medline: 16636713]
33. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall, Inc; 1986.
34. McGowan P. The efficacy of diabetes patient education and self-management education in type 2 diabetes. Can J Diabetes 2011 Jan;35(1):46-53. [doi: 10.1016/S1499-2671(11)51008-1]
35. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [Medline: 12867348]
36. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262 [FREE Full text] [Medline: 11769298]
37. Norris SL, Lau J, Smith SJ, Schmid CH, Engelgau MM. Self-management education for adults with type 2 diabetes: a meta-analysis of the effect on glycemic control. Diabetes Care 2002 Jul;25(7):1159-1171. [Medline: 12087014]
38. Elder JP, Ayala GX, Harris S. Theories and intervention approaches to health-behavior change in primary care. Am J Prev Med 1999 Nov;17(4):275-284. [Medline: 10606196]
39. National Institute for Health and Care Excellence. NICE. A review of the use of the Health Belief Model (HBM), the Theory of Reasoned Action (TRA), the Theory of Planned Behavior (TPB) and the Trans Theoretical Model (TTM) to study and predict health related behavior change URL: http://www.nice.org.uk/nicemedia/live/11868/44524/44524.pdf [WebCite Cache ID 6vdaV1Gvq]
40. Abidi SS. Healthcare knowledge management: the art of the possible. In: Riano D, editor. Knowledge Management for Health Care Procedures. Lecture Notes in Computer Science, vol 4924. Berlin, Heidelberg: Springer; 2008.
41. Abidi S, Abidi S. An ontology-driven personalization framework for designing theory-driven self-management interventions. Process Support and Knowledge Representation in Health Care. Springer International Publishing, pp 2013:97.
42. Abidi SR, Cox J, Abusharekh A, Hashemian N, Abidi SS. A digital health system to assist family physicians to safely prescribe NOAC medications. Stud Health Technol Inform 2016;228:519-523. [Medline: 27577437]

43. Maghsoud-Lou E, Christie S, Abidi SR, Abidi SS. Protocol-driven decision support within e-Referral systems to streamline patient consultation, triaging and referrals from primary care to specialist clinics. J Med Syst 2017 Sep;41(9):139. [doi: 10.1007/s10916-017-0791-7] [Medline: 28766103]
44. Baig W, Abidi S, Abidi SS. An ontological model of behaviour theory to generate personalized action plans to modify behaviours. Stud Health Technol Inform 2016;228:399-403. [Medline: 27577412]
45. Stevens R, Goble CA, Bechhofer S. Ontology-based knowledge representation for bioinformatics. Brief Bioinform 2000 Jan 01;1(4):398-414. [doi: 10.1093/bib/1.4.398]
46. W3C. 2017. OWL Web Ontology Language Overview URL: https://www.w3.org/TR/owl-features/ [WebCite Cache ID 6vdaci8Pa]
47. Fernández-López M, Gómez-Pérez A, Juristo N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. 1997 Mar 24 Presented at: AAAI-97 Spring Symposium Series; March 24-26, 1997; Stanford University, EEUU p. 24-26.
48. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
49. d'Aquin M, Schlicht A, Stuckenschmidt H, Sabou M. Ontology modularization for knowledge selection: experiments and evaluations. In: Wagner R, Revell N, Pernul G, editors. Database and Expert Systems Applications. DEXA 2007. Lecture Notes in Computer Science, vol 4653. Berlin, Heidelberg: Springer; 2007.
50. Pathak J, Johnson TM, Chute CG. Survey of modular ontology techniques and their applications in the biomedical domain. Integr Comput Aided Eng 2009 Aug;16(3):225-242 [FREE Full text] [doi: 10.3233/ICA-2009-0315] [Medline: 21686030]
51. Stanford Center for Biomedical Informatics Research, Stanford University School of Medicine. 2017. Protégé URL: https://protege.stanford.edu/products.php [accessed 2017-12-11] [WebCite Cache ID 6vdajZeul]
52. Gómez-Pérez A. Evaluation of ontologies. Int J Intell Syst 2001;16(3):391-409. [doi: 10.1002/1098-111X(200103)16:3<391::AID-INT1014>3.0.CO;2-2]
53. The University of Manchester. 2017. FaCT++ reasoner URL: http://owl.cs.manchester.ac.uk/tools/fact/ [WebCite Cache ID 6vdbXhNEc]
54. GitHub. DWISE_V1.owl URL: https://github.com/niche-dal/dwise/blob/master/DWISE_V1.owl [accessed 2018-04-05] [WebCite Cache ID 6yRMpJhwI]
55. Van Woensel W, Roy PC, Abidi SR, Abidi SS. A mobile and intelligent patient diary for chronic disease self-management. Stud Health Technol Inform 2015;216:118-122. [Medline: 26262022]
56. Russell A, Abidi S, Abidi SS. Shared decision making: using theories and technology to engage the patient in their health journey. Stud Health Technol Inform 2014;205:303-307. [Medline: 25160195]
57. Abidi SS. Medical knowledge morphing: towards case-specific integration of heterogeneous medical knowledge resources. 2005 Presented at: 18th International Symposium on Computer-Based Medical Systems; June 23-24, 2005; Dublin, Ireland. [doi: 10.1109/CBMS.2005.74]
58. Abidi S, Cox J, Abidi SS, Shepherd M. Using OWL ontologies for clinical guidelines based comorbid decision support. 2012 Presented at: 45th Hawaii International Conference on System Sciences (HICSS); January 4-7, 2012; Maui, HI, USA.
59. Abidi S, Abidi SS, Hussain S, Shepherd M. Ontology-based modeling of breast cancer follow-up clinical practice guideline for providing clinical decision support. 2007 Presented at: 12th World Congress on Medical Informatics; August 20-24, 2007; Brisbane, Australia.
60. Abidi S, Abidi SS, Hussain S, Butler L. Ontology-based modeling and merging of institution-specific prostate cancer clinical pathways. In: Knowledge Management for Healthcare Procedures, LNAI 5626,. Berlin, Heidelberg: Springer Verlag; 2008.
61. Jafarpour B, Abidi S, Abidi SS. 2016. Exploiting OWL reasoning capabilities to develop semantic-web based clinical practice guideline execution engines URL: https://pdfs.semanticscholar.org/b942/5befbc48c356826aafbc5a081974de1a86d5.pdf [accessed 2018-04-02] [WebCite Cache ID 6yNU3Jngt]
62. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
63. Abidi SR, Stewart S, Shepherd M, Abidi SS. Usability evaluation of family physicians' interaction with the Comorbidity Ontological Modeling and ExecuTion System (COMET). Stud Health Technol Inform 2013;192:447-451. [Medline: 23920594]
64. Abidi S, Vallis M, Piccinini H, Imran SA, Abidi R. A digital framework to support providers and patients in diabetes related behavior modification. Stud Health Technol Inform 2017;235:589-593. [Medline: 28423861]
65. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: Sage Publications; 1998.
66. Canadian Diabetes Association. Diabetes in Canada and supports available URL: https://www.diabetes.ca/getmedia/410c30ab-cb8d-45b6-b1f1-1694f4f89c49/2015-report-on-diabetes-driving-change-english-chapter-1.pdf.aspx [accessed 2017-12-11] [WebCite Cache ID 6vda7UKZ4]

Abbreviations

BCI: Behavior Change Institute
BCO: Behavior Change Ontology
CDA: Canadian Diabetes Association
CDE: certified diabetes educators
CPG: clinical practice guidelines
DMC: Diabetes Management Centers
DWISE: Diabetes Web-Centric Information and Support Environment
HbA_{1c}: glycated hemoglobin
PCP: primary care practitioner
SCT: social cognition theory
OWL: Web Ontology Language
SMART: Specific, Measurable, Action-Oriented, Relevant, Timely
TAP: Think Aloud Protocol

Edited by G Eysenbach; submitted 11.12.17; peer-reviewed by B Davis, R Pankomera; comments to author 17.01.18; revised version received 04.02.18; accepted 08.02.18; published 18.04.18.

Please cite as:
Abidi S, Vallis M, Piccinini-Vallis H, Imran SA, Abidi SSR
Diabetes-Related Behavior Change Knowledge Transfer to Primary Care Practitioners and Patients: Implementation and Evaluation of a Digital Health Platform
JMIR Med Inform 2018;6(2):e25
URL: http://medinform.jmir.org/2018/2/e25/
doi:10.2196/medinform.9629
PMID:29669705

Original Paper

Automated Modular Magnetic Resonance Imaging Clinical Decision Support System (MIROR): An Application in Pediatric Cancer Diagnosis

Niloufar Zarinabad[1,2], MSc, PhD; Emma M Meeus[1,2,3], MSc, PhD; Karen Manias[1,2], MD, MRCPCH; Katharine Foster[2], MD, FRCR; Andrew Peet[1,2], PhD, FRCPCH

[1]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, United Kingdom

[2]Birmingham Children Hospital NHS Trust, Birmingham, United Kingdom

[3]Physical Sciences of Imaging in Biomedical Sciences Doctoral Training Centre, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Andrew Peet, PhD, FRCPCH
Institute of Cancer and Genomic Sciences
University of Birmingham
Edgbaston
Birmingham,
United Kingdom
Phone: 44 0121 333 8234
Fax: 44 0121 333 8241
Email: a.peet@bham.ac.uk

Abstract

Background: Advances in magnetic resonance imaging and the introduction of clinical decision support systems has underlined the need for an analysis tool to extract and analyze relevant information from magnetic resonance imaging data to aid decision making, prevent errors, and enhance health care.

Objective: The aim of this study was to design and develop a modular medical image region of interest analysis tool and repository (MIROR) for automatic processing, classification, evaluation, and representation of advanced magnetic resonance imaging data.

Methods: The clinical decision support system was developed and evaluated for diffusion-weighted imaging of body tumors in children (cohort of 48 children, with 37 malignant and 11 benign tumors). Mevislab software and Python have been used for the development of MIROR. Regions of interests were drawn around benign and malignant body tumors on different diffusion parametric maps, and extracted information was used to discriminate the malignant tumors from benign tumors.

Results: Using MIROR, the various histogram parameters derived for each tumor case when compared with the information in the repository provided additional information for tumor characterization and facilitated the discrimination between benign and malignant tumors. Clinical decision support system cross-validation showed high sensitivity and specificity in discriminating between these tumor groups using histogram parameters.

Conclusions: MIROR, as a diagnostic tool and repository, allowed the interpretation and analysis of magnetic resonance imaging images to be more accessible and comprehensive for clinicians. It aims to increase clinicians' skillset by introducing newer techniques and up-to-date findings to their repertoire and make information from previous cases available to aid decision making. The modular-based format of the tool allows integration of analyses that are not readily available clinically and streamlines the future developments.

(JMIR Med Inform 2018;6(2):e30) doi:10.2196/medinform.9171

KEYWORDS

clinical decision support; real-time systems; magnetic resonance imaging

Introduction

Magnetic resonance imaging (MRI) is a fast-growing clinical imaging modality and has become the modality of choice for the evaluation of disease and treatment management across multiple therapeutic areas. It has increasingly been used in oncology, central nervous system diseases, musculoskeletal disorders, and cardiovascular disease due to its superior soft-tissue imaging capabilities, lack of ionizing radiation, and noninvasive nature [1-4].

MRI technology constantly advances with new magnetic resonance applications being pioneered, investigated, mainstreamed, and added to clinical applications and capabilities. Nevertheless, clinical interpretation remains largely by qualitative expert review. In addition, new advanced and computationally intensive medical quantitative image analysis techniques are constantly being developed and validated. These techniques have allowed the discovery of specific biomarkers of both disease and treatment response and have exposed clinicians to new information in a computable format [5-8]. However, the growing and versatile amount of magnetic resonance–derived information can form an insurmountable obstacle to the individual clinician; in particular, the use of quantitative MRI biomarkers requires further improvement in accessibility and presentation to aid decision making [9].

In the past decade, clinical decision support (CDS) systems have increasingly gained attention, and the routine uptake of these intelligent systems is becoming more common [10-17]. Introduction of CDSs has provided clinicians and health care investigators with a platform for extraction of relevant information to aid decision making, prevent errors, and enhance health care. CDSs include a range of options from computerized alerts, reminders, and clinical guidelines to diagnostic support and clinical workflow through computer-assisted diagnosis tools (CAD) [18-26]. There are several clinically implemented or research-based CADs available for medical image analysis [23,26-29]. However, majority of them lack at least one of the following: (1) a user-friendly graphical interface to be used by clinicians in their clinical routine; (2) system performance is often not compared with radiologist diagnosis in the absence of the tool or when the tool is utilized; (3) are not MRI based; (4) are designed for one particular disease; and (5) are just a single postprocessing tool or analysis algorithm, which also provides a likelihood for a disease and does not offer decision support for the clinicians (ie, in form of only providing additional structured information for comparison with available other relevant diagnosis). These types of solutions have shown to suffer from high false positives [9].

Availability of a user-friendly and flexible MRI CAD that encompasses a variety of medical image analysis techniques and postprocessing methods and can act as a CDS could facilitate the uptake of new advanced magnetic resonance techniques in the real-time clinical setting; it could also allow health care investigators to interrogate their data in a scientifically informative and convenient manner to determine a robust and efficient diagnosis. The aim of this study was to design, develop, and evaluate a medical image region of interest analysis tool and repository (MIROR) platform for conventional magnetic resonance data aimed at improving clinical performance through the provision of real-time diagnostic support for clinicians.

To the best of our knowledge, there is only the International Network for Pattern Recognition of Tumours Using Magnetic Resonance Decision Support System validated and available for the analysis of magnetic resonance spectroscopy (MRS) data [30]. However, this CDS is developed for diagnosing and grading adult brain tumors and is based on MRS only. There is no CDS for both MRI and MRS analysis with a robust user interface for clinical routine use that is capable of creating and updating a validated repository for different diagnostic problems.

Methods

Clinical Decision Support System Design

Features available in the presented version of the MIROR are (1) a clinician-friendly graphical user interface; (2) measurement of morphologic properties such as size, shape, volume, length dimensions, and center-of-mass location of the region of interest (ROI); (3) an integrated magnetic resonance diffusion-weighted imaging (DWI) analysis application based on intravoxel incoherent motion (IVIM) model; (4) statistical data analysis of the ROI overlaid on standard MRI images (such as T1-weighted and T2-weighted scans or the advanced quantitative maps) to provide decision support in forms of comparison with other differential diagnosis and several different image volumes to aid diagnosis and determination of prognosis; and (5) a self-archiving repository of the extracted data and features. Availability of the latter 2 options in combination with the first 3 will move the designed tool from a CAD toward becoming a CDS for MRI data. MIROR also allows investigators to further grow, advance, and combine different analysis techniques and types of imaging sequences to extend the tool to a more sophisticated decision support, dependent on their individual center's needs to better inform diagnosis. MIROR's self-archiving, evolving repository is the core of its decision support. This unique feature of the MIROR distinguishes it from pervious CADs and CDSs. First, the repository's continuous development allows for improvement in the predication accuracy for the available biomarkers and disease in the database; second, it permits provision of a decision support system compatible to additional disease types by means of importing and appending the repository.

We used a modular and open architecture design [23] in the design and implementation of MIROR to be able to adapt to the constant increase and development in the MRI sequences; it will also make room for consequent advances in the related analysis applications and allow future development of additional new workflows. Additionally, we used Mevislab software (v. 2.7.1, MeVis AG- Fraunhofer-MEVIS) [31,32], a research-based rapid prototyping platform for medical image processing, for development of the MIROR to achieve the latter. Post processing, quantitative and statistical analysis functionalities embedded in MIROR were either developed using Python (v. 2.7, embedded within Mevislab) or were imported from the Mevislab library. The MIROR repository was developed using

Python. Each individual independent module of MIROR was developed, evaluated, and tested by different groups within the team in their own life cycle and schedules before addition to the final product. A hierarchical structure of MIROR infrastructure is represented in Figure 1. MIROR self-archiving, evolving repository is the core of its decision support. This feature of the MIROR distinguishes it from pervious computer-assisted diagnosis tools (CADs) and CDSs.

Based on the recommendations of the American medical informatics association [33], an evidence-adaptive approach was employed in the design of the MIROR by utilizing its knowledge base to derive from—and reflect on—the most up-to-date evidence from the research literature and practice-based sources [34]. The statistical and quantitative analysis module embedded in MIROR are developed based on the literature and local practice-based research and will continue to update in future releases. MIROR is an evolving database of available diagnosis data gathered from routine clinical practice. Outcomes of this repository data analysis can inform future clinical investigations, reflect on the clinical practice, and consequently impact on the MIROR statistical and quantitative analysis module. Conversely, practice-based experience can inform the choice of MRI sequence and parametric maps to be used for analysis and clinical evaluation (Figure 2).

MIROR can import all file formats supported by the National Library of Medicine Insight Segmentation and Registration Toolkit, such as digital imaging and communications in medicine (DICOM) files, Neuroimaging Informatics Technology Initiative files, JPEG as well as text files, and comma spreadsheets. Built-in Mevislab modules and Python were used to create this functionality. Using the module raw images, postprocessed quantitative maps and data files can be imported from the clinical data warehouses, such as hospital picture archiving and communication system (PACS), the local servers, or MIROR for future analysis and visualization (Figure 1).

Comparing the MIROR architecture with previously developed CDSs with a clinical data base and domain expert knowledge base, MIROR does not connect to hospital electronic health record (EHR) system or any Internet-based database or medical knowledge representations or guidelines [10,35,36]. Having said so, one should note that imaging data can be imported to MIROR through connection to hospital PACS and therefore can be considered as a semi-integrated CDSs [37]. Moreover, currently available active and robust CDSs benefit from EHR data with very large and historical dataset that changes continuously and contains hidden knowledge. MIROR was designed based on a similar architecture applied to a repository containing a constantly updated independent database. The updating of the database allows the advanced MRI biomarkers to be revised whenever new data are available.

One of the main strengths of MIROR is its ability to allow for integration of new advanced and computationally intensive quantitative analyses that are not readily available to be used in routine clinical practice under the Advance Quantitative analysis module (Figure 1). In this study, the analysis of multi b-value (b=diffusion weightings) magnetic resonance DWI is embedded in MIROR. The analysis was developed using the well-established and not clinically available intravoxel incoherent motion (IVIM) model, which has been shown to have clinical value in many different tumor types [38-40] as well as in other pathologies [41,42]. Although IVIM provides a similar measure to clinically available apparent diffusion coefficient (ADC), derivation of the additional parameters allows the separation of the perfusion contribution from the true diffusion, resulting in a greater insight to the underlying tissue microenvironment [43-45].

Figure 1. Flow diagram showing hierarchical structure of the medical image region of interest analysis tool and repository (MIROR) infrastructure. Dashed blue lines indicate direct connection of the module output to the front-end display, solid lines are the connections between internal clinical decision support (CDS) modules, and green dashed lines represent the feedback system to the repository. PACS: picture archiving and communication system; ROI: region of interest; MRI: magnetic resonance imaging. MRI: magnetic resonance imaging; DICOM: digital imaging and communications in medicine.

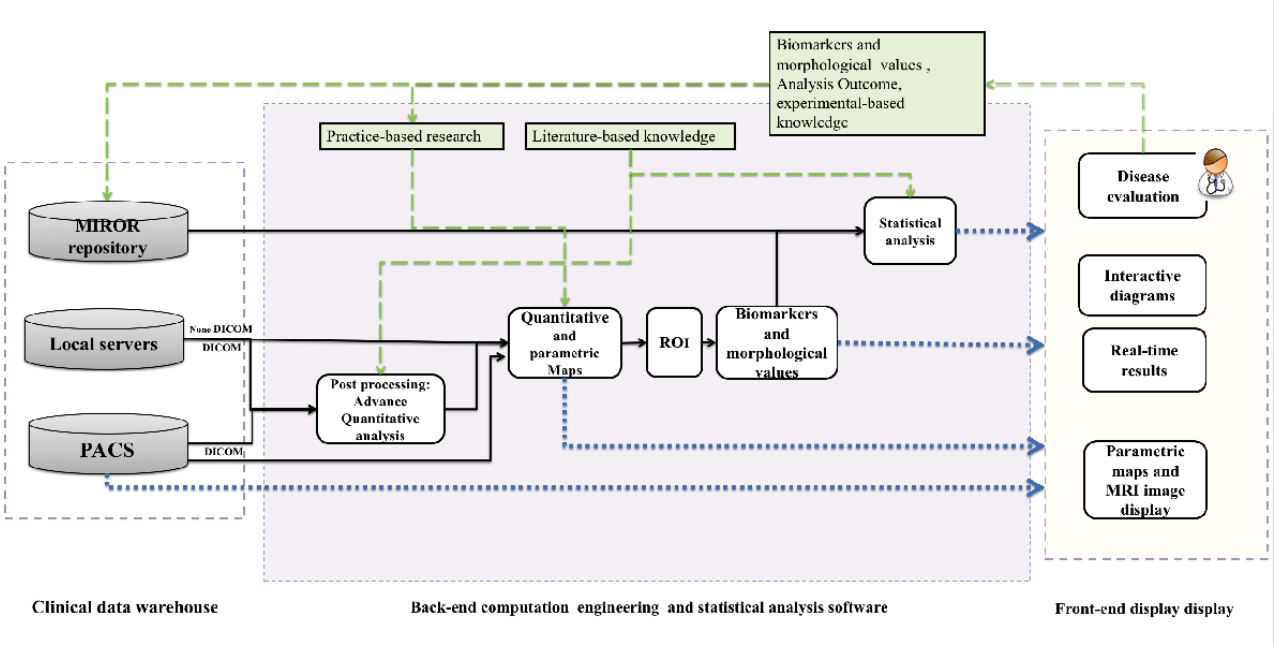
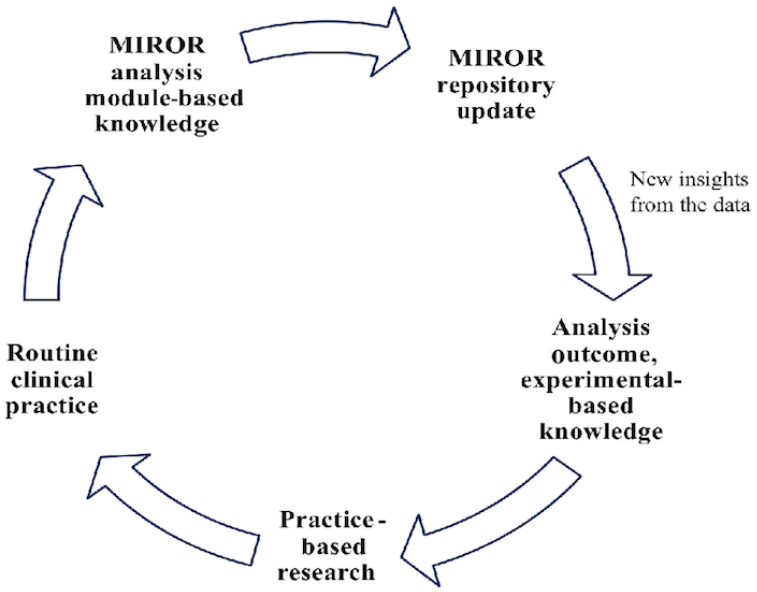


Figure 2. Medical Image Region of interest analysis tool and Repository (MIROR) evidence-adaptive protocol. An evidence-adaptive approach was utilized in the design of the MIROR by utilizing its knowledge base to derive from and reflect on the most up-to-date evidence from the research literature and practice-based sources.



This analysis allows the computation of tissue diffusion coefficient (IVIM-*D*), pseudo-diffusion coefficient (IVIM-*D**), and perfusion fraction (IVIM-*f*) [46]. However, as a relatively new analysis method, it is currently not available as part of the scanner software packages. The MIROR IVIM module, which is developed using python programming, loads the raw DWI DICOM file and based on it determines whether the IVIM analysis is feasible, based on the number of the b values. The output from the module is the IVIM parameter maps, which can then be used for further ROI-based statistical analysis.

The MIROR-embedded statistical analysis module provides specialists with instruments for (1) analyzing and interpreting individual patient MRI data and (2) comparing it with results of previous cases directly from the MIROR-evolving repository via powerful statistical techniques to future inform the investigation (Figure 1). MIROR provides statistical analysis of the advanced parametric maps or the standard MRI sequence image's ROI (ie, diffusion maps produced in MIROR, other imported advance quantitative maps, or T1-weighted and T2-weighted scans). It measures the volume of the defined ROI, creates a histogram for it, performs statistical analysis on the ROI values, and extracts and stores histogram parameters such as entropy, median, mean, different quantiles, skewness, and kurtosis in its repository. Unlike previously reported CAD and CDS tools, using MIROR the medical expert has multiple options, including (1) selection of the population of interest from the repository to work only with data from a specific condition, (2) choosing the MRI biomarker/variable (eg, diffusion, perfusion), and (3) the statistical variable of interest from a complete set of basic and advanced features that cover both clinical and research needs (basic statistics mean, median, variance, standard deviation, quantile, histogram analysis, etc).

Note that MIROR is a nonregion-specific MRI CDS, and its novel quantitative image analysis and statistical analysis modules are designed to analyze any region of the body and aid in resolving different demanding diagnostic problems.

The frontend of the MIROR is a clinician's user-friendly graphical interface that displays MRI images as well as quantified parametric maps and allows clinicians to define their ROI. It also provides real-time morphological and statistical results for compassion with the repository data so as to aid clinical evaluation of the disease.

Medical Image Region of Interest Analysis Tool and Repository Application to Pediatric Tumor Evaluation

MIROR is currently being developed, evaluated, and used at Birmingham Children's Hospital to determine its role in facilitating noninvasive diagnosis in children presenting with solid body tumors in clinical practice.

Solid masses in children represent a diagnostic dilemma, as neoplastic and non-neoplastic lesions can appear similar on conventional imaging. Although in some cases the clinical history and physical examination findings indicate a likely diagnosis, the majority of cases require further evaluation with MRI to assess the extent of the lesion and make a specific diagnosis. It is often difficult to determine whether a lesion is benign or malignant or identify specific tumor type based on conventional MRI alone.

Table 1. Body tumor patient cohort demographics.

Tumor	Median age (range)	Gender	Diagnosis	Patients, n
Benign	3.63 (0.03-14.22)	Female=6; Male=7	Liver hemangioma	1
			Ganglioneuroma	4
			Hematocolpos	1
			Lipoma	1
			Infantile myofibromatosis	1
			Mesoblastic nephroma	2
			Hematocolpos	1
			Vascular malformation	1
			Ovarian immature teratoma	1
Malignant	3.94 (0.03-11.82)	Female=16; Male=21	Clear cell sarcoma of kidney	1
			Ewing's sarcoma	1
			Germ cell tumor	1
			Hepatoblastoma	4
			Neuroblastoma	11
			Osteosarcoma	1
			Rhabdoid tumor	2
			Rhabdomyosarcoma	3
			Wilms tumor	13

This study evaluated the impact of information provided by MIROR in aiding clinicians to distinguish between benign and malignant solid body pediatric tumor types using DWI. Accuracy testing involved examination of MIROR for a cohort of real patient cases with recent visits to Birmingham Children's Hospital and comparison of the MIROR outcome with the radiologist's initial opinion and final diagnosis derived based on the opinion of the clinical multidisciplinary team of experts together with pathology.

Magnetic Resonance Imaging Data

The body tumor patient cohort studied consisted of children (aged 0-16 years) with solid tumors, undergoing diagnostic MRI with multi b-value DWI at Birmingham Children's Hospital from 2012 to September 2016. A total of 48 children were enrolled, of whom 37 had malignant tumors and the rest had benign tumors. Details of the malignant and benign body tumors along with patients' demographics are presented in Table 1.

We performed the MRI on a Siemens Avanto 1.5 T MRI scanner (Siemens Healthcare, Erlangen, Germany). The diffusion-weighted MRI protocol used an echo-planar imaging sequence in an axial acquisition plane with a field-of-view (FOV) 221 to 350 × 172 to 317 mm^2, matrix size 122 to 192 × 128 to 192, slice thickness of 5 mm, and in-plane resolution of 1.56 × 1.56 mm^2. For all subjects, 6 b-values: 0, 50, 100, 150, 600, and 1000 s/mm^2 were acquired in 3 orthogonal directions with TR/TE=3200 to 9900/92 ms and number of averages=3. The signal to noise of the MRI dataset was approximately 30 (SD 10) for b1000 and 60 (SD 10) for b0 images.

Depending on their ability to cooperate, children were awake, sedated, or under general anesthesia. MRI acquired with different diffusion weightings (b-values) was used to compute ADC (computed from b_0 and b_{1000}), IVIM-*D*, IVIM-*D**, and IVIM-*f* maps. These values are a quantitative measure of diffusion related to tissue cellularity [45,47] and can be useful for tumor characterization. Clinicians currently only use ADC maps in a qualitative manner to help tumor characterization, commenting on restriction of diffusion as a possible marker of malignancy. Advanced quantitative diffusion parameters (ie, IVIM-*D*, IVIM-*D**, and IVIM-*f*) and means for direct and real-time statistical analysis of these variables are unavailable clinically, despite the growing body of evidence for their potential value in noninvasive diagnosis.

Discrimination Between Benign and Malignant Tumor Types

To discriminate between malignant and benign tumors, the authors made use of a leave-one-out cross-validation method combined with a {displaystyle C_{30}^{100}approx 3times 10^{25}.} threshold-based classification approach to determine the potential of individual parameters determined by MIROR. To achieve this, one case of the cohort was assigned in turn as the validation case, with the remaining ones used for training to inform the outcome. The selected validation case was then iteratively changed until all cases had been evaluated exactly once. On the basis of the threshold approach, if the value of the case under study lay within 1 standard deviation of the mean of a statistical parameter of the training tumor group (ie, benign or malignant), it was assigned to that particular group. However, if the value of the parameter under study fell within both tumor

group regions, we made use of second-adjusted classification based on distance from the mean of the statistical parameter of the groups in the training set. To further evaluate the significance of the statistical parameters and information provided by MIROR, k-nearest neighbor (KNN) and support vector machine (SVM) pattern recognition techniques, followed by leave-one-out cross-validation, were utilized to test the accuracy of derived data in distinguishing between tumor types, using all parameters as classification features. KNN was chosen for its simplicity and performance on basic recognition problems; it has been a ubiquitous classification method with good scalability. SVM outperforms conventional pattern recognition methods, especially when the number of training data is small and number of input variables is large [48].

To account for the data skewness and imbalanced distribution of the 2 groups, synthetic minority oversampling technique has been used to allow for building a larger decision region that contains nearby instances of the minority class [49] when KNN and SVM are used.

Feature selection was performed before classification by means of calculating the significance level of the histogram derived parameters between the tumor groups. Then, benign and malignant tumors' histogram permanents (ie, Median; 2nd, 5th, 10th, 15th, 25th, 75th, 85th, 90th, and 98th centile values; kurtosis; skewness; and entropy) were compared using the nonparametric Mann-Whitney U test. The authors used the Bonferroni correction method. Parameters showing a significant difference between the 2 groups were used for classification. All statistical analyses were performed using SPSS Statistics (v. 23, Chicago, Illinois) software.

All patients were consented for research to the UK Children's Cancer and Leukemia Group, Functional Imaging Group database, a UK National Health Service Research Ethics committee-approved study (Reference number 04/MRE04/41, Health Research Authority East Midlands—Derby, UK, Ethical Review Board Chair, Dr Peter Korczak). Informed participation and publication consent was given by parents/guardians.

MIROR is aimed at improving clinical practice through the provision of real-time diagnostic support. To ensure achievement of the latter, each individual application and module has been tested in such an environment by allowing clinicians to interrogate it about the most important clinical questions and provide feedback. We used an iterative process of design, testing, and revision of the MIROR by a diverse team, including medical informatics experts, clinical content experts, and end users to ensure reliable translation of the tool to clinical practice. Experts in MRI and medical imaging, including PhD researchers specializing in MRI and data processing, a physician, and a senior consultant radiologist pilot tested MIROR iteratively during the development and refinement of the tool.

Results

MRI datasets and ADC produced by the scanner were imported to MIROR from a local PACS. An ROI was drawn around the entire solid tumor for each case on a high-resolution image by one clinician, which was then checked by another (KM and KF) before being transferred to a matched parametric map (eg, ADC and IVIM maps). The entire tumor volume, including cystic and necrotic areas, was included in the ROI to determine representative data for heterogeneous tumors [50]. A histogram of the drawn ROI was constructed; the mean, median, 2nd-98th percentile values, skewness, kurtosis, and entropy of the histograms were calculated for all tumors, recorded, and stored on a database. Figure 3 demonstrates a screenshot of MIROR for all of the described stages. Advance IVIM parametric maps D and f are also shown in Figure 4.

Figure 5 represents an example of MIROR's decision support module for evaluating benign and malignant cases.

To establish differences between the malignant and benign lesions histogram parameters, the authors compared the histogram-derived parameters in the repository. Analysis showed apparent differences between malignant and benign tumors, with lower ADC values and higher skewness and kurtosis in malignant lesions. There was no significant difference between 85th (P=.12), 90th (P=.22), 95th (P=.82), or 98th centile (P=.41) ADC values between benign and malignant tumors (Table 2). Malignant tumors demonstrated statistically significantly lower mean (P=.03), median (P=.005), 2nd (P=.04), 5th (P=.02), 10th (P=.01), 15th (P=.005), 25th (P=.004), and 75th centile (P=.03) ADC values, higher kurtosis (P<.01), more positively skewed histograms (P<.001), and higher entropy (P=.03). These results are in agreement with similar studies published for adults [51-53].

The feasibility of MIROR to provide distinctive surplus information, which would further aid diagnosis, was evaluated using histogram-derived parameters with the statistically significant differences between the 2 groups. The accuracy of individual statistical parameters to discriminate between specific benign or malignant tumors is presented in Table 3. Due to the high correlation between the ADC centiles, only 15% and 75% centiles were used for classification.

Figure 3. Medical image region of interest analysis tool and repository (MIROR) user-interface patient view. This figure represents data for a malignant tumor case. Here, the region of interest (ROI) is drawn on a high-resolution image and overlaid on the corresponding parametric apparent diffusion coefficient (ADC) map. Measurement of morphologic properties of the ROI, zooming, scaling, rotating of the estimated object surface, and histogram analysis of the overlaid ROI on voxel-by-voxel parametric maps, is supported to enhance quality assessment.

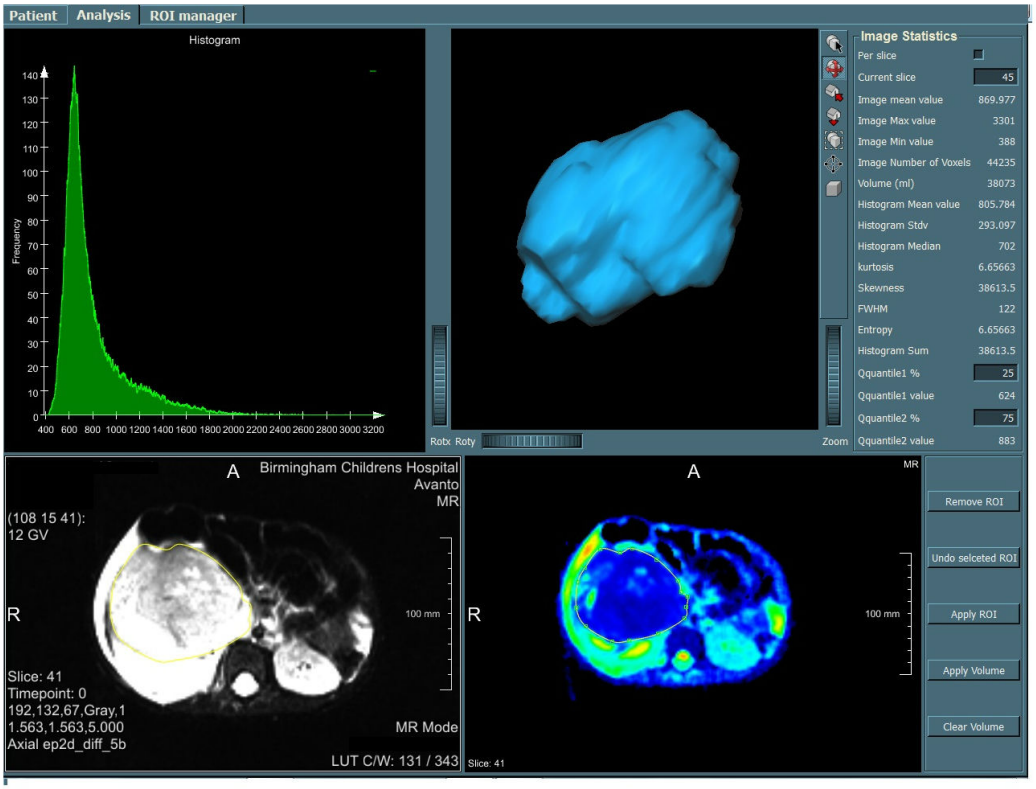


Figure 4. Medical image region of interest analysis tool and repository (MIROR) user-interface intravoxel incoherent motion (IVIM) maps tabs. This figure represents data for a malignant tumor case. Here, the region of interest (ROI) is drawn on a high-resolution image and overlaid on the corresponding parametric map IVIM-D and IVIM-f. Measurement of morphologic properties of the ROI, zooming, scaling, rotating of the estimated object surface and histogram analysis of the overlaid ROI on voxel-by-voxel parametric maps, is supported to enhance quality assessment.

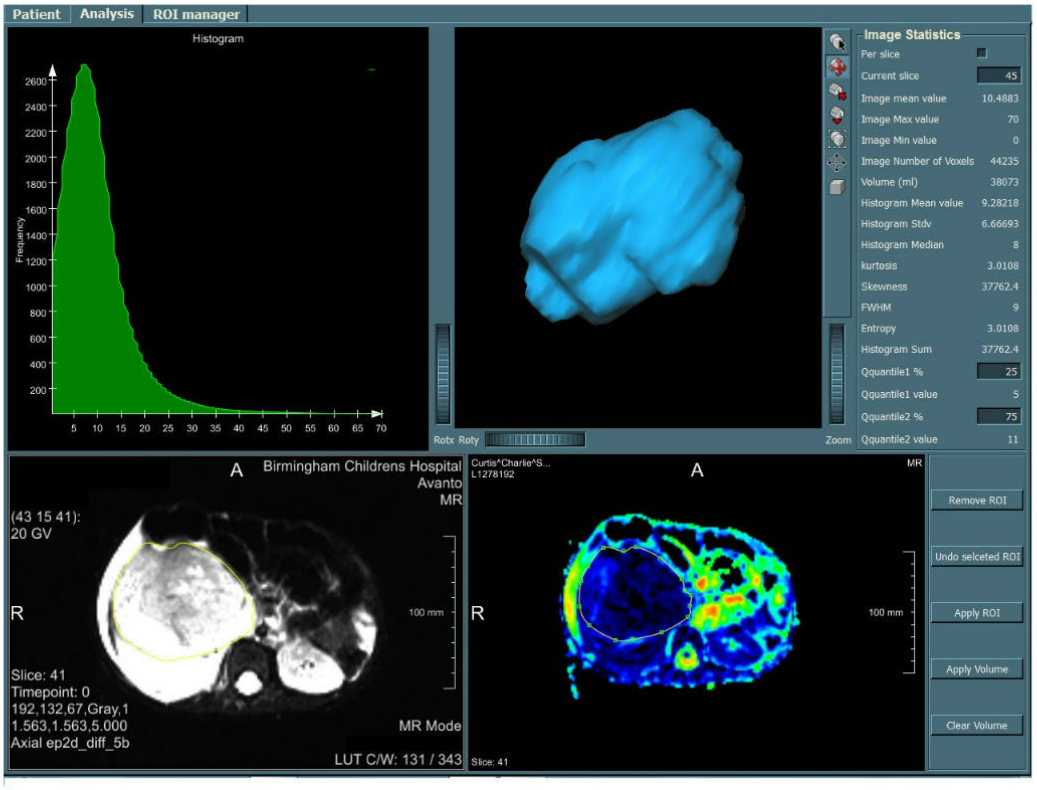
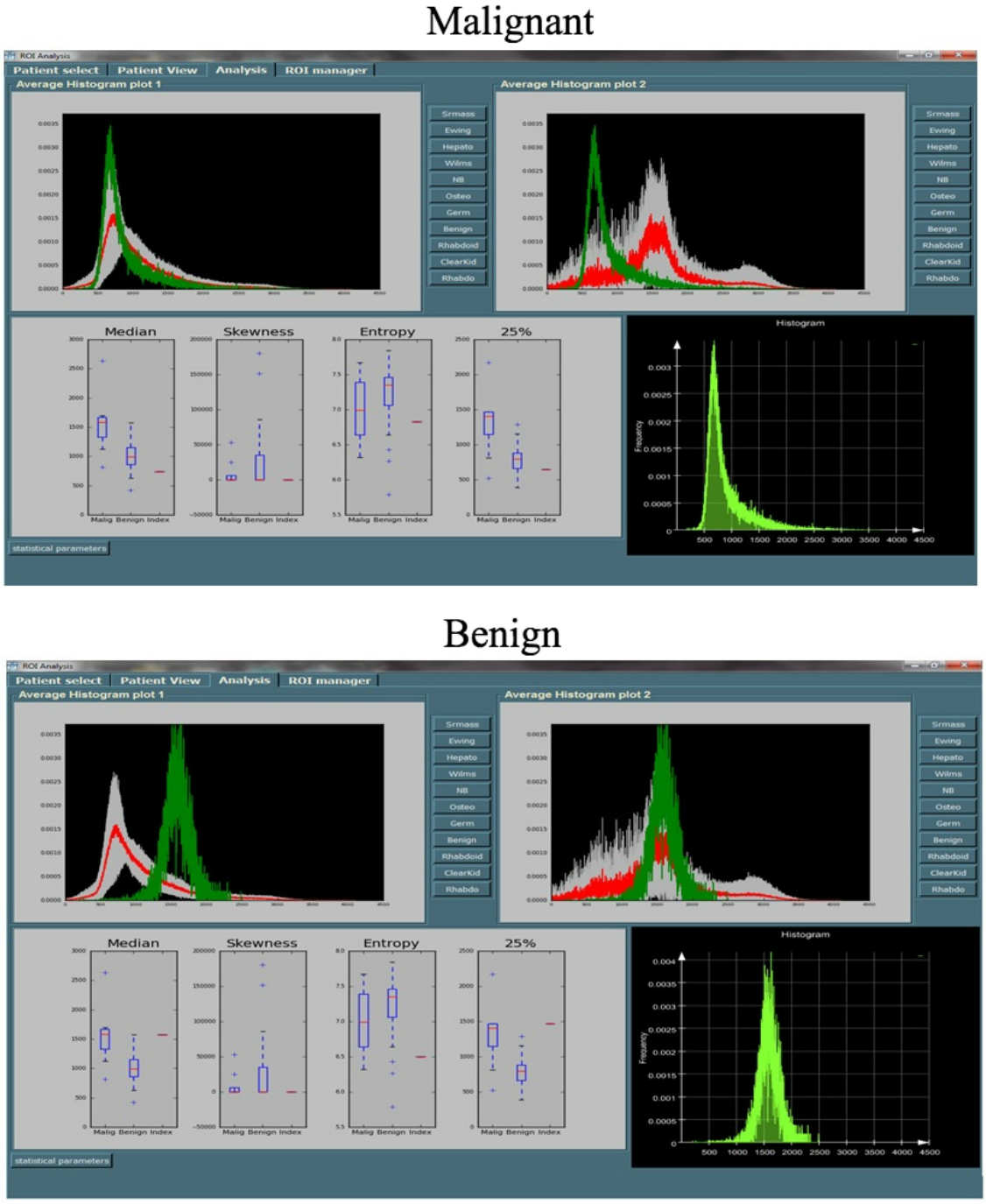


Figure 5. Medical image region of interest analysis tool and repository (MIROR) user-interface analysis tab. This figure represents MIROR use as a decision support system for benign and malignant cases. Here, the green histogram line represents the index case under examination; the red line and gray area represent the mean and standard deviation of the selected tumor group for comparison, respectively. The box plot compares median, skewness, entropy, and 25th percentile values of the index case with the tumor types in the database.



In the studied cohort, skewness, kurtosis, entropy, different percentiles, mean, and median provided important distinctive additional information, which was not available by using qualitative approaches only. Using kurtosis, entropy and 15th percentile for threshold-based classification, 100% of malignant cases were correctly assigned. Mean, median, and skewness had an accuracy of 0.97 in classifying malignant cases. In the benign category, kurtosis, entropy, and 75th percentile achieved full accuracy. Mean and median had an accuracy of 0.91 in classifying benign tumors. Overall, kurtosis and entropy had the highest sensitivity and specificity (sensitivity=1, specificity=1) in discriminating between benign and malignant tumors. Note that the threshold classification was performed based on a single feature input and a 2-step classification process, with a less strict rule in second layer to reclassify the cases in the ambiguous group. Using all of the above extracted features, more advanced pattern recognition techniques, and 10-fold cross-validation, an accuracy of 0.89 (sensitivity=0.97, specificity=0.5, area under the curve [AUC]=0.78) and 0.93 (sensitivity=0.97, specificity=0.58, AUC=0.84) was obtained by SVM and KNN, respectively. Figure 6 shows a comparison of the classifiers using 10-fold cross-validation receiver operating characteristic analysis.

Table 2. Comparison of apparent diffusion coefficient (ADC) histogram parameters between malignant and benign pediatric tumors using the Mann-Whitney *U* test.

Parameters	*P* value
Mean	.03[a]
Median	.005[a]
Kurtosis	<.01[a]
Skewness	<.01[a]
Entropy	.03[a]
2% percentile	.04[a]
5% percentile	.02[a]
10% percentile	.01[a]
15% percentile	.005[a]
25% percentile	.004[a]
75% percentile	.03[a]
85% percentile	.12
90% percentile	.22
95% percentile	.82
98% percentile	.41

[a]Statistical significance $P<.05$.

Table 3. Accuracy of individual statistical parameters along with sensitivity and specificity of the analysis obtained by medical image region of interest analysis tool and repository (MIROR) to discriminate between benign or malignant tumors using 2-step threshold classifications.

Marker	Mean	Median	Kurtosis	Skewness	Entropy	15% percentile	75% percentile
Malignant	1098 (SD 295)	996 (SD 262)	2.1 (SD 0.09)	0.02 (SD 0.004)	7.1 (SD 0.42)	710 (SD 201)	1319 (SD 329)
Benign	1443 (SD 462)	1442 (SD 511)	2.031 (SD 0.11)	0.0007 (SD 0.01)	6.85 (SD 0.4)	1072 (SD 406)	1683 (SD 538)
Sensitivity	0.94	0.97	1	0.97	1	1	0.97
Specificity	0.91	0.91	1	0.58	1	0.83	1
Accuracy	0.932	0.9455	1	0.6955	1	0.9165	0.987

To further evaluate MIROR in terms of its added clinical value, radiologist initial diagnosis from the first MRI scans was compared with the final diagnosis obtained from histopathology, and the outcome of the 3 statistical analyses based on MIROR provided information for classifying the tumor types for this cohort of patients (Figure 7). For the 2-step thresholding classification, outcome of the first classification layer is presented to include the ambiguous group. The ambiguous groups for KNN and SVM were identified by thresholding their predication probabilities at above 0.8 and above 0.5 for the accurate and ambiguous assignment of cases, respectively.

A higher amount of uncertainty was observed in the initial diagnosis of the benign group. The benign group diagnostic uncertainty rate decreased when we used the information provided by MIROR. Moreover, the false diagnosis rate for both the malignant and benign groups was reduced compared with the radiologist's initial report with all 3 analysis methods. Additional statistical information provided to clinicians by MIROR can allow for a better and more informed noninvasive discrimination of benign and malignant body tumors in children.

Net reclassification improvement (NRI) [54] for KNN, SVM, and the 2-step thresholding methods using the histogram parameters were calculated to evaluate the level of improvement achieved by these methods compared with the radiologist's initial reading . The same is presented in Figure 8. KNN had the highest incremental value (NRI=0.35) among all the methods. Histogram parameters on average had a NRI of 0.19 in comparison with the radiologist's initial read.

Figure 6. Receiver operating characteristic (ROC) curves of support vector machine and k-nearest neighbor (KNN) in discriminating benign from malignant tumors using medical image region of interest analysis tool and repository (MIROR)-derived parameters. Area under the curve (AUC) was 0.78 for support vector machine (SVM) and 0.84 for KNN.

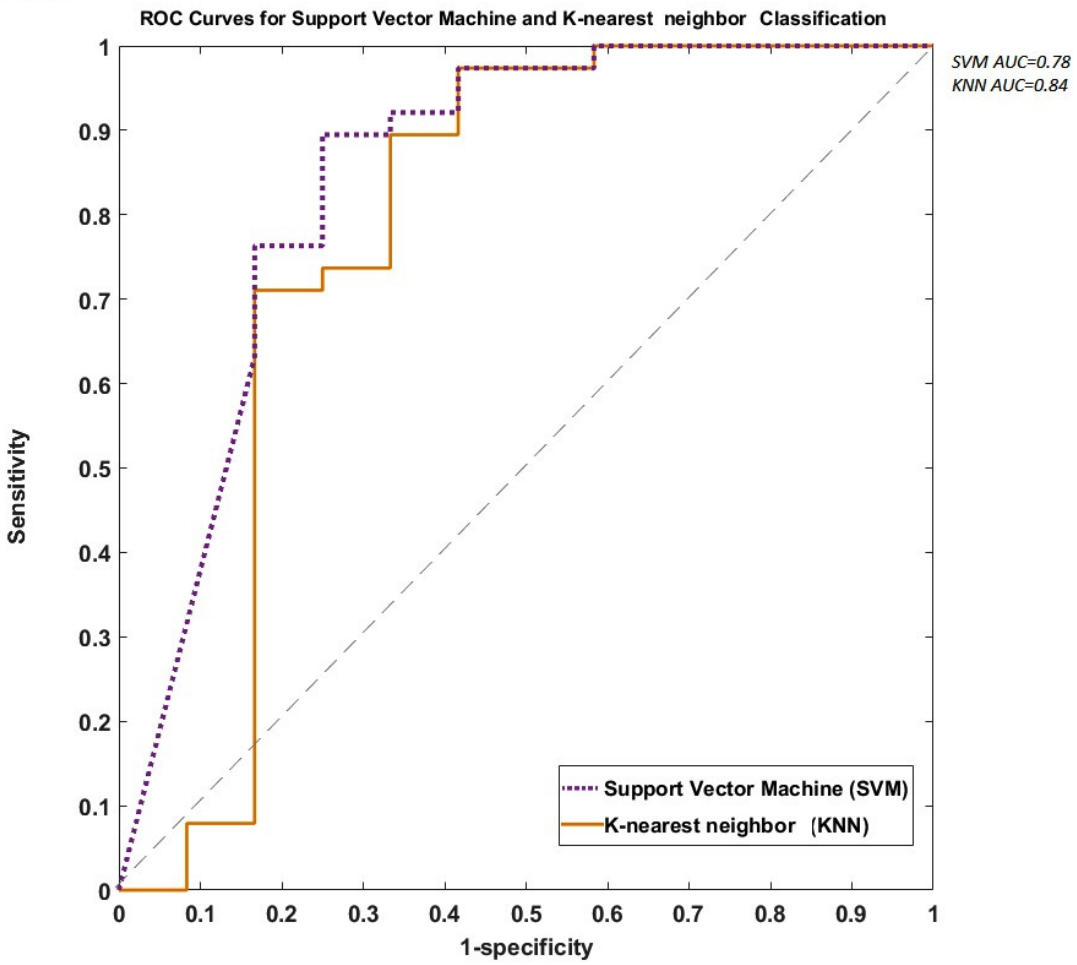


Figure 7. Radiologist's initial diagnosis compared with final diagnosis after histopathology for different tumor types, along with the comparison between Medical Image Region of interest analysis tool and Repository (MIROR) performance evaluated by support vector machine (SVM), k-nearest neighbor (KNN), and 2-step threshold classification methods.

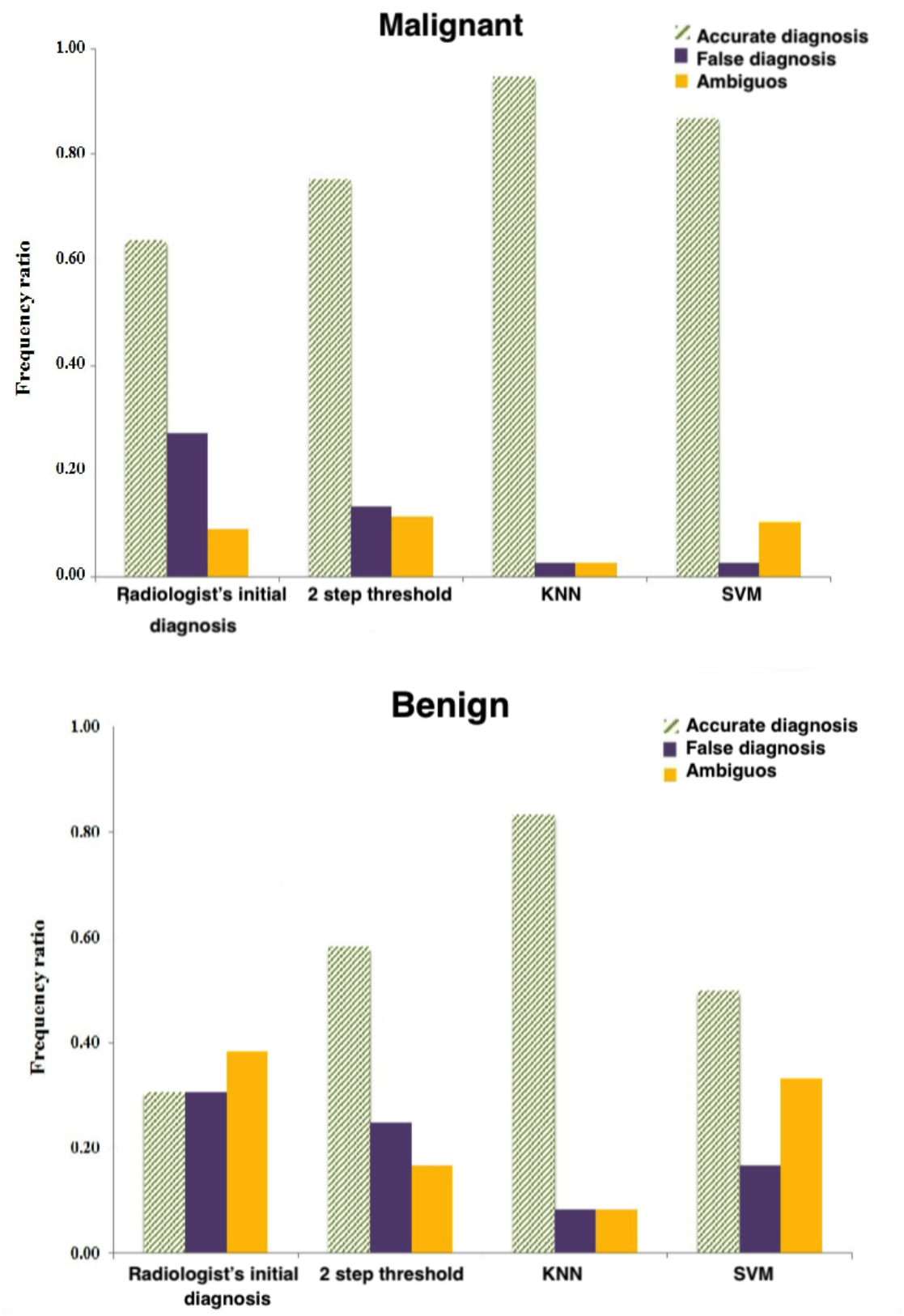
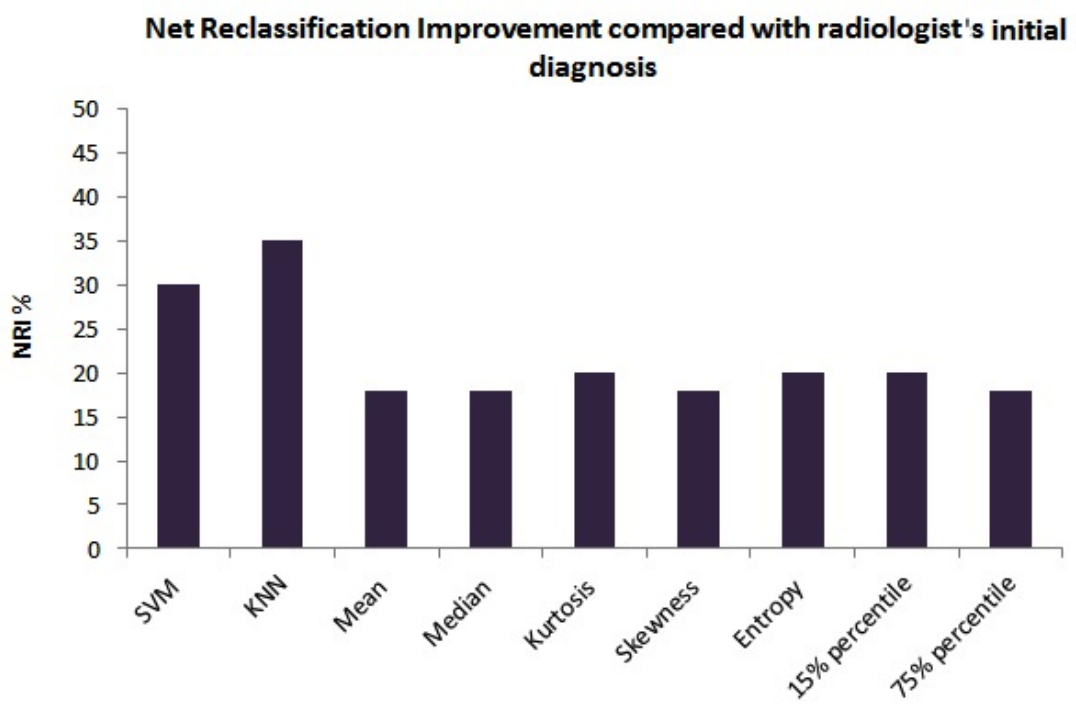


Figure 8. Net reclassification improvement (NRI) for k-nearest neighbor (KNN), support vector machine (SVM), and the 2-step thresholding methods compared with radiologist's initial read.



Discussion

Continuous developments of magnetic resonance systems have transformed this domain from a pure imaging system to a sophisticated precise metric system that generates a substantial amount of information and data. The complex structure of the clinical data generated often does not ease the difficulty in discriminating between different diagnoses and promotes adoption of intelligent CDSs [55,56]. CDSs allow clinicians and specialists to get insight into the data, test hypotheses, draw conclusions, and directly interact with all the available information. One should note that CDSs should aim to facilitate optimal human performance by harnessing the most advanced imaging and analysis techniques in conjunction with the end user's own decision-making skills and abilities.

Radiologists are moving toward quantitative imaging techniques that are difficult to apply and complex to interpret [16,57]. MIROR is a real-time CDS, which can guide clinicians through the implementation and analysis of advanced and new imaging techniques and allow for these new methodologies to find clinical acceptance through translational applications.

MIROR as a diagnostic tool allows its users to extract specific region morphological features, request specific quantified metrics and features (as a biomarker), and compare with relevant findings available in its repository to gain maximal statistical power with regard to outcome prediction for the input case into the support system. MIROR can direct users to refine their search patterns looking for particular diagnoses, even if they themselves are not immediately aware of the significance of these findings.

Use of modular programming in the development of MIROR enforces logical boundaries between magnetic resonance analysis applications, thereby improving maintainability [58,59]. Modularity has also allowed development and validation of individual analysis techniques in separate studies to ensure achievement of the important feature of any CDSs, which is its accuracy and appropriateness of the system's result.

In terms of its added clinical value and its impact on providing clinical evidence, MIROR will assist clinicians to better understand the pathophysiological difference between the different tumor types and provide information that could help them to better understand the mechanisms of diseases to improve the diagnosis and prognosis of tumors. MRI biomarkers provide information on both the tumor and its interaction with its environment and can potentially provide new information, which is not available from histology or tumor genetics. Analysis of cancer imaging big data will allow uncovering the relation and structure of cancer disease from an angle that has not previously been viewed.

Although we concentrate on developing these advanced MRI methods as a noninvasive diagnostic aid, they provide important information on tissue properties. Apparent diffusion coefficient shows a strong inverse correlation to cellularity—a key feature of tumors and tumor aggressiveness. Likewise, there is an increasing understanding that IVIM-*f* is related to tumor vascularity, which is again an important pathophysiological property of the tumor. Making these advanced MRI techniques available to clinicians in their multidisciplinary team meetings, where imaging and histopathology are evaluated together for individual cases, is an important goal and will allow an improved understanding of pathophysiology for these tumors in vivo and ex vivo.

Limitations

As a part of future research, we plan to work on functionality and intelligent scaling of quantitative applications of MIROR by further enhancing its statistical capabilities and extension to more embedded advanced quantitative analysis modules such as magnetic resonance spectroscopy and integration of real-time interactive machine learning to optimize the use of available data in MIROR based on the guidelines [60].

It should be noted that a major limitation of the study was that MIROR was trained and tested with a rather limited set of data analyzed retrospectively. A major focus of future work will be the validation of the tool based on a prospective dataset in real time and in a multicenter clinical setting to reinforce the credibility, usability, and efficiency of the proposed CDSs applications.

Conclusions

The proposed CDS is a diagnostic tool and repository that allows the interpretation and analysis of magnetic resonance images to be more accessible and comprehensive for clinicians. The process and experiences described here provide a model for development of the other CDSs attempting to perform a nonregion-specific quantitative analysis of MRI data. MIROR aims to increase clinician's skillset by introducing newer techniques and up-to-date findings to their repertoire and make information from previous cases available to aid decision making. The modular-based format of the tool allowed integration of analyses that are not readily available clinically, and streamlines future developments. Pipelines for new analysis applications are available or already in development and will be shortly available under the MIROR platform.

Acknowledgments

This study is funded through an NIHR Research Professorship, 13-0053. The authors acknowledge funding from the CRUK and EPSRC Cancer Imaging Programme at the Children's Cancer and Leukaemia Group in association with the MRC and Department of Health (England; C7809/A10342) and the Engineering and Physical Sciences Research Council (EPSRC, EP/F50053X/1). The authors would also like to acknowledge the magnetic resonance research radiographers at Birmingham Children's Hospital for scanning the patients in this study.

Authors' Contributions

NZ developed application of the software, implemented suggested interface components, enabled deployment of the system, identified problems requiring alteration and implemented them upon agreed revisions, conducted data analysis, and drafted the manuscript. EM implemented diffusion application algorithm, tested MIROR, and made suggestions for revisions of MIROR. KM evaluated MIROR, extracted patient data, assisted in drafting the manuscript, and made suggestions for revisions of MIROR. KF, KM, and AP provided guidance and recommendations for modification of the system for this clinical application and participated in early setup of the project. AP conceived of the study, wrote the original grant proposal, supervised development, made suggestions for revisions, and provided final approval of MIROR. All authors read, provided suggestions for revising, and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Smith KA, Carrino JA, Carrino J. MRI-guided interventions of the musculoskeletal system. J Magn Reson Imaging 2008 Feb;27(2):339-346. [doi: 10.1002/jmri.21274] [Medline: 18219687]
2. Morton G, Ishida M, Schuster A, Hussain S, Schaeffter T, Chiribiri A, et al. Perfusion cardiovascular magnetic resonance: comparison of an advanced, high-resolution and a standard sequence. J Cardiovasc Magn Reson 2012 Jun 09;14:34 [FREE Full text] [doi: 10.1186/1532-429X-14-34] [Medline: 22682016]
3. Hamirani YS, Kramer CM. Cardiac MRI assessment of myocardial perfusion. Future Cardiol 2014 May;10(3):349-358 [FREE Full text] [doi: 10.2217/fca.14.18] [Medline: 24976472]
4. Kwon T, Jeong IG, Yoo S, Lee J, Hong S, You D, et al. Role of MRI in indeterminate renal mass: diagnostic accuracy and impact on clinical decision making. Int Urol Nephrol 2015 Apr;47(4):585-593. [doi: 10.1007/s11255-015-0928-x] [Medline: 25681121]
5. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia 2009 Feb;11(2):102-125 [FREE Full text] [Medline: 19186405]
6. Patterson DM, Padhani AR, Collins DJ. Technology insight: water diffusion MRI--a potential new biomarker of response to cancer therapy. Nat Clin Pract Oncol 2008 Apr;5(4):220-233. [doi: 10.1038/ncponc1073] [Medline: 18301415]
7. Waldman AD, Jackson A, Price SJ, Clark CA, Booth TC, Auer DP, National Cancer Research Institute Brain Tumour Imaging Subgroup. Quantitative imaging biomarkers in neuro-oncology. Nat Rev Clin Oncol 2009 Aug;6(8):445-454. [doi: 10.1038/nrclinonc.2009.92] [Medline: 19546864]

8. O'Connor JP, Jackson A, Parker GJ, Roberts C, Jayson GC. Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. Nat Rev Clin Oncol 2012 Feb 14;9(3):167-177. [doi: 10.1038/nrclinonc.2012.2] [Medline: 22330689]
9. Mendelson DS, Rubin DL. Imaging informatics: essential tools for the delivery of imaging services. Acad Radiol 2013 Oct;20(10):1195-1212 [FREE Full text] [doi: 10.1016/j.acra.2013.07.006] [Medline: 24029051]
10. Shahmoradi L, Farzanehnejad AR, Arji G. Guideline-based clinical decision support systems as an inseparable tool for better cancer care management. Iran J Public Health 2016 Jul;45(7):962-963 [FREE Full text] [Medline: 27517010]
11. Turgeon J, Michaud V. Clinical decision support systems: great promises for better management of patients' drug therapy. Expert Opin Drug Metab Toxicol 2016 Sep;12(9):993-995. [doi: 10.1517/17425255.2016.1171317] [Medline: 27021287]
12. Wright A, Sittig DF. A four-phase model of the evolution of clinical decision support architectures. Int J Med Inform 2008 Oct;77(10):641-649 [FREE Full text] [doi: 10.1016/j.ijmedinf.2008.01.004] [Medline: 18353713]
13. Berrouiguet S, Barrigón ML, Brandt SA, Ovejero-García S, Álvarez-García R, Carballo JJ, MEmind Study Group, et al. Development of a Web-Based Clinical Decision Support System for Drug Prescription: Non-Interventional Naturalistic Description of the Antipsychotic Prescription Patterns in 4345 Outpatients and Future Applications. PLoS One 2016;11(10):e0163796 [FREE Full text] [doi: 10.1371/journal.pone.0163796] [Medline: 27764107]
14. Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine Learning and Decision Support in Critical Care. Proc IEEE Inst Electr Electron Eng 2016 Feb;104(2):444-466 [FREE Full text] [doi: 10.1109/JPROC.2015.2501978] [Medline: 27765959]
15. Schiff GD, Bates DW. Can electronic clinical documentation help prevent diagnostic errors? N Engl J Med 2010 Mar 25;362(12):1066-1069. [doi: 10.1056/NEJMp0911734] [Medline: 20335582]
16. Stivaros SM, Gledson A, Nenadic G, Zeng XJ, Keane J, Jackson A. Decision support systems for clinical radiological practice -- towards the next generation. Br J Radiol 2010 Nov;83(995):904-914 [FREE Full text] [doi: 10.1259/bjr/33620087] [Medline: 20965900]
17. Wang X, Bie R, Sun Y, Wu Z, Zhou M, Cao R, et al. The architecture of an automatic eHealth platform with mobile client for cerebrovascular disease detection. JMIR Mhealth Uhealth 2013 Aug 09;1(2):e20 [FREE Full text] [doi: 10.2196/mhealth.2550] [Medline: 25098861]
18. Billis AS, Papageorgiou EI, Frantzidis CA, Tsatali MS, Tsolaki AC, Bamidis PD. A decision-support framework for promoting independent living and ageing well. IEEE J Biomed Health Inform 2015 Jan;19(1):199-209. [doi: 10.1109/JBHI.2014.2336757] [Medline: 25073180]
19. Tsolaki E, Kousi E, Svolos P, Kapsalaki E, Theodorou K, Kappas C, et al. Clinical decision support systems for brain tumor characterization using advanced magnetic resonance imaging techniques. World J Radiol 2014 Apr 28;6(4):72-81 [FREE Full text] [doi: 10.4329/wjr.v6.i4.72] [Medline: 24778769]
20. Njie GJ, Proia KK, Thota AB, Finnie RK, Hopkins DP, Banks SM, Community Preventive Services Task Force. Clinical decision support systems and prevention: a community guide cardiovascular disease systematic review. Am J Prev Med 2015 Nov;49(5):784-795 [FREE Full text] [doi: 10.1016/j.amepre.2015.04.006] [Medline: 26477805]
21. Pesce F, Diciolla M, Binetti G, Naso D, Ostuni VC, Di Noia T, et al. Clinical decision support system for end-stage kidney disease risk estimation in IgA nephropathy patients. Nephrol Dial Transplant 2016 Jan;31(1):80-86. [doi: 10.1093/ndt/gfv232] [Medline: 26047632]
22. Bosch JG, Nijland F, Mitchell SC, Lelieveldt BP, Kamp O, Reiber JH, et al. Computer-aided diagnosis via model-based shape analysis: automated classification of wall motion abnormalities in echocardiograms. Acad Radiol 2005 Mar;12(3):358-367. [doi: 10.1016/j.acra.2004.11.025] [Medline: 15766696]
23. Pietka E, Kawa J, Badura P, Spinczyk D. Open architecture computer ‐ aided diagnosis system. Expert Syst 2010;27(1):17-39. [doi: 10.1111/j.1468-0394.2009.00524.x]
24. Ho TW, Huang CW, Lin CM, Lai F, Ding JJ, Ho YL, et al. A telesurveillance system with automatic electrocardiogram interpretation based on support vector machine and rule-based processing. JMIR Med Inform 2015 May 07;3(2):e21 [FREE Full text] [doi: 10.2196/medinform.4397] [Medline: 25953306]
25. Pietka E, Gertych A, Pospiechâ Euro Kurkowska S, Cao F, Huang HK, Gilzanz V. Computer-assisted bone age assessment: graphical user interface for image processing and comparison. J Digit Imaging 2004 Sep;17(3):175-188 [FREE Full text] [doi: 10.1007/s10278-004-1006-6] [Medline: 15175931]
26. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. IEEE Trans Med Imaging 2014 May;33(5):1083-1092. [doi: 10.1109/TMI.2014.2303821] [Medline: 24770913]
27. Sonka M, Bosch JG, Lelieveldt BP, Mitchell SC, Reiber JH. Computer-aided diagnosis via model-based shape analysis: cardiac MR and echo. Int Congr Ser 2003 Jun;1256:1013-1018. [doi: 10.1016/S0531-5131(03)00261-9]
28. Dorrius MD, Jansen-van der Weide MC, van Ooijen PM, Pijnappel RM, Oudkerk M. Computer-aided detection in breast MRI: a systematic review and meta-analysis. Eur Radiol 2011 Aug;21(8):1600-1608 [FREE Full text] [doi: 10.1007/s00330-011-2091-9] [Medline: 21404134]
29. Kim YJ, Park JW, Kim JW, Park C, Gonzalez JP, Lee SH, et al. Computerized automated quantification of subcutaneous and visceral adipose tissue from computed tomography scans: development and validation study. JMIR Med Inform 2016 Feb 04;4(1):e2 [FREE Full text] [doi: 10.2196/medinform.4923] [Medline: 26846251]

30. Pérez-Ruiz A, Julià-Sapé M, Mercadal G, Olier I, Majós C, Arús C. The INTERPRET Decision-Support System version 3.0 for evaluation of magnetic resonance spectroscopy data from human brain tumours and other abnormal brain masses. BMC Bioinformatics 2010 Nov 29;11:581 [FREE Full text] [doi: 10.1186/1471-2105-11-581] [Medline: 21114820]
31. Merkel B, Harz MT, Konrad O, Hahn HK, Peitgen HO. A novel software assistant for clinical analysis of MR spectroscopy with MeVisLab. In: Proceedings of SPIE, Volume 6915, 69152R. 2008 Presented at: Medical Imaging; 2008; San Diego, CA, United States URL: https://www.researchgate.net/profile/Markus_Wenzel2/publication/252297720_A_novel_software_assistant_for_clinical_analysis_of_MR_Spectroscopy_with_MeVisLab/links/53cf8a3f0cf2fd75bc59d6ea/A-novel-software-assistant-for-clinical-analysis-of-MR-Spectroscopy-with-MeVisLab.pdf [doi: 10.1117/12.772638]
32. Egger J, Tokuda J, Chauvin L, Freisleben B, Nimsky C, Kapur T, et al. Integration of the OpenIGTLink network protocol for image-guided therapy with the medical platform MeVisLab. Int J Med Robot 2012 Sep;8(3):282-290 [FREE Full text] [doi: 10.1002/rcs.1415] [Medline: 22374845]
33. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc 2001;8(6):527-534 [FREE Full text] [Medline: 11687560]
34. de Clercq PA, Blom JA, Korsten HH, Hasman A. Approaches for creating computer-interpretable guidelines that facilitate decision support. Artif Intell Med 2004 May;31(1):1-27. [doi: 10.1016/j.artmed.2004.02.003] [Medline: 15182844]
35. Thomas KW, Dayton CS, Peterson MW. Evaluation of internet-based clinical decision support systems. J Med Internet Res 1999;1(2):E6 [FREE Full text] [doi: 10.2196/jmir.1.2.e6] [Medline: 11720915]
36. Nejat F, El Khashab M, Rutka JT. Initial management of childhood brain tumors: neurosurgical considerations. J Child Neurol 2008 Oct;23(10):1136-1148 [FREE Full text] [doi: 10.1177/0883073808321768] [Medline: 18952580]
37. Sen A, Banerjee A, Sinha AP, Bansal M. Clinical decision support: converging toward an integrated architecture. J Biomed Inform 2012 Oct;45(5):1009-1017 [FREE Full text] [doi: 10.1016/j.jbi.2012.07.001] [Medline: 22789390]
38. Federau C, Meuli R, O'Brien K, Maeder P, Hagmann P. Perfusion measurement in brain gliomas with intravoxel incoherent motion MRI. AJNR Am J Neuroradiol 2014 Feb;35(2):256-262 [FREE Full text] [doi: 10.3174/ajnr.A3686] [Medline: 23928134]
39. Cho GY, Moy L, Zhang JL, Baete S, Lattanzi R, Moccaldi M, et al. Comparison of fitting methods and b-value sampling strategies for intravoxel incoherent motion in breast cancer. Magn Reson Med 2015 Oct;74(4):1077-1085 [FREE Full text] [doi: 10.1002/mrm.25484] [Medline: 25302780]
40. Hectors SJ, Wagner M, Besa C, Bane O, Dyvorne HA, Fiel MI, et al. Intravoxel incoherent motion diffusion-weighted imaging of hepatocellular carcinoma: is there a correlation with flow and perfusion metrics obtained with dynamic contrast-enhanced MRI? J Magn Reson Imaging 2016 Oct;44(4):856-864 [FREE Full text] [doi: 10.1002/jmri.25194] [Medline: 26919327]
41. Wirestam R, Brockstedt S, Lindgren A, Geijer B, Thomsen C, Holtås S, et al. The perfusion fraction in volunteers and in patients with ischaemic stroke. Acta Radiol 1997 Nov;38(6):961-964. [Medline: 9394649]
42. Luciani A, Vignaud A, Cavet M, Nhieu JT, Mallat A, Ruel L, et al. Liver cirrhosis: intravoxel incoherent motion MR imaging--pilot study. Radiology 2008 Dec;249(3):891-899. [doi: 10.1148/radiol.2493080080] [Medline: 19011186]
43. Chandarana H, Kang SK, Wong S, Rusinek H, Zhang JL, Arizono S, et al. Diffusion-weighted intravoxel incoherent motion imaging of renal tumors with histopathologic correlation. Invest Radiol 2012 Dec;47(12):688-696. [doi: 10.1097/RLI.0b013e31826a0a49] [Medline: 22996315]
44. Lemke A, Laun FB, Klauss M, Re TJ, Simon D, Delorme S, et al. Differentiation of pancreas carcinoma from healthy pancreatic tissue using multiple b-values: comparison of apparent diffusion coefficient and intravoxel incoherent motion derived parameters. Invest Radiol 2009 Dec;44(12):769-775. [doi: 10.1097/RLI.0b013e3181b62271] [Medline: 19838121]
45. Klauss M, Lemke A, Grünberg K, Simon D, Re TJ, Wente MN, et al. Intravoxel incoherent motion MRI for the differentiation between mass forming chronic pancreatitis and pancreatic carcinoma. Invest Radiol 2011 Jan;46(1):57-63. [doi: 10.1097/RLI.0b013e3181fb3bf2] [Medline: 21139505]
46. Meeus EM, Novak J, Withey SB, Zarinabad N, Dehghani H, Peet AC. Evaluation of intravoxel incoherent motion fitting methods in low-perfused tissue. J Magn Reson Imaging 2017 May;45(5):1325-1334 [FREE Full text] [doi: 10.1002/jmri.25411] [Medline: 27545824]
47. Humphries PD, Sebire NJ, Siegel MJ, Olsen OE. Tumors in pediatric patients at diffusion-weighted MR imaging: apparent diffusion coefficient and tumor cellularity. Radiology 2007 Dec;245(3):848-854. [doi: 10.1148/radiol.2452061535] [Medline: 17951348]
48. Abe S. Support Vector Machines for Pattern Classification, 2nd ed. London; New York: Springer; 2010.
49. Zarinabad N, Wilson M, Gill SK, Manias KA, Davies NP, Peet AC. Multiclass imbalance learning: improving classification of pediatric brain tumors from magnetic resonance spectroscopy. Magn Reson Med 2017 Jun;77(6):2114-2124 [FREE Full text] [doi: 10.1002/mrm.26318] [Medline: 27404900]
50. Woo S, Cho JY, Kim SY, Kim SH. Histogram analysis of apparent diffusion coefficient map of diffusion-weighted MRI in endometrial cancer: a preliminary correlation study with histological grade. Acta Radiol 2014 Dec;55(10):1270-1277. [doi: 10.1177/0284185113514967] [Medline: 24316663]

51. Xu XQ, Hu H, Su GY, Liu H, Hong XN, Shi HB, et al. Utility of histogram analysis of ADC maps for differentiating orbital tumors. Diagn Interv Radiol 2016;22(2):161-167 [FREE Full text] [doi: 10.5152/dir.2015.15202] [Medline: 26829400]
52. Suo ST, Chen XX, Fan Y, Wu LM, Yao QY, Cao MQ, et al. Histogram analysis of apparent diffusion coefficient at 3.0 T in urinary bladder lesions: correlation with pathologic findings. Acad Radiol 2014 Aug;21(8):1027-1034. [doi: 10.1016/j.acra.2014.03.004] [Medline: 24833566]
53. Bull JG, Saunders DE, Clark CA. Discrimination of paediatric brain tumours using apparent diffusion coefficient histograms. Eur Radiol 2012 Feb;22(2):447-457. [doi: 10.1007/s00330-011-2255-7] [Medline: 21918916]
54. Jewell ES, Maile MD, Engoren M, Elliott M. Net reclassification improvement. Anesth Analg 2016 Mar;122(3):818-824. [doi: 10.1213/ANE.0000000000001141] [Medline: 26891393]
55. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, Aerts HJ, Bendriem B, et al. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology 2011 Jun;259(3):875-884 [FREE Full text] [doi: 10.1148/radiol.10100800] [Medline: 21325035]
56. Sullivan DC. Imaging as a quantitative science. Radiology 2008 Aug;248(2):328-332. [doi: 10.1148/radiol.2482080242] [Medline: 18641239]
57. Peh WC. Pitfalls in Diagnostic Radiology. Verlag Berlin Heidelberg: Springer; 2015:978-973.
58. Hughes J. Why functional programming matters. Comput J 1989;32(2):98-107. [doi: 10.1093/comjnl/32.2.98]
59. Parnas DL. On the criteria to be used in decomposing systems into modules. Commun ACM 1972;15(12):1053-1058. [doi: 10.1145/361598.361623]
60. 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]

Abbreviations

ADC: apparent diffusion coefficient
AUC: area under the curve
CADs: computer-assisted diagnosis tools
CDS: clinical decision support
DICOM: digital imaging and communications in medicine
DWI: diffusion-weighted imaging
IVIM: intravoxel incoherent motion
KNN: k-nearest neighbor
MIROR: medical image region of interest analysis tool and repository
MRI: magnetic resonance imaging
PACS: picture archiving and communication system
ROI: region of interest
SVM: support vector machine

Edited by G Eysenbach; submitted 25.10.17; peer-reviewed by CS Hung, JJ Ding; comments to author 17.12.17; revised version received 10.01.18; accepted 26.01.18; published 02.05.18.

Please cite as:
Zarinabad N, Meeus EM, Manias K, Foster K, Peet A
Automated Modular Magnetic Resonance Imaging Clinical Decision Support System (MIROR): An Application in Pediatric Cancer Diagnosis
JMIR Med Inform 2018;6(2):e30
URL: http://medinform.jmir.org/2018/2/e30/
doi:10.2196/medinform.9171
PMID:29720361

Original Paper

Capturing a Patient-Reported Measure of Physical Function Through an Online Electronic Health Record Patient Portal in an Ambulatory Clinic: Implementation Study

Jing Li[1], MPH; Jinoos Yazdany[1], MD, MPH; Laura Trupin[1], MPH; Zara Izadi[1], Pharm D; Milena Gianfrancesco[1], PhD, MPH; Sarah Goglin[1], MD; Gabriela Schmajuk[1,2], MD, MSc

[1]Division of Rheumatology, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States
[2]VA Medical Center, San Francisco, CA, United States

Corresponding Author:
Jing Li, MPH
Division of Rheumatology
Department of Medicine
University of California, San Francisco
Suite 270
3333 California Street
San Francisco, CA, 94118
United States
Phone: 1 415 290 4997
Fax: 1 415 476 9030
Email: jing.li2@ucsf.edu

Abstract

Background: Despite significant interest in the collection of patient-reported outcomes to make care more patient-centered, few studies have evaluated implementation efforts to collect patient-reported outcomes from diverse patient populations

Objective: We assessed the collection of patient-reported outcomes from rheumatoid arthritis patients in an academic rheumatology clinic, using a paper and an online form through the electronic health record patient portal.

Methods: We identified patients seen between 2012-2016 with ≥2 face-to-face encounters with a rheumatology provider and International Classification of Diseases codes for RA, ≥30 days apart. In 2013, our clinic implemented a paper version of the Patient Reported Outcome Measurement Information System (PROMIS) physical function form that was administered to patients upon their check-in at the clinic. In 2015, an online version of the form became available by way of the electronic health record patient portal to patients with active portal accounts. We compared the proportion of visits with documented PROMIS scores across age, race and ethnicity, and language and examined trends over time using a control chart.

Results: We included 1078 patients with rheumatoid arthritis with 7049 in-person encounters at the rheumatology clinic over 4 years, with an average of 168 visits per month. Of the included patients, 80.4% of patients (867/1078) were female and the mean age was 58 (SD 16) years. The overall PROMIS physical function score documentation increased from 60.4% (1081/1791) of visits in 2013 to 74.4% (905/1217) of visits in 2016. Online score documentation increased from 10.0% (148/1473) in 2015 to 19.3% (235/1217) in 2016. African American patients were least likely to have a PROMIS physical function score recorded (55/88, 62.5% compared to 792/990, 80.0% for other racial or ethnic groups; P<.001). Compared with white patients, both African American and Hispanic patients were less likely to have active online electronic health record portal accounts (44/88, 50% and 90/157, 57.3% respectively, compared to 437/521, 83.9% of white patients; P<.001) and, once activated, less likely to use the online survey (6/44, 13.6% and 16/90, 17.8% respectively, compared to 135/437, 30.9% of white patients; P=.02). There was no significant difference in the proportion of any PROMIS physical function forms recorded between non-English vs English preferred patients. No significant differences were found across age or gender.

Conclusions: PROMIS physical function form completion improved overall from 2012-2016 but lagged among racial and ethnic minorities and non-English preferred patients. Future studies should address issues of portal access, enrollment, satisfaction, and persistence and focus on developing PRO implementation strategies that accommodate the needs and preferences of diverse populations.

(JMIR Med Inform 2018;6(2):e31) doi:10.2196/medinform.8687

KEYWORDS

electronic health record; patient-reported outcomes; rheumatoid arthritis

Introduction

The effective use of patient-reported outcomes (PROs) data is anticipated to play a critical role in improving health care delivery, patient experiences with care, and outcomes. In rheumatoid arthritis (RA), a complex chronic condition characterized by joint pain and inflammation, validated PROs have been used over the past several decades to assess levels of RA disease activity and functional status [1,2]. PROs have successfully informed treatment decisions and facilitated shared decision-making, patient engagement and goal-setting in RA [3-6]. Routine assessment of PROs is now recommended by American College of Rheumatology (ACR) guidelines, and quality measures to encourage the regular collection of RA PROs have been endorsed by the National Quality Forum [7,8]. Despite significant interest in the collection of PROs to make care more patient-centered, few studies have evaluated implementation efforts to collect PROs in real-world practice settings that serve diverse patient populations [9].

Different approaches to collecting RA PROs have been used, including paper questionnaires, telephone interviews, and, more recently, digital health approaches such as electronic health record (EHR) patient portals. Online patient portal-based collection of PROs is appealing in health care settings because, by utilizing the existing infrastructure of the EHR, they have the potential to decrease the burden of data collection and entry for clinic staff and providers. In addition, PRO information collected through the EHR enables tracking of outcomes at the individual patient level over time, a feature that is useful for patients and providers in assessing whether a key treatment goal, maintaining functional capacity, is being achieved.

In this study, we assessed the proportion of RA patients who completed a physical function PRO form prior to an in-person visit, and the fraction of those who used the online EHR portal to report PROs once that functionality was implemented. Because we hypothesized that socio-demographic factors might influence how patients chose to complete PRO surveys, we also examined completion patterns by age, gender, race and ethnicity, and preferred language. Finally, we describe the challenges encountered in implementing RA PROs in our health system.

Methods

Clinical Context and Workflows

The University of California, San Francisco (UCSF) is an academic health center with over 750,000 outpatient visits per year that uses an Epic EHR system. The catchment area includes much of northern California. The UCSF rheumatology clinic incorporated physical function assessments into its routine practice February 1, 2013. Workflows were designed such that a paper version of the PROMIS PF survey, the PF-10a, was given to patients upon check in to each of their return visits, with instructions to complete the form in the waiting room prior to their clinical encounter. When patients were called from the waiting room to perform vital signs, medical assistants would calculate the PROMIS score from the paper survey and input the raw score into a flowsheet in the EHR. The raw score is automatically converted into a T-score, a standardized score on a relevant reference population, where 50 is the mean and 10 is the standard deviation (SD) of that population [10].

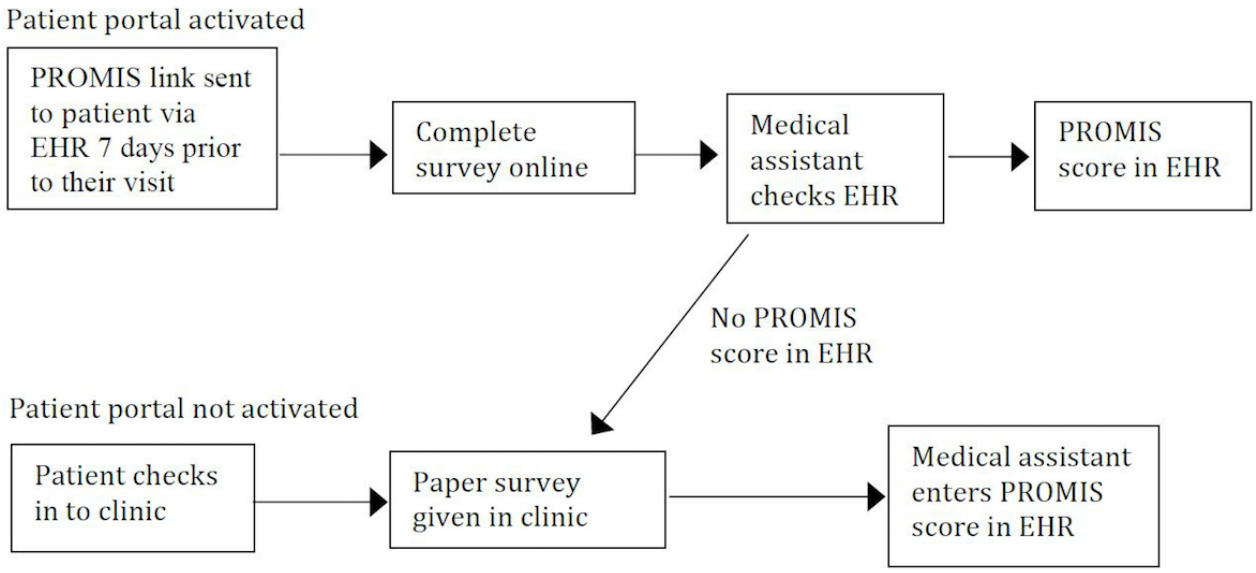
Because there was interest in automating PRO measure collection through the EHR, we collaborated with our institutional health information technology specialists to implement PROMIS PF collection through our Epic system's patient portal, called MyChart. Patients received an activation code at any in-person visit at UCSF that would allow them to log on to the online patient portal. Once the portal was activated, patients received a MyChart message from their provider 7 days prior to their appointment with a link to the PROMIS PF form (PF-12a). This workflow was implemented January 1, 2015 and medical assistants looked for a recorded online PROMIS survey score in the EHR when a patient checked in to the clinic. If a patient had a documented PROMIS PF score through the patient portal within 7 days of their appointment, they would be instructed to skip the paper-based PROMIS PF survey. PROMIS scores collected through the online portal are also automatically converted to T-score and input into a flowsheet within the EHR (Figure 1).

Patients and Data Source

The UCSF Committee on Human Research approved this study.

Data was derived from our Epic Clarity Data Warehouse and included the denominator of patients who had two or more ICD9 codes (714.0) for RA (at least 30 days apart) between June 1, 2012 and July 31, 2016. Information on the demographics for each patient was extracted (age, gender, self-reported race and ethnicity, and preferred language), online EHR portal activation status as of January 1, 2015 (the date the PROMIS PF became available through the online EHR portal), and, by the end of the study period, rheumatology visit encounter dates, disease activity scores by way of Clinical Disease Activity Index , PROMIS physical function (PF) scores, and PROMIS PF survey completion method (paper or EHR patient portal). Follow-up time was defined as the total number of months between each patient's first and last visits during the study period.

Figure 1. Flow of PROMIS documentation in the clinic. EHR: electronic health record; PROMIS: Patient Reported Outcome Measurement Information System.



PROMIS PF Measures

The PROMIS PF survey was developed by the National Institute of Health [10]. Raw scores range from 10 to 50 and can be transformed into a T-score to compare a given patient's score to the US general population mean (mean 50, SD 10). In this study, we examined use of the PF-10a (the short form 10-item paper questionnaire), which is available in multiple languages, including English, French, Spanish, and Chinese. In the EHR patient portal, we implemented the PF-12a in English only on January 1, 2015. PF-12a is a short form physical function questionnaire revised from PF-10a. PF-12a raw scores range from 12 to 60 but can be compared directly with PF-10a scores using T-scores [11].

Primary Analysis

We examined the proportion of patients completing the PROMIS PF-10a after the paper form was implemented in 2013, and the proportion of patients completing the form electronically in 2015 after the online EHR patient portal PRO form was implemented. We used descriptive statistics to summarize age, gender, race and ethnicity, language preference, and online EHR portal activation. This data, and the relationships between patient characteristics and PROMIS PF completion method, mean PROMIS PF score, and mean disease activity score (when available) were examined using analysis of variance (ANOVA) for continuous variables or chi-square tests for categorical variables.

We tested for a potential interaction between age and race and ethnicity in a logistic regression model with portal activation as outcome, categorized age (≥70 and <70), race and ethnicity and cross-product terms as independent variables. In addition, we calculated the proportion of visits per month in which either a paper (PF-10a) or online EHR portal (PF-12a) PROMIS score was recorded in the EHR from 2013 to 2016. The frequencies of PROMIS PF completed by 1) either online EHR portal or paper and 2) the online EHR portal only were plotted monthly on a quality control chart (p-chart) [12]. We calculated the proportion of patients who used and persistently used the online EHR portal to complete PROMIS PF forms. Spearman correlation coefficients were used to test the correlation between the paper and the online EHR portal PROMIS PF T-scores.

Paper versus Online PROMIS PF Survey Score Correlation

Because some patients were filling out the paper survey, while others were filling out the online EHR portal survey, and still others were switching between methods, we were interested in understanding the correlation of PRO scores when assessed by different means within a short time period. A temporary system error that occurred during a 3-month period in 2016 allowed us to examine this correlation. During this period, due to a bug in a system upgrade, patients' online EHR portal PROMIS PF scores were not visible to clinic staff at the time of patient check in (N=51 encounters). These patients were asked to complete the paper PROMIS PF in clinic in addition to the online EHR PROMIS PF survey they had already completed. Additionally, there were multiple occasions in which the online EHR portal and paper PROMIS PF were completed within 7 days by the same patient for unclear reasons (N=209 encounters). In the main analysis, paper surveys that were completed within 7 days of online EHR surveys were deleted from the analysis, and only the online EHR portal score was counted (117/1446, 8.1% of paper surveys in 2015; 143/907,15.8% of paper surveys) in 2016). However, these paper and online EHR scores within 7 days of each other provided an opportunity to compare scores across collection methods. When both scores were present, we calculated the proportion of individuals with floor (defined as worst) and ceiling (defined as best) scores for PF-10a and PF-12s and compared these proportions using a *t*-test [10].

Analyses were performed using Stata 14. For all analyses, *P* values <.05 were used as the criterion for statistical significance.

Results

We included 1078 RA patients with 7049 in-person encounters in the UCSF rheumatology clinic from June 1, 2012 to July 31, 2016. Of the patients, 80.4% (867/1078) were female, with a mean age of 58 years (SD 16; see Table 1). This group was racially and ethnically diverse and 557/1078 (51.7%) identified themselves as non-white, and 150/1078 (13.9%) reported a language other than English as their preferred language, primarily Chinese or Spanish. EHR portal account activation was a required step to complete the online portal PROMIS survey, although most patients who completed exclusively paper surveys (412/627) also had active accounts (65.7%).

Of all the patients, 78.6% (847/1078) had at least one PROMIS PF score recorded during the study period. There was an average of 168 RA in-person encounters per month over the study period. The proportion of visits with any documented PROMIS PF score increased over time, from 60.4% (1081/1791) in 2013 to 74.4% (905/1217) in 2016. In 2015 (after the online EHR portal PROMIS PF survey became available), 10.0% (148/1473) of visits had an associated online EHR portal PROMIS PF score, rising to 19.3% (235/1217) in 2016 (see Figure 2).

We explored patient factors associated with method of completion of the PROMIS PF (see Table 1). African American patients were less likely to have any PROMIS PF recorded compared with other groups (55/88, 62.5% compared to 792/990 80.0% for other racial and ethnic groups; $P<.001$). Both African American and Hispanic patients were less likely to have active EHR portal accounts compared with white patients (44/88, 50.0% and 90/157, 57.3% 90/157 respectively, compared to 437/521, 83.9% 437/521 of white patients; $P<.001$) and once activated, less likely to have completed the online PROMIS survey (6/44, 13.6% and 16/90, 17.8% respectively, compared to 135/437, 30.9% of white patients; $P=.02$). There was no significant difference in the proportion of any PROMIS PF recorded between non-English vs English preferred patients. However, non-English preferred patients were less likely to have active EHR portal accounts (79/150, 47.3% vs 734/928, 79.1% of English preferred patients; $P<.001$) or to use the online survey (10/71, 14.1% vs 210/734, 28.6% of English preferred patients; $P=.03$), likely because the online survey existed only in English.

Patients who had completed at least one online EHR portal PROMIS PF survey had 3.4 points higher mean PROMIS T-score, although this difference was not statistically significant (42.3 vs 38.9; $P=.7$). Patients who only completed the paper PROMIS PF survey had a significantly higher mean disease activity score (12.9 vs 10.4; $P<.001$). We also found that patients with longer follow-up and more visits per year were more likely to have any PROMIS PF score recorded ($P<.001$). No significant differences in PROMIS completion rate overall or online survey use were found across age or gender.

There were 775 (775/1078, 71.9%) patients who activated their online patient portal, 84.5% of which were activated before January 2015, when the online PROMIS PF survey was available. This rate did not vary by group (paper only, online, no PROMIS). There were no significant differences in patient characteristics when we compared patients who contributed visits before versus after January 2015 (when the online PROMIS survey became available, data not shown). We used multivariate logistic regression to assess the possibility of an interaction between age and race and ethnicity in the patients' portal activation status. We found that both age and race and ethnicity were associated with significant differences (younger, Caucasian patients were more likely to have active portals, $P<.001$ and $P<.001$, respectively), but that there was no age-race and ethnicity interaction.

Only 220 patients completed the online PROMIS PF survey at least once during the study period. We examined whether patients who completed the online survey once continued to use the online version of the PROMIS survey for future visits. Of these 220 patients, 84 (38.2%) used the online survey intermittently; 112 (112/220, 50.9%) used the online survey only once and reverted to paper surveys thereafter, and the remainder had no additional visits. We found no patients who used the online EHR portal exclusively (after its implementation) over time.

During the 3-month period in which an EHR programming error resulted in clinic staff not being able to view the online EHR portal PROMIS PF scores at the time of patient check-in, 51 patients completed the paper PROMIS PF in the clinic even though they had already completed the online EHR portal version in the days prior to the appointment. We found an additional 157 patients (209 encounters) who completed both paper and online EHR portal scores within 7 days of each other at some point during the study period. This gave us the opportunity to compare paper and online PROMIS scores from the same patients, for the same visit. Among these patients, the online EHR portal PROMIS PF score had a mean T-score of 40.5 (SD 10.7, range 16.1-66.4), and the paper PROMIS PF score had a mean T-score of 43.2 (SD 9.6, range 23.4-61.7), with a Spearman's correlation of $r=0.68$ ($P<.001$). The paper PROMIS PF had a significant 2.7 point higher mean T-score by paired t-test ($P<.001$).

Table 1. Patient characteristics by PROMIS survey completion methods, n (%).

	All Patients (n=1078)	Paper Survey Only (n=627)	At Least 1 Online EHR[a] Portal Survey (n=220)	No PROMIS[b] Survey Completed (n=231)	*P* value[c]
Age (years), mean (SD)	58 (16)	58 (16)	56 (15)	58 (17)	.7
Gender, n (%)					**.4**
Female	867 (80.4)	508 (81.0)	181 (82.3)	178 (77.1)	
Male	211 (19.6)	119 (19.0)	39 (17.7)	53 (22.9)	
Race or Ethnicity, n (%)					**<.001**
White	521 (48.3)	281 (44.8)	135 (61.4)	105 (45.4)	
African American	88 (8.2)	49 (7.8)	6 (2.7)	33 (14.3)	
Asian	149 (13.8)	103 (16.4)	26 (11.8)	20 (8.7)	
Hispanic	157 (14.6)	105 (16.8)	16 (7.3)	36 (15.6)	
Other or Multiple	163 (15.1)	89 (14.2)	37 (16.8)	37 (16)	
Preferred Language, n (%)					**<.001**
English	928 (86.1)	519 (82.8)	210 (95.4)	199 (86.2)	
Other[d]	150 (13.9)	108 (17.2)	10 (4.6)	32 (13.8)	
Online EHRs portal activated prior to January 2015, n (%)	655 (60.8)	343 (54.7)	189 (85.9)	123 (53.2)	<.001
Online EHR portal activated prior to end of study period, n (%)	775 (71.9)	412 (65.7)	220 (100.0)	143 (61.9)	<.001
Visits per patient, per year, mean (SD)	3.8 (2.0)	4.0 (1.9)	4.5 (1.8)	2.4 (2.0)	<.001
Follow-up months, mean (SD)	23 (0.5)	24 (0.6)	30 (0.8)	10 (1)	.05
Disease activity score (CDAI)[e], mean (SD)	12.0 (11.0)	12.9 (11.4)	10.4 (10.1)	19.4 (3.2)	<.001
PROMIS scores, mean (SD)	40.1 (10.8)	38.9 (11.1)	42.3 (10.0)[f]	N/A[g]	.7

[a]EHR: electronic health record.

[b]PROMIS: Patient Reported Outcome Measurement Information System.

[c]*P* values were tested by ANOVA tests for continuous variables or chi-square for categorical variables.

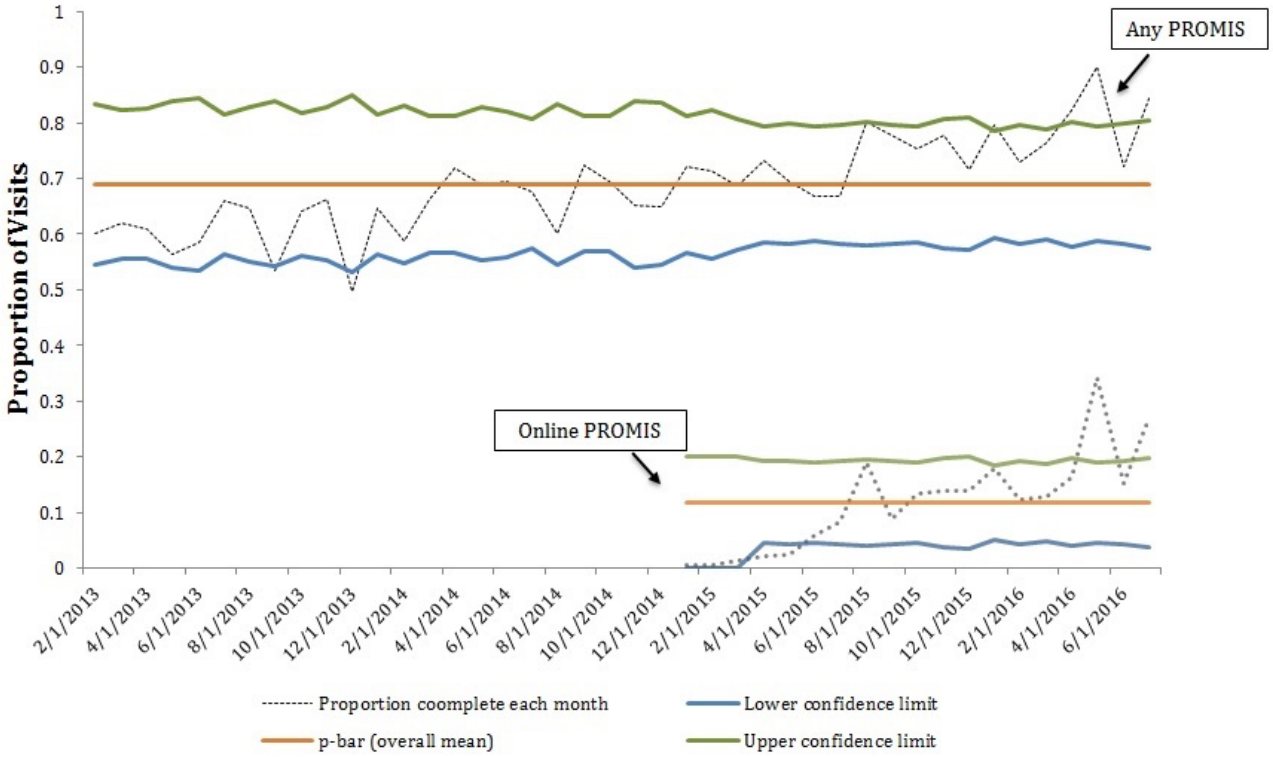
[d]Other included unknown or declined, n=5.

[e]681 patients with Clinical Disease Activity Index (CDAI) recorded: 480 paper survey only; 198 at least one online survey; 3 no survey completed.

[f]260 paper records were excluded: 209 PROMIS Physical Function scores because of duplicate online score from the same visit; 51 scores because of the temporary systematic error.

[g]Not applicable.

Figure 2. Percent completion of PROMIS by month. Both overall PROMIS and online PROMIS completion are presented. The black dotted line shows the proportion of patient encounters with a documented PROMIS by month. The upper control and low control limits vary as the denominator of patient encounters changed every month. The p-bar (overall mean) shows the average PROMIS documentation rate. More than 6 values are seen above the p-bar over the time, indicating a positive improvement in both overall PROMIS and online PROMIS documentation.



Discussion

To our knowledge, this is the first study to describe utilization of an online EHR patient portal for the collection of PROs as part of the routine care of a population of patients with RA. We found that over a 4-year period, overall PROMIS PF survey completion increased by 23.2%. Still, implementation was suboptimal, with more than 20.0% of in-person encounters lacking a PROMIS PF score. PROMIS measurement lagged particularly among racial and ethnic minorities. The proportion of patients completing PROMIS PF surveys through the online EHR portal almost doubled from the time of its implementation in 2015, yet only accounted for only 19.3% of the all PROMIS PF measurements. We found no persistence in the use of the online EHR portal for PROMIS reporting over time.

Few studies have described PRO collection via an online portal. One study in an ambulatory cancer care setting described a very successful implementation of electronic collection of PROs to measure common cancer symptoms (including physical function). PROs were measured by way of an online portal survey prior to clinical visits and were found to be an effective basis for referral for psychosocial and supportive care [13]. However, this study included only patients who activated and enrolled in the online portal and did not provide information on patients unable or unwilling to do so.

Our study found that most patients had an active online portal account before the PROMIS PF survey was available by way of the online EHR portal in January 2015, although only a small fraction completed the online EHR portal PROMIS survey, and that use lagged among racial and ethnic minorities. There are multiple possible reasons for this difference by race and ethnicity. First, there may be a lack of support and training to assist and encourage patients with portal activation, which is crucial to the successful implementation of an online measurement system [14]. Our findings are consistent with prior studies showing that racial and ethnic minorities are generally less likely to enroll and utilize online EHR portals [15-18]. Second, qualitative studies exploring possible reasons for reduced use of online EHR portals among African American and Hispanic patients have highlighted technical barriers and worries that use of a portal could undermine in-person relationships with healthcare providers [19]. Third, language proficiency may have hindered online EHR portal access—only 14.1% of patients with a preferred language other than English completed the online survey, compared to 28.6% of patients preferring English. Providing materials including enrollment information, websites and surveys, in other languages and for low literacy patients will be an important advance for increasing online EHR portal use [20,21]. Interestingly, although prior studies have reported that older adults from racial and ethnic minorities are significantly less likely to access and use an online patient portal [22], we found no evidence of a significant interaction between age and race and ethnicity on online EHR portal activation use in our study.

We found that patients who had completed at least one online survey seemed to have a lower disease burden compared to patients who completed only paper surveys (as evidenced by

marginally better functional status and lower disease activity scores). Although this study wasn't designed to study why this was the case, we hypothesize that completing the online survey may be more burdensome or difficult for patients with more active disease. Future qualitative studies could investigate this further.

Perhaps of greatest concern, our study found that no patients used the online EHR portal consistently for PRO measure collection over time; over half of patients abandoned the online portal for PRO reporting after a single use. To our knowledge, this lack of persistence in use has not been reported previously. Reasons for this could be multiple, including the lack of integration into routine care progress or dissatisfaction with the online PRO collection process or time constraints. Disruptions in the EHR that resulted in clinic staff not being able to access online PRO scores due to a system error, necessitating double entry of PRO scores, may have frustrated patients and caused them to abandon online EHR portal use. Future work on online EHR portals should focus both on barriers to enrollment and barriers to persistence in use.

In our comparison of paper and online PROMIS PF T-scores for the same patient and same visit, we found moderate levels of correlation between the PF-10a (paper) and PF-12a (online) versions (r=0.68; P<.001). Although it is valid to compare results of these 2 scores [10], the small difference that we detected is likely due to a difference in psychometric properties of the PF-10a and the PF-12a. Specifically, there was a significant difference in ceiling effects between the two versions, with the paper version having a higher ceiling effect (8.9% vs 0%; P<.001). No floor effects were observed.

The limitations of this study include a lack of information about patients' internet and computer access, and lack of a measure of education or healthy literacy level. Studies have shown that patients with low health literacy experience basic technological barriers such as difficulty using a mouse or finding specific keys on the keyboard in addition to "routine" technological barriers such as mistyping and navigation issues [23]. Future assessments should aim to capture this important variable.

Our study shows that the potential benefits of online EHR portal collection of PROs have not yet been realized in the UCSF rheumatology clinic. One goal of online collection was to decrease the burden of data collection and data entry. In our case, the clinic workflow still required that medical assistants assess whether a patient already completed an online version of the survey at the time of patient check-in. This process was both time-consuming and faulty (as evidenced by the numerous patients who completed both online and paper surveys within 7 days of an in-person visit). Another goal of online PRO collection was to enable automated tracking of outcomes over time. However, 1.5 years after implementation, uptake of the online EHR portal is poor at only 19.3% of visits. Further study is needed to investigate and address these issues, and pragmatic trials should test strategies to optimize the collection of PROs by addressing patient, provider, and health-system factors.

In summary, despite increasing completion of PROMIS PF surveys, we found only a fraction of patients who were offered online EHR portal use completed their PROMIS PF survey online, and none used it persistently. Disparities exist across race and ethnicity and language in access to the online EHR portal and in PRO completion once the portal is activated. Future studies should address issues of portal access, enrollment, satisfaction and persistence, and focus on developing PRO implementation strategies that accommodate the needs and preferences of diverse populations.

Acknowledgments

This work is supported by the Agency for Healthcare and Research Quality [R01 HS024412] and the National Institutes of Health [K23 AR063770 (GS)]. Drs. Yazdany and Schmajuk are also supported by the Russell/Engelman Medical Research Center for Arthritis and an independent research grant from Pfizer. Dr. Yazdany is supported by the Robert L. Kroc Chair in Rheumatic and Connective Tissue Diseases (I). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl 1994 Sep;41:86-89. [Medline: 7799394]
2. van Tuyl LHD, Boers M. Patient-reported outcomes in core domain sets for rheumatic diseases. Nat Rev Rheumatol 2015 Dec;11(12):705-712. [doi: 10.1038/nrrheum.2015.116] [Medline: 26324860]
3. Malysheva O, Bedrich A, Kuipers JG, Kleine H, Wolff B, Baerwald CG. Use of clinical scores to guide therapeutic decisions in patients with rheumatoid arthritis in daily care. Clin Exp Rheumatol 2015;33(2):255-258. [Medline: 25896474]
4. Ovretveit J, Keller C, Hvitfeldt FH, Essén A, Lindblad S, Brommels M. Continuous innovation: developing and using a clinical database with new technology for patient-centred care--the case of the Swedish quality register for arthritis. Int J Qual Health Care 2013 Apr;25(2):118-124. [doi: 10.1093/intqhc/mzt002] [Medline: 23360809]

5. Newman ED, Lerch V, Billet J, Berger A, Kirchner HL. Improving the quality of care of patients with rheumatic disease using patient-centric electronic redesign software. Arthritis Care Res (Hoboken) 2015 Apr;67(4):546-553 [FREE Full text] [doi: 10.1002/acr.22479] [Medline: 25417958]
6. Lavallee DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD, et al. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. Health Aff (Millwood) 2016 Apr;35(4):575-582. [doi: 10.1377/hlthaff.2015.1362] [Medline: 27044954]
7. Singh JA, Saag KG, Bridges SL, Akl EA, Bannuru RR, Sullivan MC, American College of Rheumatology. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Care Res (Hoboken) 2016 Jan;68(1):1-25 [FREE Full text] [doi: 10.1002/acr.22783] [Medline: 26545825]
8. National Quality Forum . NQF-Endorsed measures for musculoskeletal conditions. 2015 URL: http://www.qualityforum.org/Publications/2015/01/NQF-Endorsed_Measures_for_Musculoskeletal_Conditions.aspx [accessed 2017-08-10] [WebCite Cache ID 6sce0S7VK]
9. Harle CA, Lipori G, Hurley RW. Collecting, Integrating, and Disseminating Patient-Reported Outcomes for Research in a Learning Healthcare System. EGEMS (Wash DC) 2016;4(1):1240 [FREE Full text] [doi: 10.13063/2327-9214.1240] [Medline: 27563683]
10. Wahl E, Gross A, Chernitskiy V, Trupin L, Gensler L, Chaganti K, et al. Validity and Responsiveness of a 10-Item Patient-Reported Measure of Physical Function in a Rheumatoid Arthritis Clinic Population. Arthritis Care Res (Hoboken) 2017 Mar;69(3):338-346. [doi: 10.1002/acr.22956] [Medline: 27332620]
11. Oude Voshaar MAH, ten Klooster PM, Glas CAW, Vonkeman HE, Krishnan E, van de Laar MAFJ. Relative performance of commonly used physical function questionnaires in rheumatoid arthritis and a patient-reported outcomes measurement information system computerized adaptive test. Arthritis Rheumatol 2014 Oct;66(10):2900-2908 [FREE Full text] [doi: 10.1002/art.38759] [Medline: 24964773]
12. Duclos A, Voirin N. The p-control chart: a tool for care improvement. Int J Qual Health Care 2010 Oct;22(5):402-407. [doi: 10.1093/intqhc/mzq037] [Medline: 20675711]
13. 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]
14. van der Vaart R, Drossaert CHC, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MAFJ. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
15. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
16. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient race/ethnicity and shared medical record use among diabetes patients. Med Care 2012 May;50(5):434-440. [doi: 10.1097/MLR.0b013e318249d81b] [Medline: 22354209]
17. Garrido T, Kanter M, Meng D, Turley M, Wang J, Sue V, et al. Race/ethnicity, personal health record access, and quality of care. Am J Manag Care 2015 Feb;21(2):e103-e113 [FREE Full text] [Medline: 25880485]
18. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]
19. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. "I Want to Keep the Personal Relationship With My Doctor": Understanding Barriers to Portal Use among African Americans and Latinos. J Med Internet Res 2016 Oct 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
20. Bates DW, Bitton A. The future of health information technology in the patient-centered medical home. Health Aff (Millwood) 2010 Apr;29(4):614-621 [FREE Full text] [doi: 10.1377/hlthaff.2010.0007] [Medline: 20368590]
21. Luque AE, van Kenken A, Winters P, Keefer MC, Sanders M, Fiscella K. Barriers and Facilitators of Online Patient Portals to Personal Health Records Among Persons Living With HIV: Formative Research. JMIR Res Protoc 2013 Jan 22;2(1):e8 [FREE Full text] [doi: 10.2196/resprot.2302] [Medline: 23612564]
22. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Apr 25. [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
23. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]

Abbreviations

ACR: American College of Rheumatology
ANOVA: analysis of variance

EHR: electronic health record
PF: physical function
PROMIS: Patient Reported Outcome Measurement Information System
PRO: patient-reported outcome
RA: rheumatoid arthritis

Edited by G Eysenbach; submitted 15.08.17; peer-reviewed by S Matsuda, D Lavallee, S Wissig, J Edwards; comments to author 30.11.17; revised version received 25.01.18; accepted 19.02.18; published 09.05.18.

Please cite as:
Li J, Yazdany J, Trupin L, Izadi Z, Gianfrancesco M, Goglin S, Schmajuk G
Capturing a Patient-Reported Measure of Physical Function Through an Online Electronic Health Record Patient Portal in an Ambulatory Clinic: Implementation Study
JMIR Med Inform 2018;6(2):e31
URL: http://medinform.jmir.org/2018/2/e31/
doi:10.2196/medinform.8687
PMID:29743158

Original Paper

An Early Model for Value and Sustainability in Health Information Exchanges: Qualitative Study

Sue S Feldman[1], RN, MEd, PhD

Department of Health Services Administration, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Sue S Feldman, RN, MEd, PhD
Department of Health Services Administration
University of Alabama at Birmingham
SHP #590K
1716 9th Avenue South
Birmingham, AL,
United States
Phone: 1 205 975 0809
Email: suefeldman1009@gmail.com

Abstract

Background: The primary value relative to health information exchange has been seen in terms of cost savings relative to laboratory and radiology testing, emergency department expenditures, and admissions. However, models are needed to statistically quantify value and sustainability and better understand the dependent and mediating factors that contribute to value and sustainability.

Objective: The purpose of this study was to provide a basis for early model development for health information exchange value and sustainability.

Methods: A qualitative study was conducted with 21 interviews of eHealth Exchange participants across 10 organizations. Using a grounded theory approach and 3.0 as a relative frequency threshold, 5 main categories and 16 subcategories emerged.

Results: This study identifies 3 core current perceived value factors and 5 potential perceived value factors—how interviewees predict health information exchanges may evolve as there are more participants. These value factors were used as the foundation for early model development for sustainability of health information exchange.

Conclusions: Using the value factors from the interviews, the study provides the basis for early model development for health information exchange value and sustainability. This basis includes factors from the research: fostering consumer engagement; establishing a provider directory; quantifying use, cost, and clinical outcomes; ensuring data integrity through patient matching; and increasing awareness, usefulness, interoperability, and sustainability of eHealth Exchange.

(JMIR Med Inform 2018;6(2):e29) doi:10.2196/medinform.9299

KEYWORDS

health information exchange; medical informatics; information systems; value proposition; health informatics

Introduction

Background

The last decade has been one of understanding the contribution of the health information exchange to health care's Triple Aim: improved care, lowered costs, and increased patient satisfaction. To that end, eHealth Exchange (formerly Nationwide Health Information Network [NwHIN]) was established in 2009 as the nation's mechanism of health information exchange. However, onboarding was slow, and the US government soon realized that internal electronic exchange within an organization was not enough. Motivated by incentive funding provided by the Health Information Technology for Economic and Clinical Health (HITECH) Act, many states or regions have health information exchanges (HIEs), and many electronic health record (EHR) vendors are capable of health information exchange with disparate organizations. For the purposes of this paper, HIE refers to a single organization or group of organizations facilitating the act of electronic health information exchange. Additionally, eHealth Exchange is a vehicle facilitating health information exchange for HIEs.

While no solution, including eHealth Exchange, will singlehandedly address every health information exchange scenario, eHealth Exchange, as our nation's HIE, is an environment and a component toward the ability to exchange records with any provider, at any time, for any patient. Isolated use cases and studies have tried to quantify the economic value of health data exchange across an HIE in general [1,2] and eHealth Exchange more specifically [3,4], and some have reported cost savings in terms of laboratory and radiology testing, emergency department expenditures, and admissions [2,5-7], one of which claims, "little generalizable evidence currently exists regarding benefits attributable to HIE" [7]. Additionally, models that consider both current and perceived value are needed to help move away from isolated use case examples and statistically quantify value and sustainability. As shown in the literature, value is not a singular focus, and therefore a method and model of statistically quantifying value that considers multiple factors is important.

The Sequoia Project

The Sequoia Project, which partially funded this study, is a nonprofit membership corporation whose goal is to improve the health and welfare of all Americans by supporting and advancing health data exchange that is trusted, scalable, and enhances quality of care and health outcomes by supporting comprehensive longitudinal health records. The Sequoia Project seeks to expand trusted, secure, and interoperable exchange of health information across the nation by fostering cross-industry collaboration and consensus agreement among public and private organizations who wish to function as interconnected networks. Current eHealth Exchange participation includes over 100 organizations, representing about 33% of all US hospitals, over 17,000 medical groups, over 8200 pharmacies, over 1000 dialysis centers, and over 100 million patients.

Theoretical Orientation: Group Forming Networks

While over 100 organizations participate in eHealth Exchange, there are few regional clusters/networks within which medical information is able to be queried and retrieved. Having regional clusters/networks would facilitate the transportation of vital information needed to provide a comprehensive clinical picture, exponentially (according to the premise of group forming networks) increasing the value of eHealth Exchange to *all* organizations. The Healthcare Information and Management Systems Society (HIMSS) [8] suggests that more needs to be done to show the business value of health data exchange and suggests value in terms of creating a health care data economy whereby people are willing to pay for and sell data, stakeholders could control data and exchange with others, and the surrounding ecosystem includes measures of interoperability that are meaningful to patients and providers.

This value equation has been seen in other network of networks configurations, described as group forming networks, or Reed's Law. Reed describes 3 types of networks: a one-to-many network, in which a central entity shares information with a large number of members (eg, through a Web portal); a one-to-one network, where single members are connected to other individuals to conduct a number of transactions (eg, email); and a flexible communication network, which renders it possible to connect not only pairs of participants but groups as well. Metcalfe's Law has also been used to describe network value but does not account for the power of group connections, in this case HIE networks or groups [9].

Under Reed's Law, value grows such that the whole network (eHealth Exchange) is greater than the sum of the individual participants or clusters/networks (statewide or vendor HIE networks) [10]. This environment exponentially increases the number of health data exchange transactions that can occur and broadens the geographical reach of the individual and collective networks, thereby providing more accurate, current, and comprehensive information at the point of care. Furthermore, the expansion of accountable care models and retail medical clinics (eg, CVS MinuteClinic or Walgreens Healthcare Clinic) present additional opportunities for onboarding to state or regional HIEs, thereby bringing additional groups to eHealth Exchange (assuming the state or regional HIEs are themselves onboarded to eHealth Exchange). The use of such a network of networks could aid widespread achievement of the Triple Aim, widespread use of health information exchange in general and eHealth Exchange in particular, and increase the value of individual and collective factors. The purpose of this study was to explore the various factors associated with real and perceived value to provide a basis for early model development for health information exchange value and sustainability.

At a high level, Reed's Law suggests that all connections result in some degree of value. A white paper from Brookings Center for Technology Innovation [11] provides more detail on this with a model showing connections between patients, payers, medical data providers, and health care providers. For example, more data between patients and health care providers could result in better care, and better care could result in lower costs between patients and payers.

Methods

The study design incorporated 21 semistructured 1-hour phone interviews and document analyses to understand the perceived current and potential value factors of eHealth Exchange participation. Each interview was recorded and transcribed. Transcriptions were imported into ATLAS.ti (ATLAS.ti Scientific Software Development GmbH) for data organization and analysis. The findings from the interviews were used to form the basis of an early model for health information exchange value and sustainability.

Interviewees were recruited by email invitation and were purposefully selected based on their participation in the decision-making process to onboard to eHealth Exchange. All interviewees were consented and their identity, location, organization, and role within the organization will be kept confidential; interviewees came from the hospital system (7), statewide HIE (2), regional HIE (1), vendor (2), and federal government (9) sectors.

The following is an example of selected interview questions (a full listing can be found in Multimedia Appendix 1):

- What technical advances will need to happen for more organizations to join eHealth Exchange?

- What technical issues need to be solved to impact sustainability?
- What are the current reasons for maintaining your participation in eHealth Exchange?
- What needs to happen for HIE to impact improved care delivery, reduce costs, etc?
- What public policies need to happen for HIE to be a standard of care?

Using a grounded theory inductive approach [12], 16 conceptual categories and 73 subcategories emerged with relative frequency (RF) counts ranging from 0.20 to 6.77. RF, in this case, is the proportion of responses (as in frequency of a response) in the particular category across all interviews divided by the number of interviewees. For example, if 10 interviewees talk about the number of records being exchanged using eHealth Exchange as having to do with use, the RF would be 3.00. Using 3.00 as a cutoff, 5 main categories and 16 subcategories are described in the findings.

Results

Overview

The findings of this qualitative study reveal that a majority of eHealth Exchange participants have onboarded since 2014, even though eHealth Exchange originated in 2009 (as NwHIN). Overall, interviewees demonstrated much confusion regarding vendor HIEs, regional HIEs, statewide HIEs, and eHealth Exchange. At times in the interview process, interviewees needed to be recentered that the interview was specific to eHealth Exchange and not about other HIEs such as those contained within vendor systems. When interviewees were asked about the alignment of policy to health data exchange initiatives, many commented that public policy and legislation need to catch up to the willingness of providers to exchange information and of consumers to have their information exchanged.

Using RF≥3.00 as a top-tier cutoff for data reporting, Table 1 and Table 2 show 5 conceptual categories and 16 subcategories. To readily show the issues of greatest importance, Table 1 is organized in descending order (RF=6.77 to RF=3.00). To correspond to the narrative detail in this section, Table 2 is organized with the subcategory data grouped by category.

eHealth Exchange Concerns and Challenges

The primary concerns expressed by interviewees related to interoperability (RF=6.17), level of implementation (RF=4.60), and increasing statewide or regional HIE to eHealth Exchange connectivity (RF=3.17).

Interoperability

When interviewees discussed reconciling technology and usability, they pointed out that eHealth Exchange is not plug and play and lamented the lack of direct communication from vendors about their system requirements. One interviewee summed up what many expressed: “Make it [eHealth Exchange] as interoperable as banking.” Even still, interoperability will require constant consensus building, improvement, and course corrections to keep pace with innovations.

Table 1. Overall findings of the interviews in descending order (relative frequency [RF] ≥3.00).

Category	Subcategory	RF
Value	Value in better care	6.77
Use	Increase eHealth Exchange use	6.75
eHealth Exchange concerns or challenges	Interoperability	6.17
Technical	Technical standards	6.13
Technical	Patient matching	5.80
Value	Value in avoiding duplication	4.93
eHealth Exchange concerns	Level of implementation	4.60
Value	Value in lowering costs	4.47
Technical	Data usability	4.37
Use	Who is using eHealth Exchange	4.30
Technical	Data integrity	3.87
Value	Intangible value	3.53
Use	Actual eHealth Exchange use time	3.30
eHealth Exchange concerns	Increase statewide health information exchange to eHealth Exchange connectivity	3.17
Governance	Data Use and Reciprocal Support Agreement	3.10
Use	Number of records exchanged using eHealth Exchange	3.00

Table 2. Findings of the interviews by descending order by category (relative frequency [RF] ≥3.00).

Category	RF
eHealth Exchange concerns or challenges	
Interoperability	6.17
Level of implementation	4.60
Increase statewide and regional health information exchange to eHealth Exchange connectivity	3.17
Governance	
Data Use and Reciprocal Support Agreement	3.10
Technical	
Technical standards	6.13
Patient matching	5.80
Data usability	4.37
Data integrity	3.87
Use	
Increase eHealth Exchange use	6.75
eHealth Exchange participants	4.30
Actual eHealth Exchange use time	3.30
Number of records exchanged using eHealth Exchange	3.00
Value	
Value in better care	6.77
Value in avoiding duplication	4.93
Value in lowering costs	4.47
Social Security Administration Disability Determination	3.53

Level of Implementation

Interview analysis suggested that few organizations are fully implemented, which would mean that they are performing queries, receiving and consuming the EHR usable information, and connected to federal partners. In terms of meaningful use, interviewees referred primarily to using eHealth Exchange as a vehicle for care transition summaries. Others described exchanging with federal partners as the level of implementation. When questioned further about implementation, many interviewees discussed other HIE networks used to exchange clinical data (eg, regional HIEs, vendor HIEs, specialized practice HIEs). Whether or not these HIEs were eHealth Exchange participants, a majority of the interviewees have implemented eHealth Exchange at the federal partner level for Social Security Administration (SSA) disability determination and/or the Veterans Health Administration (VHA).

Increasing Statewide and Regional Health Information Exchange to eHealth Exchange Connectivity

Most interviewees thought that increasing statewide or regional HIE to eHealth Exchange connectivity would depend on a less cumbersome process to gain more traction. Regardless of processes that need streamlining (eg, testing and sign-in), interviewees suggested that statewide or regional HIEs should be the first level of connection to eHealth Exchange, then organizations and health systems should connect to their statewide or regional HIE. One interviewee stated, “I would say 90% to 100% of the time it [data from the statewide HIE] impacts the way that I deal with every single patient. There’s something on there that either changes the care that I would deliver... and because I’m aware of [a] clinical context, I’m just going to deal with that patient a little bit differently.” Such comments support the value of building a network of networks.

Governance

Many interviewees thought that the Data Use and Reciprocal Support Agreement (DURSA), put in place by the Office of the National Coordinator for Health Information Technology (ONC) and carried forward by The Sequoia Project, was comprehensive and saved them significant legal counsel expenses (RF=3.10) to ensure that best practices, legislative regulations, and common sense are employed. Regional or statewide HIEs reported spending very few resources on DURSA review, which may be partially due to previous familiarity with the agreement. Overall, many interviewees said that there is a certain level of understanding and confidence that “we are all playing by the same rules.”

Technical Standards

The technical issues most often expressed by interviewees were technical standards (RF=6.13), patient matching (RF=5.80), data usability (RF=4.37), and data integrity (RF=3.87). Importantly, no one mentioned technology as a barrier but rather

raised selective technical areas that can be viewed as a natural consequence of the growth process.

Several interviewees commented about forward and backward compatibility between the 2010, 2011, and 2014 specifications. To provide some context around these comments, some organizations, such as SSA, support multiple versions, but this is not widespread. Organizations do not upgrade to the latest technical specifications in lockstep, so there will always be the need for forward and backward compatibility. For content standards, interviewees discussed the need for more granularity and more consistent interpretation of the standards. Two interviewees commented on the diversity of options for documenting data from the continuity of care document, although they acknowledged tighter specifications have resulted in improvements. Finally, some interviewees felt that vendors contribute to the lack of clarity with regard to standards—technical and content—and The Sequoia Project could help by setting universal standards and ensuring consistency in the interpretation and application of the standards between organizations and vendors.

Patient Matching

Interviewees mentioned that accurate patient matching is a critical component to seamless health data exchange across eHealth Exchange. One interviewee stated the alternative very simply: "The fallout from inaccurate patient matching is too risky." Interviewees linked patient matching to interoperability and data integrity saying that they may have made strides with patient matching for internal exchange within their organization, but more needs to be done specifically related to patient matching for external exchange across organizations (ie, eHealth Exchange). One interviewee suggested adoption of a nationwide patient matching strategy with standardized and vendor-agnostic patient demographic elements. Of those suggesting solutions, many mentioned a central patient list with a record locator and a unique health identifier (not the social security number).

Data Usability

Two interviewees suggested that trust can be critical to how usable data are used. For example, if a clinician suspects that data may not be accurate for the patient (perhaps due to inaccurate patient matching), the data will be discounted and not perceived as useful. One of these interviewees when on to say that while the data may be accurate, there may be no need for those particular data. Importantly, interviewees with more HIE experience (regional or statewide) expressed that they feel the data they get *are* usable and helpful.

Data Integrity

As one interviewee noted, "A fundamental and critical success factor for HIE is the ability to accurately link multiple records for the same patient across the disparate systems of the participating organizations." Another interviewee added that this becomes an issue of patient safety when data are incorrectly merged, sometimes between the wrong patients, and the absence of accurate patient matching was seen by many interviewees as the root problem behind data integrity.

Increasing eHealth Exchange Use

A majority of interviewee comments about use had to do with increasing eHealth Exchange use and usability (RF=6.75), understanding who is using eHealth Exchange (RF=4.30), the actual use time (RF=3.30), and the number of records exchanged using eHealth Exchange (RF=3.00).

Many interviewees felt that their organization's prior experience with data exchange had resulted in increased use of eHealth Exchange; however, that use was primarily for SSA disability determination. Many felt that SSA use was high for 2 reasons: it did not require initiation from the user and there was concrete revenue tied to its use. Other than SSA disability determination, some interviewees noted that their organization did not have any set primary purpose for eHealth Exchange and thought that might be a contributing factor to low use.

One interviewee said that in their organization, it is possible that users are not even sure if they are using Epic or eHealth Exchange to query for records, as the query goes first to Epic and then to the eHealth Exchange without signaling the transition. This scenario, for this organization, runs about 8:1 Epic to Epic versus eHealth Exchange; another organization cited a 10:1 Epic to Epic versus eHealth Exchange ratio. In other words, Epic records are returned 8 or 10 times more frequently than eHealth Exchange records.

Some interviewees suggested that if insurance companies became eHealth Exchange participants, use would increase. Some organizations created their own connectivity with HIEs that existed prior to eHealth Exchange (or even NwHIN) and have not transitioned over. Other interviewees mentioned usability: "Asking for the data is one thing; getting usable information is quite another."

eHealth Exchange Participants

Interview data suggest that, outside of SSA and VHA, users are organization-to-organization rather than organization-to-HIE (statewide or regional). Many interviewees suggested that the lack of a provider directory contributes to low use. One interviewee stated, "Just knowing who your eHealth Exchange neighbors are might increase the propensity to initiate an eHealth Exchange query."

Actual eHealth Exchange Use Time

The interviewees were from organizations that had been eHealth Exchange (or NwHIN) participants ranging from 1 to 7 years. However, findings suggest that the length of eHealth Exchange (or NwHIN) participation did not necessarily reflect the length of time that organizations were electronically exchanging health data. Those who have been using eHealth Exchange the longest (some starting as NwHIN participants) commented that a majority of their exchanges are with SSA.

Number of Records Exchanged Using eHealth Exchange

Many interviewees commented that the number of records exchanged using eHealth Exchange would rapidly increase if legislation made querying records a standard of care. However, these interviewees were quick to point out that doing so should not limit queries to only eHealth Exchange but from any HIE,

including vendor systems such as Epic. Another issue brought up by several interviewees was enforcing data contribution: if an organization is an eHealth Exchange participant, they need to contribute data. Depending on the organization and regulations for sensitive data, this may be more complicated than it sounds.

In terms of actual records exchanged across eHealth Exchange, Table 3 lists in ascending order the average records transacted each month as reported by the interviewee.

Value

While responses varied, it was apparent that all interviewees perceived value in being an eHealth Exchange participant. However, when queried for concrete value statements, interviewees mostly pointed to revenue generated from SSA participation. Most interviewees expressed that the primary perceived value was located in better care (RF=6.77), but others cited avoiding duplication of services (eg, lab, radiology) (RF=4.93). Again, although mostly anecdotal evidence, many mentioned lower costs of care as one of the value factors (RF=4.47). SSA disability determination (RF=3.53) was the only factor mentioned with a value that interviewees felt they could quantify for their organizations. It is critical to note that many interviewees who have been conducting health information exchange through regional HIEs anecdotally report better care, duplication avoidance, etc. These interviewees draw from these known experiences and perceive that this same value can and will happen at a national level with eHealth Exchange.

Value in Better Care

Several interviewees commented that although they think use of eHealth Exchange will result in better care, "its use must become the standard of care." A network diagram constructed from interview data shows linkages to developing eHealth Exchange use as a standard of care. As shown in Figure 1, interviewees identified 8 core contributors to making eHealth Exchange a standard of care:

- Increased use
- Increased marketing
- Solidified sustainability
- Ability to get accurate, current, and needed data
- Provider directory
- Increased statewide HIE connectivity
- Organizational leadership commitments
- Consistent and clear standards

Value in Avoiding Duplication of Services

Several interviewees commented that in order to avoid duplication of services, eHealth Exchange must get the patient matching right. One interviewee suggested that avoiding duplication of services could actually be motivated through a bottom-up approach with "the patient say[ing] that they just had that test, can you please check eHealth Exchange?"

Value in Lowering Costs

Interviewees discussed the perceived value of eHealth Exchange in lowering overall costs of health care, and 1 interviewee commented, "While eHealth Exchange can play a strong role in lowering health care costs, we may not be able to attach causality to eHealth Exchange for a while."

Social Security Administration Disability Determination

Many interviewees commented that even though they are not seeing actual quantifiable value in terms of clinical outcomes, "Being an eHealth Exchange participant is the right thing to do for medicine." One interviewee commented: "Revenue is not directly tied to why we're part of the eHealth Exchange. We view participation with the eHealth Exchange as it's just a part of who we are, and what we want to do, and how we promote interoperability in the country. I have to say I have never been in a meeting where we'd say, 'Look, we're making this amount of money from the SSA.'"

SSA disability determination was the only quantifiable value factor mentioned by interviewees, and for many, the primary motivation for their organization's eHealth Exchange participation. One interviewee summed up the comments of many: "Credit to SSA for figuring out that [eHealth Exchange] was possible and then figuring out how to do it so it is of value." To provide some context, uncompensated care cost recovery is directly linked to SSA disability determination; if an SSA beneficiary is approved for Social Security Disability Insurance (SSDI), they are more likely to pay the hospital bill and seek medical care before using expensive emergency care [4]. Most interviewees estimated the cost for eHealth Exchange onboarding $100,000 to $400,000 and very dependent of the existence of previous HIE participation.

Table 3. Monthly eHealth Exchange transactions.

Type of organization	Region	Average records per month[a]
Health care system	Southwest	10
Hospital	Southwest	667
State health information exchange	Midwest	1400
Veterans Health Administration	Federal	2000
Regional health information exchange	South	4000
Social Security Administration	Federal	25,657[b]

[a]These are estimates given by interviewees and represent both inbound and outbound transactions.

[b]Author's analysis from Social Security Administration–provided data.

However, once onboarded, SSA participants estimated that the revenue generated from SSA queries largely offsets eHealth Exchange participation costs.

Current and Perceived Value Factors

The interview data revealed linkages to current and future perceived value as shown in Figures 2 and 3, respectively. As shown in Figure 2, interviewees identified 3 core current perceived value factors:

- SSA disability determination (revenue and uncompensated care cost recovery)—this is the only value factor to have been quantified
- Ease with which records are retrieved
- Reduction of administrative burden for staff needing to request records

In terms of the potential perceived value—what interviewees expect will happen as eHealth Exchange evolves and has more participants—Figure 3 shows 5 core items:

- Statewide HIE connectivity
- Avoiding test duplication
- Better care
- Ability to get accurate, current, and needed data
- Decreased costs

Health Information Exchange Model Development

The aforementioned factors that contribute to the current and potential perceived value provide the basis for model development for health information exchange value and sustainability (Figure 4). The next step is statistical testing of this model to understand the contribution of each factor in terms of dependent and mediating factors relative to value and sustainability.

Figure 1. eHealth Exchange as a standard of care. HIE: health information exchange.

Figure 2. Current perceived value proposition. SSA: Social Security Administration.

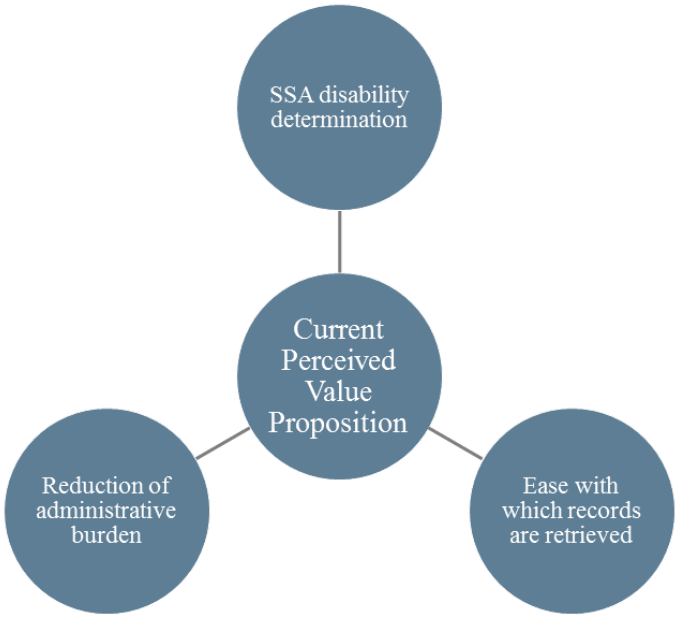


Figure 3. Potential perceived value proposition. HIE: health information exchange.

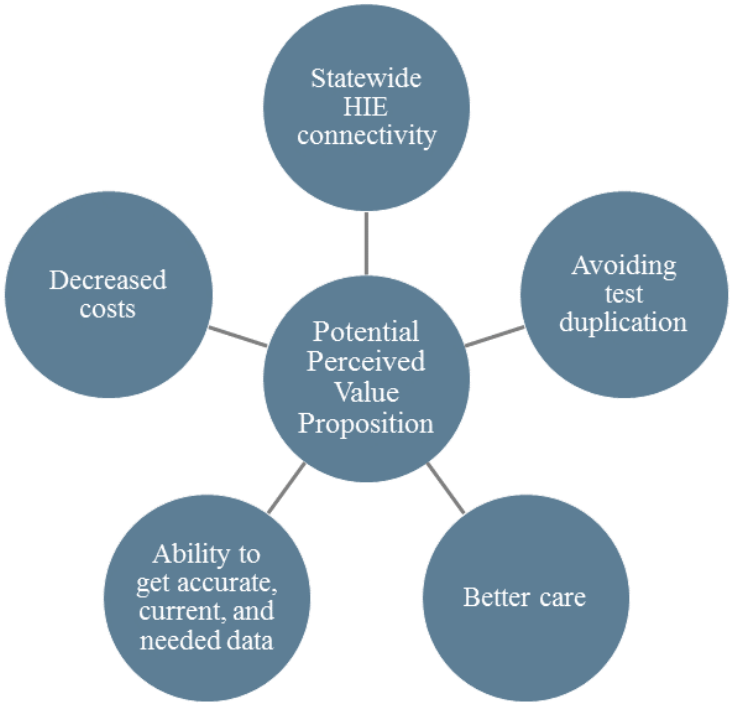
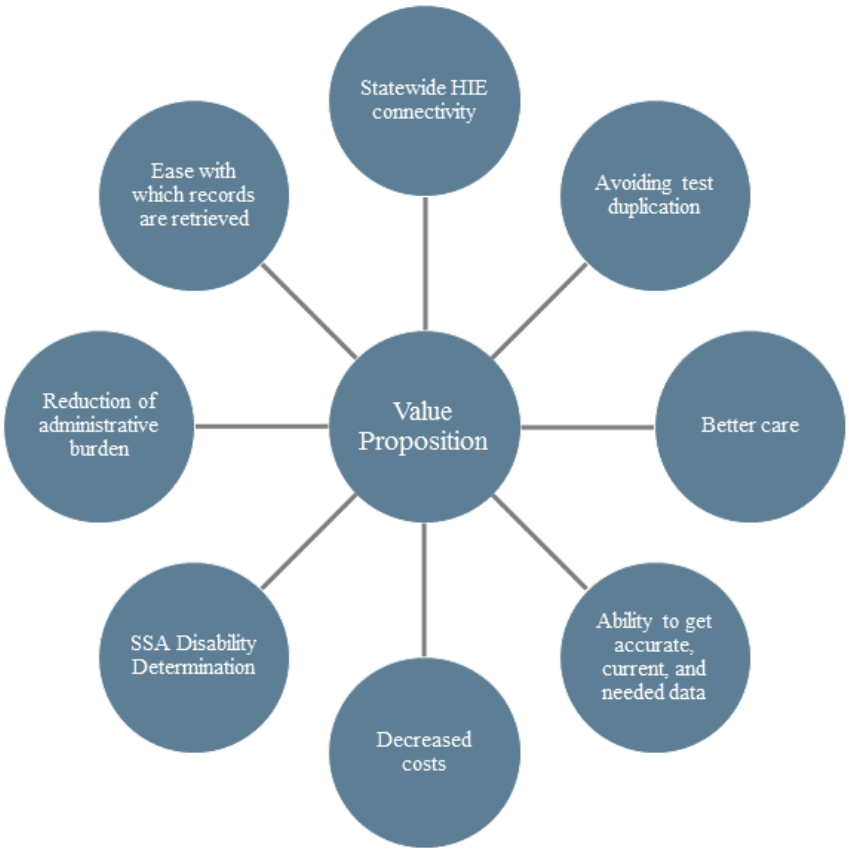


Figure 4. Model development for health information exchange value and sustainability. HIE: health information exchange; SSA: Social Security Administration.



Discussion

Principal Findings

Despite attention from policymakers and industry professionals, interoperability—the seamless exchange of health information among organizations for total patient care—remains elusive. Additionally, sharing health information across state lines is complicated by differing consent models (opt-in vs opt-out). As such, eHealth Exchange participants are not waiting for a perfect network but rather are willing to participate in what they can achieve now and readily avail themselves to the advantages of group forming networks (Reed's Law): group-to-group connections where each single connection creates a much larger network for health information exchange. This differs from those who are withholding eHealth Exchange onboarding until a tipping point of value has been achieved. It is likely that those who are current participants will see much earlier and much greater return on their investment and, more importantly, will be able to quantify elements of the Triple Aim. In ways that may not yet be apparent, such positioning could offer a strategic advantage to providing health care to patients from anywhere in the United States. By contrast, the primary findings from this study that factor into value could also impact or influence future value, especially if there is no further maturity of eHealth Exchange. It is also important to understand that even with complete interoperability, there could still be a lack of complete medical information, leading to a lack of trust in any of the information.

Group Forming Networks

eHealth Exchange has demonstrated usefulness in facilitating the development of group forming networks, as this enables health care providers to connect not only to each other but also to federal entities that have a vested interest in improving care quality (eg, VHA).

Value and Sustainability

There are several opportunities to enhance the value and thus the sustainability of eHealth Exchange. The first is to improve consumer engagement by educating patients on the value of health data exchange through an HIE. Doing so will create a culture of patients who expect and demand health data sharing as a standard of care. Additionally, compiling a national provider directory or a similar mechanism for eHealth Exchange participants will enable care providers to readily identify with whom they can exchange information.

In a similar vein, interviewees expressed that increasing awareness and usefulness of eHealth Exchange would prove beneficial to increasing value and sustainability. While there is much anecdotal discussion around what participants feel is working, very little of it has been formalized with studies. Additionally, interviewees commented on the need for increased marketing endeavors. Exchangeability for current eHealth Exchange participants can be increased by focusing on onboarding statewide HIEs and organizations in states neighboring current participants.

Another method to increase the value of eHealth Exchange is to quantify use cost and clinical outcomes through studies on well-established use cases for eHealth Exchange. Other benefits worth considering include decreased duplication for laboratory or radiology services and reduced admission rates from emergency department visits.

Ensuring data integrity and patient matching are priorities, with standardized processes to ensure overall data integrity and thus confidence in the information presented at the point of care. It is recommended that The Sequoia Project combine the findings from this study with public comments received from the recently released report entitled "A Framework for Cross-Organizational Patient Identity Management" [13].

The need to advance interoperability was mentioned by nearly every interviewee. It is recommended to use policy and funding levers to create a business imperative and clinical demand for interoperability. This may require greater involvement of the federal government to align economic incentives, including but not limited to a stronger commitment from the Centers for Medicare and Medicaid Services, which could take a multitude of forms but should start with something manageable, actionable, and measurable such as requiring all emergency department visits with an ambulatory sensitive condition diagnosis to have an external HIE query.

Model Development

While the above enhancement opportunities provide guidance to eHealth Exchange, parallel discovery is needed in understanding the strength of the constructs that contribute to the model suggested in Figure 4. This study combines current and potential perceived value to provide the basis for early model development for health information exchange value and sustainability. This model then needs to be statistically tested to determine the strength of each of the constructs and to what degree they are mediating or contributing factors.

Limitations

Limitations of this study include purposeful study participant recruitment of current eHealth Exchange participants who were involved in the decision-making process to onboard to eHealth Exchange. Future research would benefit from including end users. Additionally, several characteristics of eHealth Exchange are not applicable to health information exchange as it occurs broadly through other types of HIEs, such as those that are vendor supported. As such, the findings may or may not be applicable to health information exchange broadly.

Conclusion

Organizations interested in sharing electronic health information are not waiting for perfection in the HIE infrastructure (eHealth Exchange, state or local HIE) to engage in that sharing, but rather they are identifying particular use cases to demonstrate value. They are also relying on the advantages of group forming networks to increase the value of various use cases. Engagement of the consumer is emerging as a critical component in the value equation for health data exchange. With consumers having increased awareness of health data exchange, they stand to drive the future for health data exchange becoming a standard of care. However, absent data integrity and interoperability, the value equation will continue to be built on individually identified use case factors. This study looked at value from a variety of factors that contribute to value and sustainability. Future research will test this model to better understand the strength of each factor.

Acknowledgments

This study was funded in part by The Sequoia Project. The author acknowledges and appreciates the careful, detailed, and thoughtful comments by the reviewers whose suggestions strengthened this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview questions.

[PDF File (Adobe PDF File), 25KB - medinform_v6i2e29_app1.pdf]

References

1. Vest JR. Health information exchange and healthcare utilization. J Med Syst 2009 Jun;33(3):223-231. [Medline: 19408456]
2. Frisse ME, Johnson KB, Nian H, Davison CL, Gadd CS, Unertl KM, et al. The financial impact of health information exchange on emergency department care. J Am Med Inform Assoc 2012;19(3):328-333 [FREE Full text] [doi: 10.1136/amiajnl-2011-000394] [Medline: 22058169]
3. Feldman SS, Schooley BL, Bhavsar GP. Health information exchange implementation: lessons learned and critical success factors from a case study. JMIR Med Inform 2014 Aug 15;2(2):e19 [FREE Full text] [doi: 10.2196/medinform.3455] [Medline: 25599991]
4. Feldman S, Horan T, Drew D. Understanding the value proposition of health information exchange: the case of uncompensated care cost recovery. Health Systems 2017 Dec 19;2(2):134-146. [doi: 10.1057/hs.2012.24]

5. Hersh WR, Totten AM, Eden KB, Devine B, Gorman P, Kassakian SZ, et al. Outcomes from health information exchange: systematic review and future research needs. JMIR Med Inform 2015 Dec 15;3(4):e39 [FREE Full text] [doi: 10.2196/medinform.5215] [Medline: 26678413]
6. Yaraghi N. The benefits of health information exchange platforms: measuring the returns on a half a billion dollar investment.: Center for Technology Innovation at Brookings; 2015. URL: https://pdfs.semanticscholar.org/f575/b4e3dfcc1ff3ab85e2d5fc7053ac1a780a16.pdf [accessed 2018-03-17] [WebCite Cache ID 6xzgeYAgt]
7. Rahurkar S, Vest JR, Menachemi N. Despite the spread of health information exchange, there is little evidence of its impact on cost, use, and quality of care. Health Aff (Millwood) 2015 Mar;34(3):477-483. [doi: 10.1377/hlthaff.2014.0729] [Medline: 25732499]
8. HIMSS. 2014. The business case for interoperability and health information exchange URL: http://www.himss.org/business-case-interoperability-and-health-information-exchange?ItemNumber=32781 [accessed 2017-10-25] [WebCite Cache ID 6uUGlpMqe]
9. Odlyzko A, Tilly B. A refutation of Metcalfe's Law and a better estimate for the value of networks and network interconnections. 2005. URL: http://www.dtc.umn.edu/~odlyzko/doc/metcalfe.pdf [accessed 2018-04-14] [WebCite Cache ID 6ygRhiNIj]
10. Reed D. That sneaky exponential—beyond Metcalfe's Law to the power of community building. Context 2003 Feb.
11. Yaraghi N. A sustainable business model for health information exchange platforms: the solution to interoperability in healthcare IT.: The Brookings Institution; 2015. URL: https://www.brookings.edu/wp-content/uploads/2016/06/HIE.pdf [accessed 2018-03-17] [WebCite Cache ID 6xzgm6q1m]
12. Creswell J. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Newbury Park: Sage Publications; 2008.
13. The Sequoia Project. 2015 Nov 10. A framework for cross-organizational patient identity management URL: https://sequoiaproject.org/wp-content/uploads/2015/11/The-Sequoia-Project-Framework-for-Patient-Identity-Management.pdf [accessed 2018-04-22] [WebCite Cache ID 6ysckG3ZA]

Abbreviations

DURSA: Data Use and Reciprocal Support Agreement
EHR: electronic health record
HIE: health information exchange (a single organization or group of organizations facilitating the act of electronic health information exchange)
HIMSS: Healthcare Information and Management Systems Society
HITECH: Health Information Technology for Economic and Clinical Health
NwHIN: Nationwide Health Information Network (now eHealth Exchange)
ONC: Office of the National Coordinator for Health Information Technology
RF: relative frequency
SSA: Social Security Administration
SSDI: Social Security Disability Insurance
VHA: Veterans Health Administration

Edited by G Eysenbach; submitted 29.10.17; peer-reviewed by R Burkhard, B Schooley, M Abdelhamid, I Adeleke, S Rahurkar; comments to author 24.11.17; revised version received 17.03.18; accepted 05.04.18; published 30.04.18.

Please cite as:
Feldman SS
An Early Model for Value and Sustainability in Health Information Exchanges: Qualitative Study
JMIR Med Inform 2018;6(2):e29
URL: http://medinform.jmir.org/2018/2/e29/
doi:10.2196/medinform.9299
PMID:29712623

Original Paper

Development and Validation of a Functional Behavioural Assessment Ontology to Support Behavioural Health Interventions

Gianluca Merlo[1*], MSc; Giuseppe Chiazzese[1*], MSc; Davide Taibi[1*], MSc; Antonella Chifari[1*], MSc

Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, Palermo, Italy

[*]all authors contributed equally

Corresponding Author:
Gianluca Merlo, MSc
Istituto per le Tecnologie Didattiche
Consiglio Nazionale delle Ricerche
Via Ugo La Malfa 153
Palermo, 90146
Italy
Phone: 39 0916809207
Email: gianluca.merlo@itd.cnr.it

Abstract

Background: In the cognitive-behavioral approach, Functional Behavioural Assessment is one of the most effective methods to identify the variables that determine a problem behavior. In this context, the use of modern technologies can encourage the collection and sharing of behavioral patterns, effective intervention strategies, and statistical evidence about antecedents and consequences of clusters of problem behaviors, encouraging the designing of function-based interventions.

Objective: The paper describes the development and validation process used to design a specific Functional Behavioural Assessment Ontology (FBA-Ontology). The FBA-Ontology is a semantic representation of the variables that intervene in a behavioral observation process, facilitating the systematic collection of behavioral data, the consequential planning of treatment strategies and, indirectly, the scientific advancement in this field of study.

Methods: The ontology has been developed deducing concepts and relationships of the ontology from a gold standard and then performing a machine-based validation and a human-based assessment to validate the Functional Behavioural Assessment Ontology. These validation and verification processes were aimed to verify how much the ontology is conceptually well founded and semantically and syntactically correct.

Results: The Pellet reasoner checked the logical consistency and the integrity of classes and properties defined in the ontology, not detecting any violation of constraints in the ontology definition. To assess whether the ontology definition is coherent with the knowledge domain, human evaluation of the ontology was performed asking 84 people to fill in a questionnaire composed by 13 questions assessing concepts, relations between concepts, and concepts' attributes. The response rate for the survey was 29/84 (34.52%). The domain experts confirmed that the concepts, the attributes, and the relationships between concepts defined in the FBA-Ontology are valid and well represent the Functional Behavioural Assessment process.

Conclusions: The new ontology developed could be a useful tool to design new evidence-based systems in the Behavioral Interventions practices, encouraging the link with other Linked Open Data datasets and repositories to provide users with new models of eHealth focused on the management of problem behaviors. Therefore, new research is needed to develop and implement innovative strategies to improve the poor reproducibility and translatability of basic research findings in the field of behavioral assessment.

(JMIR Med Inform 2018;6(2):e37) doi:10.2196/medinform.7799

KEYWORDS

ontology; behavioral interventions; functional behavioral assessment; eHealth care; evidence-based practice

Introduction

Background

Behavioral Interventions (BI) are assessed as effective and evidence-based strategies by several studies and meta-analyses for reducing problem behaviors identified in school-age children from [1-5]. Among BI the Functional Behavioral Assessment (FBA) is considered one of the most effective methods for identifying the antecedents and consequences that control a problem behavior [6] and for gathering information about the reason or function for a behavior [7]. In the FBA, the data obtained through indirect measures, direct observation, and experimental manipulation of environmental variables contribute in formulating a functional hypothesis.

It can then be used in designing effective intervention plans aimed at reducing the reinforcement effect that specific antecedents and consequents could have in triggering and maintaining the problem behavior. For instance, children with Attention Deficit Hyperactivity Disorder (ADHD), one of the most common syndromes generating behavior disorders, often show many disruptive behaviors during class at school. If appropriate instructional methodologies are not implemented by teachers, a child with ADHD can have difficulties in sustaining attention to a task, and this can trigger challenging classroom behavior.

For example, these include: calling out, leaving their seat, and frequent rule violations. If the FBA was applied in a similar case, health professionals would probably have hypothesized that the function “avoidance” is what motivates the child´s behavior in an attempt to get away from the frustrating task. Accordingly, they would have suggested teachers use an intervention plan composed of strategies aimed at increasing the student task-oriented behaviors. These include the following: breaking the task into smaller portions, reducing the task duration, using visual cues, and reducing the number of challenging ones.

Newcomer and Lewis [8], comparing treatment outcomes demonstrate that behavior intervention plans based on FBA information (function-based) were more effective than behavior intervention plans not based on FBA information (non-function-based). This confirms the usefulness and importance of conducting an FBA to guide intervention plans based on the conscientious and explicit use of current best evidence [9].

The general tendency of the scientific community to open and share processes and results to anyone interested could be a further opportunity to corroborate the application of FBA as Evidence-Based Practice (EBP). An EBP is a decision-making process that integrates: the best available evidence, clinical expertise, client values, and context [10]. As suggested by Kazdin [11], clinical psychology “would profit enormously from codifying the experiences of the clinician in practice so that the information is accumulated and can be drawn on to generate and test hypotheses”. Transparency, openness, and, reproducibility could be the lifeblood for the advancement of psychological sciences and the dissemination of a more open research culture [12]. However, Scott and Alter [13], reveal that only a few scientific papers about FBA with an EBP approach can be found in the literature.

In this direction, Richesson and Andrews [14] explore how computer science could support the digitization and computation of information related to clinical processes regarding representation of knowledge found in clinical studies and in particular the role of ontologies.

In computer science, an ontology is a taxonomic description of the concepts in an application domain and the relationships among them [15] aimed to promote knowledge generation, organization, reuse, integration, and analysis [16]. Ontologies are a powerful tool to accumulate knowledge in a specific domain especially when there is a lack of shared terms and procedures.

Today, the use of ontologies in biomedical research is an established practice. For example, Gene Ontology [17] provides researchers with extensive knowledge regarding the functions of genes and gene products. Also, the Open Biomedical Ontologies initiative provides a repository of controlled vocabularies to be used across different biological and medical domains [18]. However, computable information about behavioral disorders and mental illness is still dispersed. The lack of shared definition and practices makes them difficult to aggregate, share, and search for specific information when needed.

Recently, researchers have started to recognize the important role that ontologies can play in the clinical psychology context. By far the most interesting examples include the: (1) Mental Disease Ontology [19], (2) Mental Health Ontology [20], (3) Mood Disorder Ontology [21], (4) Autism Phenotype Ontology [22], (5) Ontology of Schizophrenia [23], and (6) Ontology to monitor mental retardation rehabilitation process [24].

In the domain of the description of human behavior, a successful example is the Ontology for human behavior models [25] created with the purpose of tracing what causes a person to take an action, the cognitive state associated with the behavior, and the effects of the particular action. However, at the time of writing this paper, authors have not identified ontologies specifically focused on behavioral disorders according to the FBA methods in the main international journals on medical information systems. Starting from this perspective, the definition of a Functional Behavior Ontology (FBA-Ontology) could play a key role in the adoption of an evidence-based approach among behavioral experts to fill the gap between research and practice still widely observed in clinical psychology [26]. In fact, data mining algorithms have great potential for identifying patterns in psychological data, facilitating the decision-making processes, and automatic meta-analysis.

This study presents the description and validation of the FBA-Ontology [27] as a semantic tool to support the systematic collection of behavioral knowledge and the decision-making process based on evidence and gathered data.

Methods

Methodological Approach

The FBA-Ontology was developed applying the Uschold and King [28] methodology, which comprises the following set of guidelines: (1) identification of the ontology purpose, (2) capture the concepts and the relations between the concepts, (3) coding the ontology using a formal language, and (4) evaluate the ontology from a technical point of view. Moreover, authors of the present contribution also added a human-based assessment by interviewing 84 domain experts to check the formal structure of the ontology regarding taxonomy, relationships, and axioms. The next paragraphs describe in detail each of the steps as mentioned above.

Identification of the Ontology Purpose

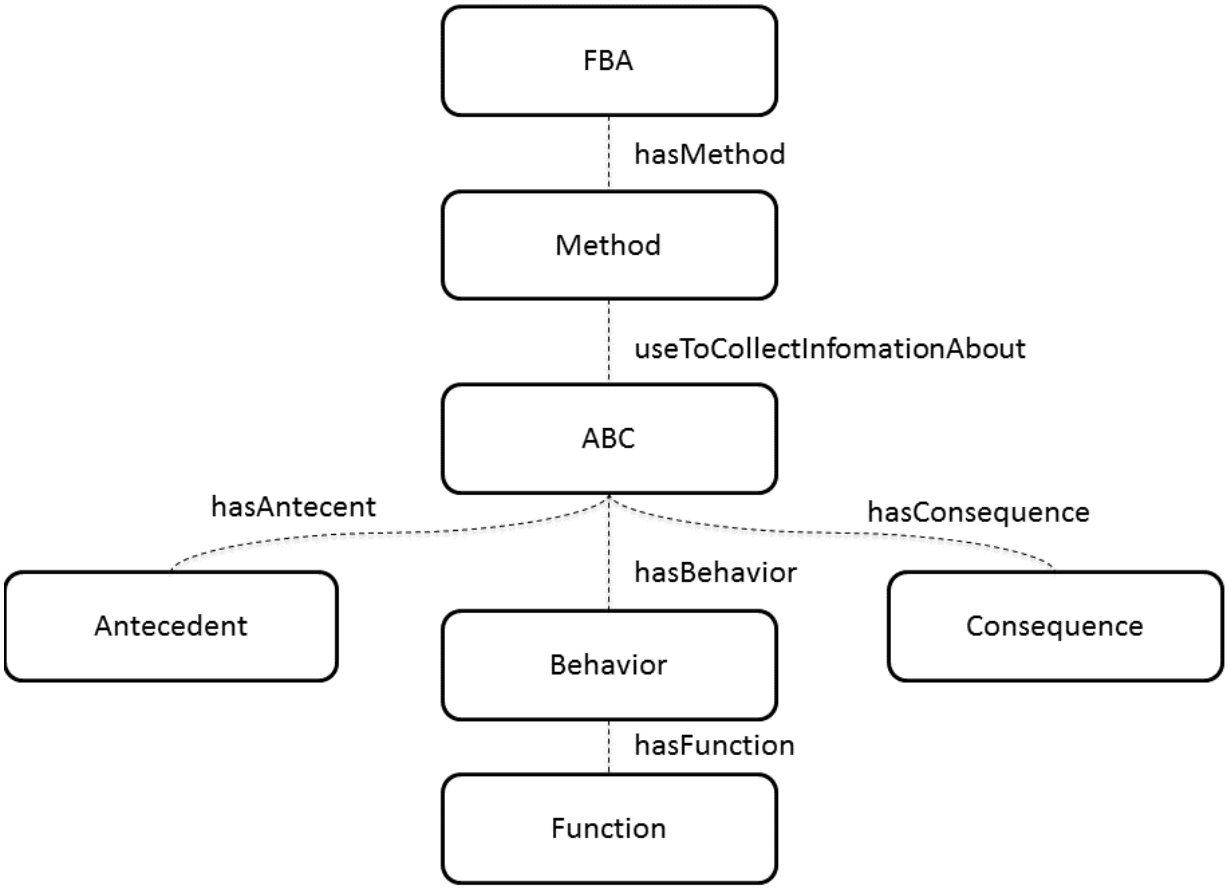
The FBA-Ontology purpose is to describe the structure and the semantics of Functional Behavior Assessment methods. Gresham et al [29] define the FBA as "a collection of methods for gathering information about antecedents, behaviors, and consequences to determine the reason (function) of behavior." The FBA derives from operant learning theories [30,31], and it is commonly used in clinical and educational contexts to design effective intervention plans. These are aimed at reducing the reinforcement effect that specific antecedents and consequents could have in triggering and maintaining the problem behavior.

Capture the Concepts and the Relations Between the Concepts

The concepts of the ontology were captured starting from the above mentioned theoretical assumptions. The FBA-Ontology is, therefore, a collection of classes and properties used to describe the whole assessment process. This includes the definition of a target behavior, the collection of the behavioral data, the hypotheses about the target behavior functions, and the planning of a behavioral intervention. In particular, the FBA-Ontology key concepts are FBA, Method, Antecedent, Behavior, Consequence, and Function (Figure 1).

According to Hanley [32], the FBA is a descriptive assessment including indirect and direct observation methods and measurements of a target behavior. The FBA-Ontology includes the *Method* class to specify the observation methods applied to the target behavior, specified in the class *Behavior*. Rating scales, questionnaires, and interviews are examples of indirect methods because they do not require direct observation of the target behavior. The direct methods are based on descriptive assessments and systematic recordings of observation sessions. The descriptive assessments provide qualitative information about variables that may trigger or maintain a target behavior, while the recording methods, such as the systematic direct observation, provide quantitative information about frequency, intensity, and duration of a targeted behavior during a specific time interval. The property *isDirect*, defined within the *Method* class, models the use of several direct or indirect methods [33].

Figure 1. Key concepts at the basis of the FBA-Ontology.



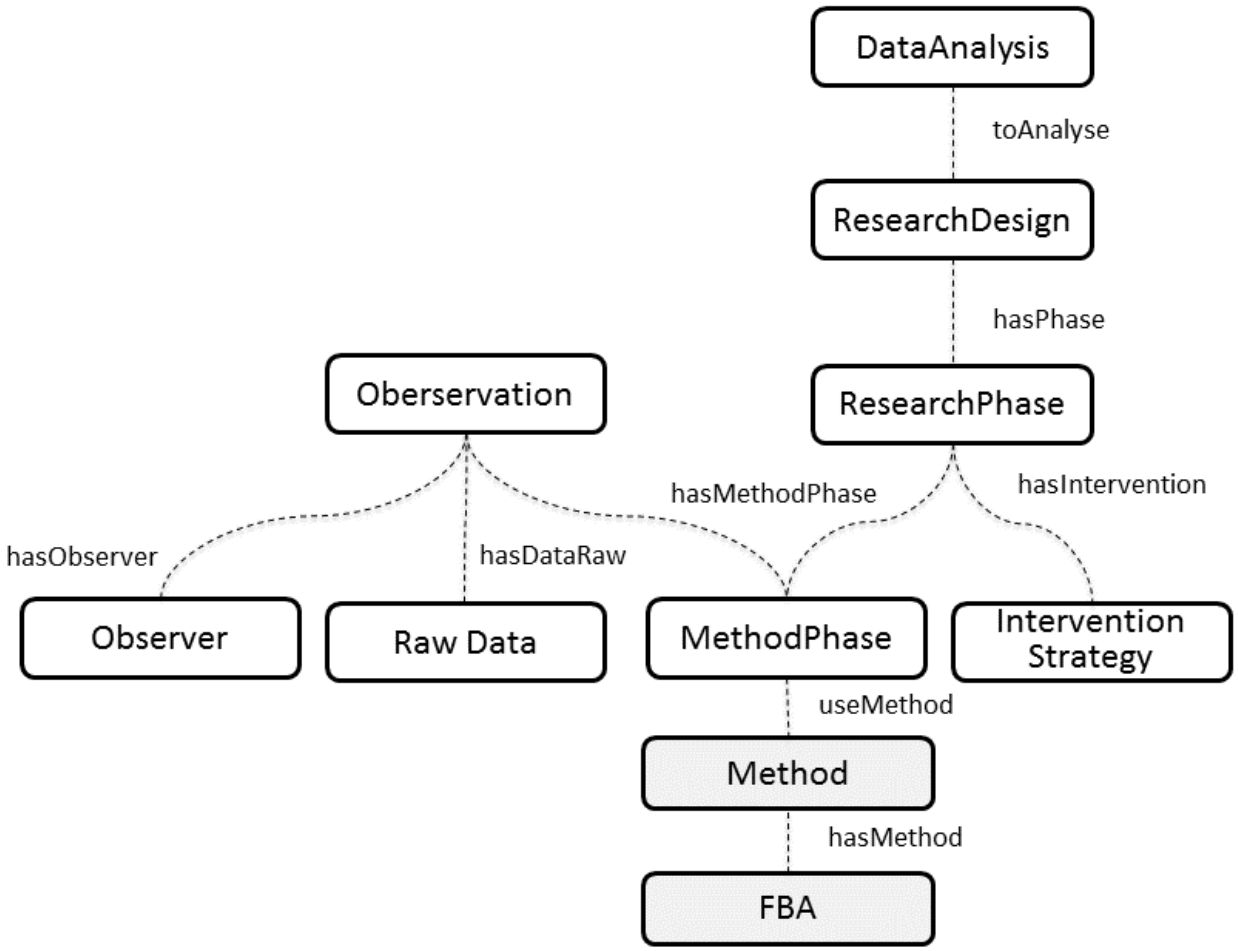
The triad of classes: *Antecedent*, *Behavior*, and *Consequence* encloses the descriptors of the target behavior (Behavior), and the variables that trigger (Antecedent) or maintain it (Consequence). The class *Function* defines what purpose the problem behavior serves for the individual. According to Iwata and colleagues [34], the four behavioral functions are: avoid or escape difficult tasks, gain adult and peer attention, access to a desired object or activity, and sensory stimulation. The FBA-Ontology embodies these functions through the purported enumerated *datatypes*, included into the *Function* class mentioned above.

Unlike experimental designs where researchers can randomly assign participants to a control and treatment group, the behavior of the subject under observation generates data that can change over time or stay steady. To evaluate whether the time series changes, the elective and most popular research approach is single case research design. Single-case research designs are a diverse and powerful set of procedures used for demonstrating causal relationships among clinical phenomena [35]. Clinicians use three main research designs: case studies, quasi-experimental designs, and experimental designs. The differences among these are regarded as the increasing level of scientific rigor, ranging from anecdotal data gathered retrospectively to the maximum level of control of the dependent variables achievable in a laboratory setting. Dallery and Raiff [36] suggest the use of single-case design as a method for optimizing behavioral health interventions and facilitating the practitioners in the planning of suitable interventions for both individuals and groups. The FBA-Ontology assumes that data about an observed behavior is collected according to the single-case research design constraints (Figure 2).

Generally, single-case designs start with a baseline phase (A) to observe the dependent variable as it appears. Once the baseline is established, the observer continues while implementing the intervention (B) to compare the time series looking for significant changes. The design just described is named AB. Other examples of single-case designs are ABA (adding a nontreatment condition to the AB design), ABAB (repeating the AB design twice) or BAB (implementing the intervention immediately for the safety of the person observed). The typology of single case designs used during the FBA process can be specified in the property *ResearchType* of the *ResearchDesign* class. In turn, the class *ResearchPhase* identifies the specific phase of the research design.

The *Observation* class describes the observed data gathered during a research phase. This class is linked both to the type *Observer* (the person who carries out the observation) and *RawData* (the data collected during the observation session). The class *InterventionStrategy* defines the set of strategies chosen to increase positive behaviors or decrease negative ones.

Figure 2. Key concepts related to research methods and data collection.



The *RawData*, once gathered during the single-case design, could be analyzed to assess the statistical effect of the intervention implemented. Many statistical methods include non-parametric tests and time-series analysis which are used to compare the data gathered during the different experimental conditions. The *DataAnalysis* class of the FBA-Ontology specifies what statistical methods are used to analyze data.

To identify the constraints related to the concepts included in the FBA-Ontology, a set of competency questions were formulated (Table 1). Competency questions are requirements that are expressed in the form of questions [37-39] using the natural language. They play an important role in both the ontology creation and validation. The competence questions support the ontology development enabling developers to identify the main elements and relationships within the selected domain. They also represent a starting point to carry out a deeper evaluation in a later stage of development [40].

Coding the Ontology Using a Formal Language

The Protégé tool was used to model the ontology and produce an OWL (Ontology Web Language) version of the ontology. The Protégé tool also includes "reasoners" that can be used to perform inferences and to verify the ontology.

The process of ontology verification is generally performed to check its syntactic quality and the presence of anomalies or pitfalls.

In a metric proposed by Burton-Jones and colleagues [41], the syntactic quality is measured by assessing whether the source code is correctly structured, and how rich the programming language features are which model the ontology.

The anomalies or pitfalls can refer to the assessment of the logic consistency of the ontology and the identification of modeling issues in comparison with well-known best practices [42]. Many automatic tools have been developed to facilitate the ontology verification. For instance, XD Analyzer checks whether the ontology satisfies a set of best practice criteria or not, showing errors, warnings, and suggestions useful to improve it. XD Analyzer is included in XD Tools [43], and it is released as a plugin for Eclipse. Another useful recent tool is OOPS! [42] a Web-based tool aimed at identifying the most common anomalies in the ontology development. It scans 21 pitfalls grouped in 4 different dimensions: human understanding, logical consistency, real word representation, and modeling issues. Many other tools are available, but their description is out of the scope of the present paper. The FBA-ontology created with Protégé was verified by using the Pellet reasoner [44] to check the logical consistency of the ontology in addition to the integrity of classes and properties defined. The Pellet reasoner has not detected any violation of constraints in the ontology definition in its results.

The Evaluation of the Ontology

The evaluation of the quality of an ontology plays a key role during the whole ontology development process. As suggested by Gomez-Perez [45], evaluating an ontology should ensure that it correctly implements the expected requirements and performs correctly in the real world. Low-quality ontologies reduce the possibility that intelligent agents can perform accurately intelligent tasks because of inaccurate, incomplete or inconsistent information [42]. The quality of an ontology can be assessed evaluating how well a semantic structure represents the knowledge about a specific domain and the relationships about the identified concepts. Sabou and Fernandez [46] use the term "ontology validation" to compare the ontology definitions with a frame of reference that the ontology would represent.

Table 1. List of competence questions formulated for the FBA-Ontology and its relative constraints.

Competency questions	Constraints
Which are the types of methods to collect information about a behavior?	Direct Indirect
How many methods can have an FBA[a]?	Unlimited number
How many functions serve a behavior?	At least 1
Which are the functions of a behavior?	Avoid or escape difficult tasks, gain adult and/or peer attention, acess to a desired object or activity, or sensory stimulation
How many antecedents for a behavior?	At least 1
How many consequences for a behavior	At least 1
How many single case research designs exist?	Many, for example: AB, ABA, ABAB[b], multiple baseline, changing criterion
How many intervention strategies can be applied to reduce the occurrence of a behavior?	Unlimited number. (Examples: token economy, response cost, shape, etc.)
How many observers can a behavior have?	Unlimited number
Who gathers data about a behavior?	Only individuals of the class Observer
How many statistical methods can be applied to analyze the raw data?	Unlimited number

[a]FBA: Functional Behavioral Assessment.

[b]AB is a design with a baseline phase with repeated measurements. ABA and ABAB are withdrawal designs. The intervention is concluded or stopped for some period of time before it is begun again.

According to these researchers, while the ontology validation is a process to evaluate how much an ontology is well-founded and corresponds accurately to the real world, the "ontology verification" aims to evaluate whether the way in which it is produced is correct.

A wide range of approaches and methodologies can be applied to perform the ontology validation. Brank et al [47] grouped the most common methods in four categories: (1) methods comparing the ontology with a golden standard [48-53], (2) methods based on the inductive evaluation of the results obtained through the application of the ontology [54-56], (3) methods comparing the ontology with resources specialized in the ontology domain [57,58], and (4) methods based on the assessment provided by expert humans [59-62].

The evaluation strategy adopted for the FBA-ontology was based on human evaluation, and it was aimed to assess whether the ontology definition is coherent with the knowledge domain. In this case, domain experts have been interviewed to check the formal structure of the ontology regarding taxonomy, relationships, and axioms.

To let the experts in the FBA domain assess the ontology authors created a questionnaire composed of 13 questions. Questions were aimed to evaluate the issues of the ontology. In the case of concepts, experts have to rate how much they agree with a set of 12 definitions using a 5-point Likert scale (from strongly agree to strongly disagree). For relations between concepts, experts have to rate the appropriateness of 6 statements describing the links between the main concepts of the ontology using a 5-point Likert scale (from strongly agree to strongly disagree). In concepts' attributes experts have to rate how strong the relationship between 6 concepts and 9 related attributes is through a 5-point Likert scale (from strongly related to unrelated).

Moreover, questions about demographics were included in the questionnaire to gather information about the sex, age, and level of expertise of the respondents. The following tables report concepts and attributes (Table 2) and relationships (Table 3) evaluated by the questionnaire items.

Table 2. Concepts and attributes of the FBA-Ontology assessed during the human-based evaluation.

Concepts	Attributes
FBA[a]	description
Behavior	description setting (ie, school, home, etc.) place
Function of a behavior	function_categories is_main_function
Typologies of behavior's functions	category (ie, social attention, avoidance, etc)
Methods to gather information about behaviors	hasDescription isDirect
Behavioral intervention	type (reactive or proactive)
Intervention strategy	interventionType
Antecedent	description
Consequence	description
Observer	role
Research design	type (ie, AB, ABA, ABAB, etc)[b]
Data analysis	statistical method results
ResearchPhase	sequence_num

[a]FAB: Functional Behavioral Assessment.

[b]AB: is a design with a baseline phase with repeated measurements. ABA and ABAB are withdrawal designs. The intervention is concluded or stopped for some period of time before it is begun again.

Table 3. Relationships between concepts of the FBA-Ontology assessed during the human-based evaluation.

Relation Name	Concepts in relation
hasPhase	Research Design-Research Phases
hasFunction	Behavior-Function
has Antecedent	Behavior-Antecedent
hasConsequence	Behavior-Consequence
hasObserver	Observer-Observation
toAnalise	DataAnalysis-ResearchDesign
hasRawData	Observation-RawData
hasIntervention	Research Phase-InterventionStrategy
hasMethod	FBA[a]-Method
useToCollectInformationAbout	Method-ABC[b]

[a]FAB: Functional Behavioral Assessment.

[b]ABC is a chart to collect information about a behavior that occur in a context.

Results

Principal Findings

A total of 29/84 (34.52%) people accessed the survey and completed the responses. The mean age of the valid subset was 32 (SD 6.34) years with a range of 24-57 years. The respondents were mainly female (89.66%). Participants worked in the FBA domain with a mean of 5 (SD 6.87) years.

Figure 3 shows how the expert of the domain assessed the 12 concept definitions provided in the first section of the survey. The majority of responses confirmed the proposed definitions. The response rate is higher for "agree" (15/29, 51.15%) and "strongly agree" (9/29, 30.75%). It is worth noticing that the concept definition 3 and 9 received the higher rate of undecided responses, respectively 9/29 (31.03%) and 7/29 (24.14%). In both cases, the items were probably ambiguous to the experts and not straightforward. The definition number 3 is: "The

function of a behavior is the reason that motivates a behavioral topography." This sentence seems to wrongly suggest that a behavioral topography depends on its' function. However, authors wanted to get a confirmation that a function determines why a certain behavior occurred. The definition number 9 is "An observer is a person who registers qualitative and quantitative information about a behavior." This item probably does not provide enough contextual information to responders. The definition could probably be improved by adding some information about FBA and the role of the observer in the data collection of single case research designs.

The experts' evaluation about the correctness of the relationships between some of the most relevant concepts of the ontology is reported in Figure 4. The majority of respondents were "agree" (15/29, 52.87%) and "strongly agree" (9/29, 32.18%) with the proposed statements. The most controversial relationship is the number 2 ("A research phase must contain a minimum one measure") that obtained the lower agreement rate of the section (18/29, 62.07%) and a large percentage of undecided (9/29, 31.03%). Once again, rather than indicating a problem with the ontology structure, the item is probably not well expressed (the verb "contain" is not self-explanatory) and lacks contextual information about the single case research design.

Finally, as shown in Figure 5, experts confirmed, cohesively, the relationships between the proposed concepts and their relative attributes. The higher response rate (24/29, 84.29%) was for the "related" and "strongly related" options that obtained respectively the 15/29 (51.72%) and the 9/29 (32.57%) of the overall responses. Just a few responses (4/29, 15.71%) report disagreements among the proposed attributes.

Figure 3. Evaluation of FBA-Ontology concepts.

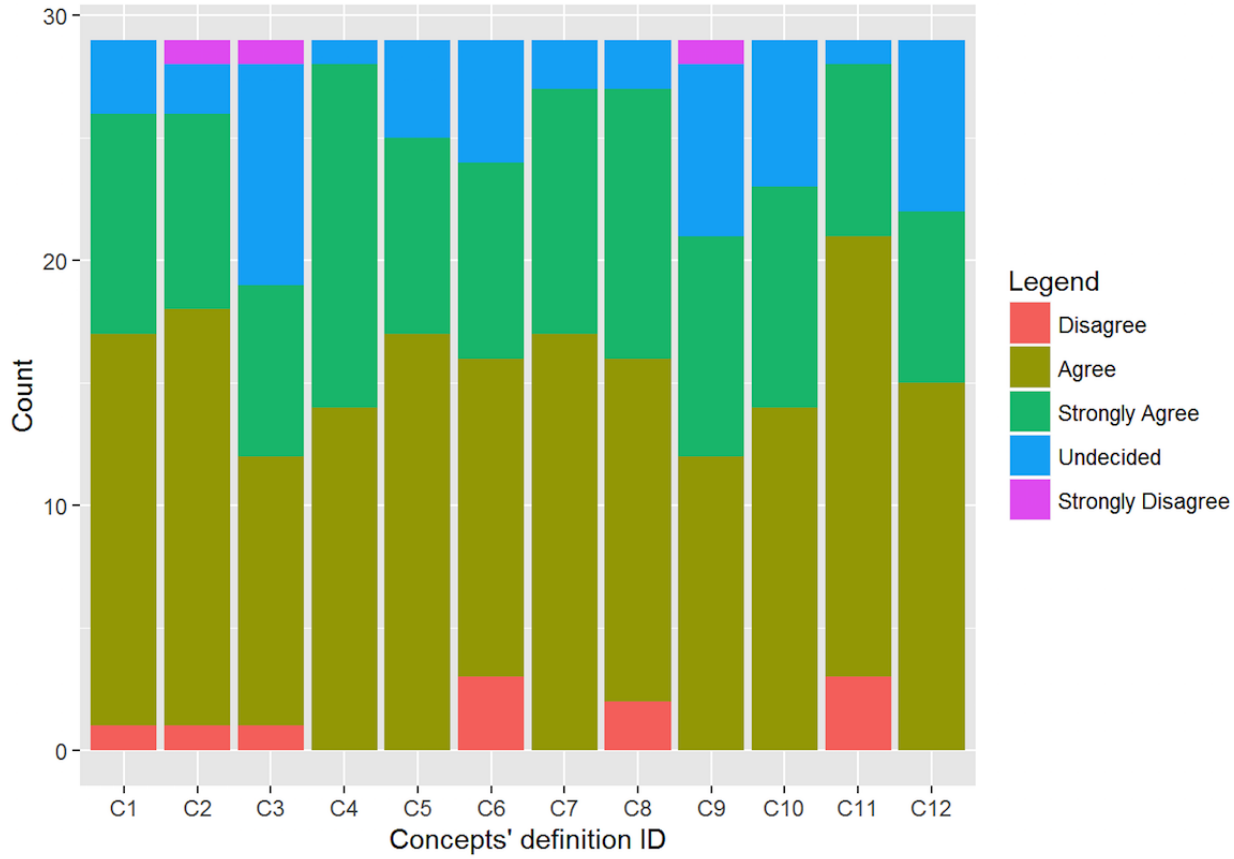


Figure 4. Evaluation of the relationships between FBA-Ontology concepts.

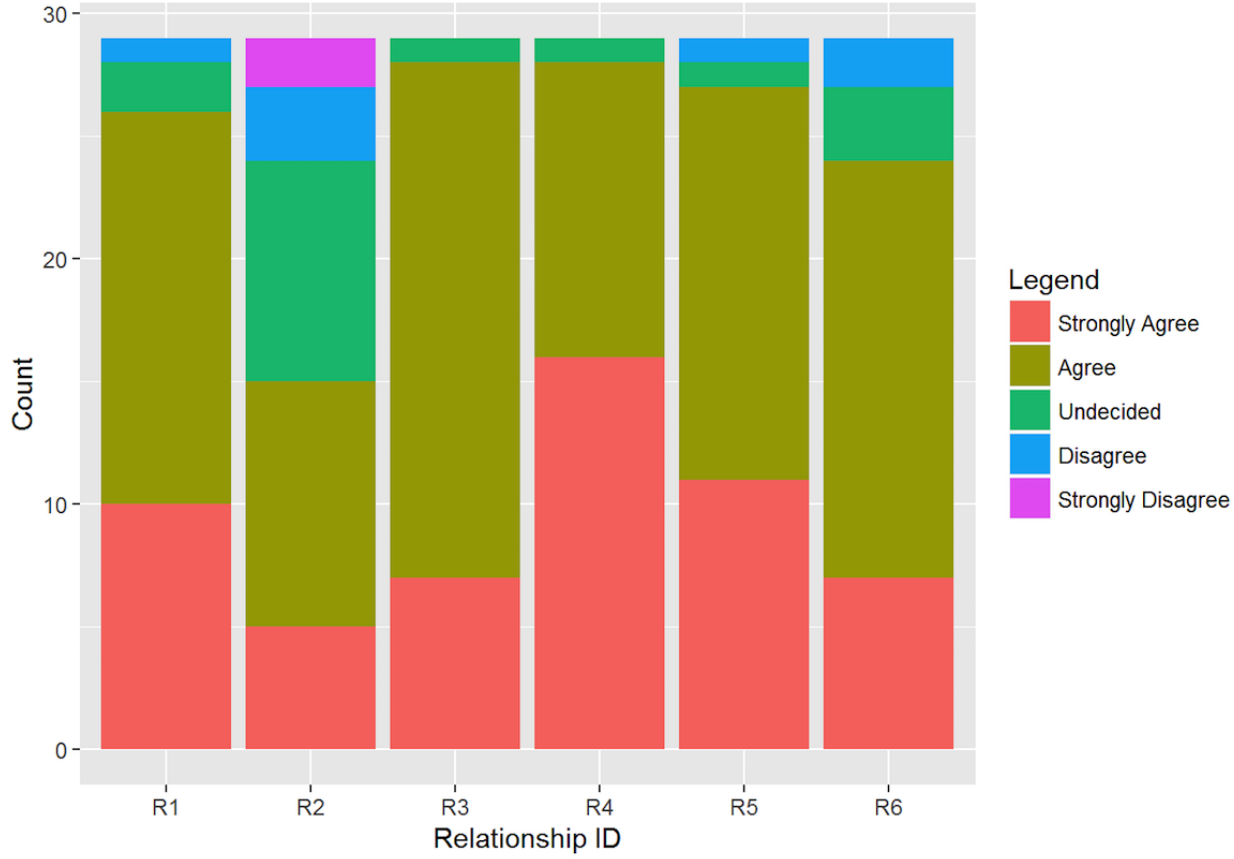
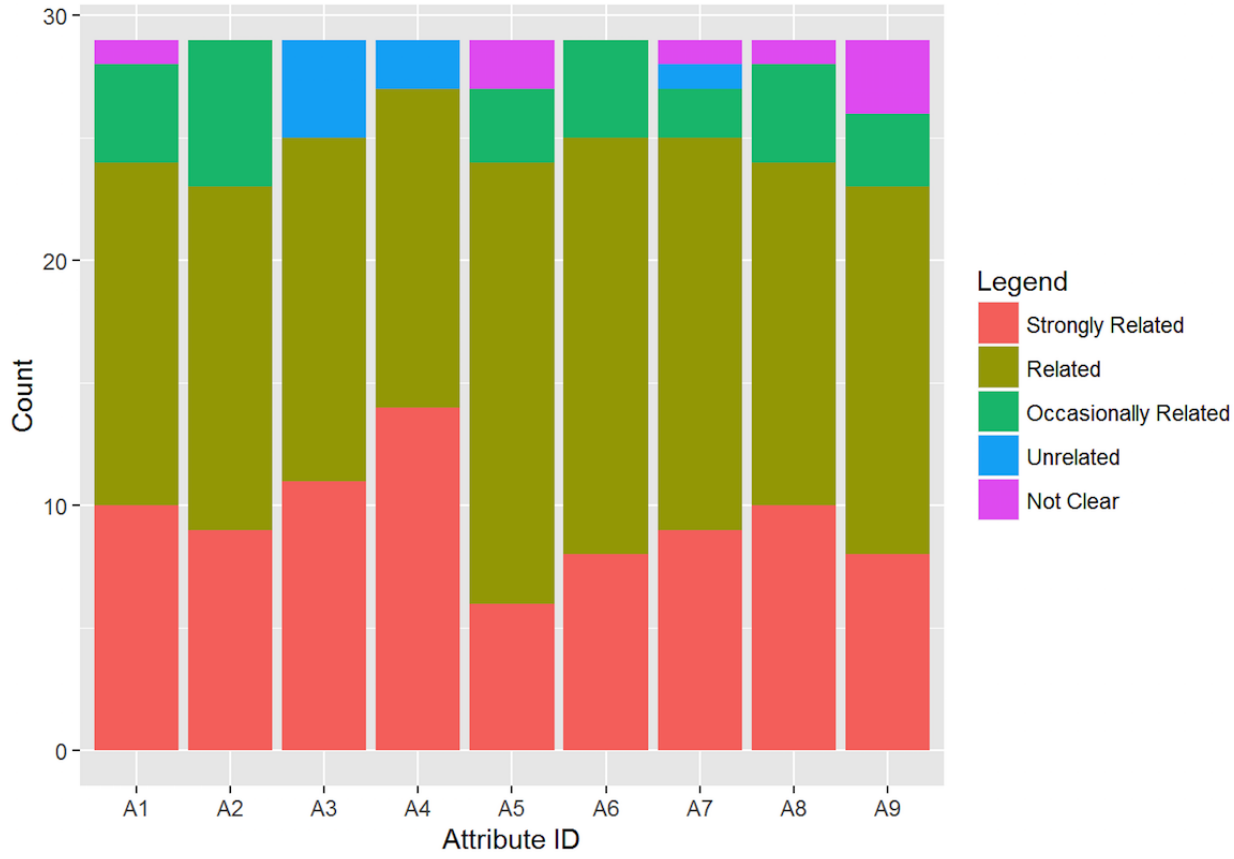


Figure 5. Evaluation of the FBA-Ontology attributes.



Discussion

Principal Findings

The FBA-Ontology describes the structure and the semantics of the FBA methods supporting the systematic collection of behavioral data, the definition of hypotheses about the function of a behavior, the consequential planning of treatment strategies, and the evidence-based evaluation of the efficacy of the applied treatments.

In the field of behavioral science, the mixing of terms and labels is frequent; this lack of common terms and shared definitions for interventions renders the aggregation of knowledge a difficult process [16]. An ontology, which provides a controlled vocabulary of agreed terms and their relationships, enables and facilitates new approaches in behavioral science. Data collected by experts are no longer collected only to be used in their research, but they can be shared, compared and integrated across experiments conducted by the whole research community. In this perspective, the FBA-Ontology represents a model able to promote the creation of new repositories, the integration, and interlinking of Linked Open Data datasets in the field of BI. It represents an open approach for sharing and exchanging data, explicating common mechanisms of action, collecting behavioral patterns, classifying contingency variables according to behavioral patterns, monitoring the statistical evidence of behavioral intervention.

Besides, the FBA-Ontology could favor the development of new applications able to support the collection of observational data in different life contexts, facilitating the interaction among practitioners and caregivers. In general, the FBA-Ontology supports the integration of several sources of data thus constituting a key element to enhance the value of the data itself. In educational settings, the presence of innovative applications could improve Lifelong Learning opportunities for teachers, parents, and clinicians to spread the use of the behavioral observation practices and the promotion of home-school relationships, to reduce the gap between research and practice. Also, the dissemination of a common communication language and the improvement of effective evidence-based decision-making processes will be advantageous from this perspective.

Concerning the ontology verification, a machine-based approach has been applied to check the logical consistency, to which the integrity of classes and properties of FBA-Ontology have not detected any violation of constraints. Moreover, the result of the questionnaire administered to domain experts confirmed that the concepts, the attributes, and the relationships between concepts defined in the FBA-Ontology are valid. These findings are particularly important for behavioral science because they contribute to improve class definitions and comparability of operational definitions, and to enable automatic and efficient meta-analysis and scientific syntheses, which, in turn, could be translated into clinical guidelines [16].

The FBA-Ontology, developed contextually to the Web Health Application for ADHD Monitoring (WHAAM) [63-65], could be a starting point to guarantee the systematic organization of behavioral knowledge and the development of future eHealth systems devoted to spreading the digital use of evidence-based assessment practices. A limitation of the work presented here concerns the lack of practical use cases in which the ontology has been adopted. This issue will be tackled in two European funded projects recently approved in the framework of the Erasmus+ program.

These projects are respectively focused on the management of social, emotional and behavioral difficulties, and the promotion of positive behaviors at school, thus offering a suitable setting to conduct further experimentations of the FBA-Ontology in a real environment. Finally, we aim to encourage not only the empirical application but also the use of computational tools and psychometric methods to provide the refinement of ontology in the future, aware that this field of study needs to be explored more in-depth.

Conclusion

In this study, we developed the FBA Ontology to promote knowledge generation, organization, reuse, integration, and analysis of behavioral data. The FBA Ontology is composed of concepts that describe the process of gathering information about behavior to determine its function and design effective intervention plans.

The study presented the assessment of the ontology by a group of experts in the domain. Results from the human-based evaluation confirmed that the ontology concepts, attributes, and relationships between concepts are valid. Moreover, the analysis provided by automatic tools has not identified anomalies in the ontology definition. Further research involving the creation and the interlink of repositories based on the behavioral data would contribute to highlight the importance of the aggregation and sharing of information in this domain.

Acknowledgments

The authors would like to thank the experts who took part in the validation of the FBA Ontology, as well as everyone who supported the writing of the article.

Authors' Contributions

All authors read and approved the final manuscript and have contributed equally to the work.

Conflicts of Interest

None declared.

References

1. Pelham WE, Wheeler T, Chronis A. Empirically supported psychosocial treatments for attention deficit hyperactivity disorder. J Clin Child Psychol 1998 Jun;27(2):190-205. [doi: 10.1207/s15374424jccp2702_6] [Medline: 9648036]
2. Ghuman JK, Arnold LE, Anthony BJ. Psychopharmacological and other treatments in preschool children with attention-deficit/hyperactivity disorder: current evidence and practice. J Child Adolesc Psychopharmacol 2008 Oct;18(5):413-447 [FREE Full text] [doi: 10.1089/cap.2008.022] [Medline: 18844482]
3. Pelham JW, Fabiano G. Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. Journal of Clinical Child & Adolescent Psychology 2008;37(1):184-214.
4. Fabiano GA, Pelham WE, Coles EK, Gnagy EM, Chronis-Tuscano A, O'Connor BC. A meta-analysis of behavioral treatments for attention-deficit/hyperactivity disorder. Clin Psychol Rev 2009 Mar;29(2):129-140. [doi: 10.1016/j.cpr.2008.11.001] [Medline: 19131150]
5. Rajwan E, Chacko A, Moeller M. Nonpharmacological interventions for preschool ADHDtate of the evidence and implications for practice. Professional Psychology: Research and Practice 2012;43(5):520.
6. Horner RH. Functional assessment: Contributions and future directions. J Appl Behav Anal 1994;27(2):401-404 [FREE Full text] [Medline: 16795831]
7. Witt J, Daly E, Noell G. Functional assessments: A step-by-step guide to solving academic and behavior problems. Longmont, Sopris West 2000.
8. Newcomer L, Lewis T. Functional behavioral assessment an investigation of assessment reliability and effectiveness of function-based interventions. Journal of Emotional and Behavioral Disorders 2004;12(3):168-181.
9. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [Medline: 8555924]
10. Slocum TA, Detrich R, Wilczynski SM, Spencer TD, Lewis T, Wolfe K. The Evidence-Based Practice of Applied Behavior Analysis. Behav Anal 2014 May;37(1):41-56 [FREE Full text] [doi: 10.1007/s40614-014-0005-2] [Medline: 27274958]
11. Kazdin AE. Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. Am Psychol 2008 Apr;63(3):146. [doi: 10.1037/0003-066X.63.3.146] [Medline: 18377105]
12. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Scientific Standards. Promoting an open research culture. Science 2015 Jun 26;348(6242):1422-1425 [FREE Full text] [doi: 10.1126/science.aab2374] [Medline: 26113702]
13. Scott T, Alter P. Examining the case for functional behavior assessment as an evidence-based practice for students with emotional and behavioral disorders in general education classrooms. Preventing School Failure: Alternative Education for Children and Youth 2017 2017;61(1):80-93.
14. Richesson RJ. Introduction to Clinical Research Informatics. Richesson R and Andrews J (eds) Clinical Research Informatics. 1st ed. London: Springer 2012:3-16.
15. Musen M. Modern architectures for intelligent systems: reusable ontologies and problem-solving methods. 1998 Presented at: AMIA Symposium; 1998; Stanford p. 46-52.
16. Larsen KR, Michie S, Hekler EB, Gibson B, Spruijt-Metz D, Ahern D, et al. Behavior change interventions: the potential of ontologies for advancing science and practice. J Behav Med 2017 Feb;40(1):6-22. [doi: 10.1007/s10865-016-9768-0] [Medline: 27481101]
17. What is the gene ontology?. URL: http://www.geneontology.org/page/introduction-go-resource [accessed 2018-05-02] [WebCite Cache ID 6z6yYvmFZ]
18. Smith, Barry. Relations in biomedical ontologies. Genome biology 2005;6(5):6-R46.
19. Hastings J. Representing mental functioning: Ontologies for mental health and disease. Representing mental functioning: Ontologies for mental health and disease. 3rd International Conference on Biomedical Ontology; University of Graz; 2012 Presented at: 3rd International Conference on Biomedical Ontology; July 21-25, 2012; Graz.
20. Hadzic M, Meifania C, Tharam S. Towards the mental health ontology. Bioinformatics and Biomedicine; IEEE International Conference 2008.
21. Haghighi M, Koeda M, Takai T, Tanaka H. Development of clinical ontology for mood disorder with combination of psychomedical information. J Med Dent Sci 2009 Mar;56(1):1-15. [Medline: 19697514]
22. McCray AT, Trevvett P, Frost HR. Modeling the autism spectrum disorder phenotype. Neuroinformatics 2014 Apr;12(2):291-305 [FREE Full text] [doi: 10.1007/s12021-013-9211-4] [Medline: 24163114]
23. Toyoshima F. Towards an Ontology of Schizophrenia. ICBO/BioCreative 2016.
24. Lovrencic S, Vidacek-Hains V, Kirinic V. Use of ontologies in monitoring mental retardation rehabilitation process. In: Katalinic B editor. DAAAM International Scientific Book. Vienna: DAAAM International Vienna; 2010:289-301.
25. Napierski D, Young A, Harper K. Towards a common ontology for improved traceability of human behavior models. 2004 Presented at: Conference on Behavior Representations in Modeling and Simulation; 19 May 2004; Arlington, Virginia.
26. Hershenberg R, Drabick DAG, Vivian D. An opportunity to bridge the gap between clinical research and clinical practice: implications for clinical training. Psychotherapy (Chic) 2012 Jun;49(2):123-134 [FREE Full text] [doi: 10.1037/a0027648] [Medline: 22642520]

27. Taibi D, Chiazzese G, Merlo G, Seta L. An ontology to support evidence-based Functional Behavioral Assessments. 2015 Presented at: 2nd International Conference on Medical Education Informatics; 2015 June 18-20; Thessaloniki p. 18-20.
28. Uschold M, King M. Towards a methodology for building ontologies. 1995 Presented at: Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing held in conjunction with IJCAI; 1995; Montreal, Canada.
29. Gresham F, Watson T, Skinner C. Functional behavioral assessment: Principles, procedures, and future direction. School Psychology Review 2001;30(2):156.
30. Skinner B. Science and human behavior. Simon, Schuster 1953.
31. Skinner B. Operant behavior. American Psychologist 1963;18(8):503.
32. Hanley GP. Functional assessment of problem behavior: dispelling myths, overcoming implementation obstacles, and developing new lore. Behav Anal Pract 2012;5(1):54-72 [FREE Full text] [doi: 10.1007/BF03391818] [Medline: 23326630]
33. Kozlowski A, Matson J. Interview and observation methods in functional assessment. Matson JL (eds) Functional assessment for challenging behaviors. New York: Springer, pp 2012:105-124.
34. Iwata BA, Dorsey MF, Slifer KJ, Bauman KE, Richman GS. Toward a functional analysis of self-injury. J Appl Behav Anal 1994;27(2):197-209 [FREE Full text] [doi: 10.1901/jaba.1994.27-197] [Medline: 8063622]
35. Nock M, Michel B, Photos V, McKay D. Single-case research designs. Handbook of research methods in abnormal and clinical psychology. 7th ed. United States of America: SAGE Publications Ltd., pp 2008:337-350.
36. Dallery J, Raiff BR. Optimizing behavioral health interventions with single-case designs: from development to dissemination. Transl Behav Med 2014 Sep;4(3):290-303 [FREE Full text] [doi: 10.1007/s13142-014-0258-z] [Medline: 25264468]
37. Grüninger M, Fox M. Methodology for the Design and Evaluation of Ontologies. 1995 Presented at: IJCAI 95- Workshop on Basic Ontological Issues in Knowledge Sharing; April 13, 1995; Montreal, Canada URL: http://citeseer.ist.psu.edu/grninger95methodology.html
38. Uschold M. Building ontologies: Towards a unified methodology. Technical report-university of Edinburgh artificial intelligence applications institute AIAI TR 1996.
39. Mariano FL, Gómez-Pérez A. Overview and analysis of methodologies for building ontologies. The Knowledge Engineering Review 2002;17(2):129-156.
40. Bezerra C, Freitas F, Santana F. Evaluating ontologies with competency questions. Web Intelligence/IAT Workshops. IEEE Computer Society 2013;284:285.
41. Burton-Jones A, Storey V, Sugumaran V, Ahluwalia P. A semiotic metrics suite for assessing the quality of ontologies. Data & Knowledge Engineering 2005;55(1):84-102.
42. Poveda-Villalón M, Suárez-Figueroa M, Gómez-Pérez A. Validating ontologies with OOPS! In: ten Teije, A. (eds) Knowledge Engineering and Knowledge Management. Heidelberg, Germany: Springer Berlin Heidelberg, pp 2012:267-281.
43. Presutti V, Blomqvist E, Daga E, Gangemi A. Pattern-based ontology design. Suárez-Figueroa MC, Gómez-Pérez A, Motta E. Gangemi A (eds) Ontology Engineering in a Networked World. Heidelberg, Germany: Springer Berlin Heidelberg, pp 2012:35-64.
44. Sirin E, Parsia B, Grau B, Kalyanpur A, Katz Y. Pellet: A practical owl-dl reasoner. Web Semanticscience, services and agents on the World Wide Web 2007;5(2):51-53.
45. Gómez Pérez A. Evaluation of ontologies. International Journal of intelligent systems 2001;16(3):391-409.
46. Sabou M, Fernandez M. Ontology (network) evaluation. Suárez-Figueroa MC, Gómez-Pérez A, Motta EGangemi A. (eds) Ontology engineering in a networked world. 1st ed. BerlSpringer Berlin Heidelberg 2012:193-212.
47. Brank J, Grobelnik M, Mladenic D. A survey of ontology evaluation techniques. 2005 Presented at: In: Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD), Ljubljana, Slovenia, , pp; 2005; Ljubljana, Slovenia p. 166-170.
48. Cheatham M, Hitzler P. String similarity metrics for ontology alignment. 2013 Presented at: The Semantic Web - ISWC 2013 12th International Semantic Web Conference; 21 - 25 October 2013; Sydney, Australia.
49. Gan M, Dou X, Jiang R. From ontology to semantic similarity: calculation of ontology-based semantic similarity. The Scientific World Journal 2013.
50. Euzenat J, Shvaiko P. Ontology matching. 2nd ed. Heidelberg: Springer 2013.
51. Maedche A, Staab S. Measuring similarity between ontologies. Benjamins VR (ed). Knowledge engineering and knowledge management: Ontologies and the semantic web 2002:251-263.
52. Shvaiko P, Euzenat J. Ontology matchingtate of the art and future challenges. IEEE Transactions on knowledge and data engineering 2013;25(1):158-176.
53. Zavitsanos E, Paliouras G, Vouros G. Gold standard evaluation of ontology learning methods through ontology. IEEE Transactions on knowledge and data engineering 2011;23(11):1635-1648.
54. Fernández M, Overbeeke C, Sabou M, Motta E. What makes a good ontology? A case-study in fine-grained knowledge reuse. Gómez-Pérez A, Yu YDing Y. (eds) The semantic web. New York: Springer-Verlag 2009:61-75.
55. Hastings J, Brass A, Caine C, Jay C, Stevens R. Evaluating the Emotion Ontology through use in the self-reporting of emotional responses at an academic conference. Journal of biomedical semantics 2014;5(1):5-38.
56. Porzel R, Malaka R. A task-based approach for ontology evaluation. 2004 Presented at: InCAI Workshop on Ontology Learning and Population; 22 - 23 August 2004; Valencia, Spain.

57. Brewster C, Alani H, Dasmahapatra S, Wilks Y. Data-driven ontology evaluation. 2004 Presented at: In: Proceedings of the Language Resources and Evaluation Conference (LREC); 26 - 28 May 2004; Lisbon, Portugal p. 164-168.
58. Gordon CL, Weng C. Combining expert knowledge and knowledge automatically acquired from electronic data sources for continued ontology evaluation and improvement. J Biomed Inform 2015 Oct;57:42-52 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.014] [Medline: 26212414]
59. Casellas N. Ontology evaluation through usability measures. In: Meersman R, Herrero P. Dillon T (eds) On the Move to Meaningful Internet Systems: OTM 2009;Workshops. Heidelberg, Germany:594-603.
60. 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 Jun 28;15(6) [FREE Full text] [doi: 10.2196/jmir.2512] [Medline: 23811542]
61. Lozano-Tello A, Gómez-Pérez A. Ontometric: A method to choose the appropriate ontology. Journal of Database Management 2004;2(15):1-18.
62. Zemmouchi-Ghomari L, Ghomari A. A New Approach for Human Assessment of Ontologies. 2013 Presented at: Third International Conference of Information Systems and Technologies, ICIST; March 2013; Tangier, Morocco.
63. Chifari A, Maia M, Bamidis P, Doherty G, Bilbow A, Rinaldi R. A Project to Foster Behavioural Monitoring in the Field of ADHD. 2013 Presented at: The Future of Education; 13 - 14 June 2013; Firenze, Italy.
64. Alves S, Bamidis P, Bilbow A, Callahan A, Chiazzese G, Chifari A, et al. Palermo, Italy: CNR - Istituto per le Tecnologie Didattiche. 2013. Web Health Application for Adhd Monitoring Context - Driven Framework URL: https://www.whaamproject.eu/images/documents/WhaamFrameworkEN.pdf [accessed 2018-05-02] [WebCite Cache ID 6z708Jj2A]
65. Spachos D, Chiazzese G, Merlo G, Doherty G, Chifari A, Bamidis P. WHAAM: A mobile application for ubiquitous monitoring of ADHD behaviors. 2014 Presented at: International Conference on Interactive Mobile Communication Technologies and Learning; 13 - 14 November 2014; Thessaloniki, Greece p. 305-309.

Abbreviations

ADHD: Attention Deficit Hyperactivity Disorder
BI: behavioral interventions
EBP: evidence-based practice
FBA: functional behavioral assessment
WHAAM: web health application for ADHD monitoring

Edited by G Eysenbach; submitted 15.05.17; peer-reviewed by HA Park, R Shrestha; comments to author 12.07.17; revised version received 05.09.17; accepted 14.03.18; published 31.05.18.

Please cite as:
Merlo G, Chiazzese G, Taibi D, Chifari A
Development and Validation of a Functional Behavioural Assessment Ontology to Support Behavioural Health Interventions
JMIR Med Inform 2018;6(2):e37
URL: http://medinform.jmir.org/2018/2/e37/
doi:10.2196/medinform.7799
PMID:

Original Paper

Patient-Physician Communication in the Era of Mobile Phones and Social Media Apps: Cross-Sectional Observational Study on Lebanese Physicians' Perceptions and Attitudes

Fady Daniel[1], MD, MSc; Suha Jabak[1], MD; Roula Sasso[2], MD; Yara Chamoun[3], MD; Hani Tamim[4], PhD

[1]Department of Internal Medicine, American University of Beirut Medical Center, American University of Beirut, Beirut, Lebanon

[2]Department of Emergency Medicine, American University of Beirut Medical Center, American University of Beirut, Beirut, Lebanon

[3]Hotel-Dieu de France Medical Center, Department of Psychiatry, Université Saint Joseph, Beirut, Lebanon

[4]Biostatistics Unit, Clinical Research Institute, American University of Beirut Medical Center, American University of Beirut, Beirut, Lebanon

Corresponding Author:
Fady Daniel, MD, MSc
Department of Internal Medicine
American University of Beirut Medical Center
American University of Beirut
Hamra, Cairo Street
Beirut,
Lebanon
Phone: 961 3614660
Fax: 961 1 370814
Email: fd21@aub.edu.lb

Abstract

Background: The increased prevalence of virtual communication technology, particularly social media, has shifted the physician-patient relationship away from the well-established face-to-face interaction. The views and habits of physicians in Lebanon toward the use of online apps and social media as forms of patient communication have not been previously described.

Objective: The aim of this study is to describe the views of Lebanese physicians toward the use of social media and other online apps as means of patient communication.

Methods: This was a cross-sectional observational study using an online survey that addressed physicians' perceptions on the use of virtual communication in their clinical practice. The study took place between April and June 2016, and was directed toward physicians at the American University of Beirut Medical Center.

Results: A total of 834 doctors received the online survey, with 238 physicians completing the survey. Most of the participants were from medical specialties. Most responders were attending physicians. Less than half of the respondents believed that Web-based apps and social media could be a useful tool for communicating with patients. Email was the most common form of professional online app, followed by WhatsApp (an instant messaging service). The majority of participants felt that this mode of communication can result in medicolegal issues and that it was a breach of privacy. Participants strictly against the use of virtual forms of communication made up 47.5% (113/238) of the study sample.

Conclusions: The majority of physicians at the American University of Beirut Medical Center are reluctant to use virtual communication technology as a form of patient communication. Appropriate policy making and strategies can allow both physicians and patients to communicate virtually in a more secure setting without fear of breaching privacy and confidentiality.

(JMIR Med Inform 2018;6(2):e18) doi:10.2196/medinform.8895

KEYWORDS

social media; communication; patient-physician communication; technology use

Introduction

The medical world is changing and the use of online communication is becoming more abundant [1]. The widespread availability of Internet-connected mobile phones has introduced new virtual ways of communication between individuals including social media apps. Patients are now emailing, text messaging, chatting, and video chatting with their physicians to speed up their health care [2]. The patient-physician relationship that has been traditionally implemented by face-to-face communication may be slowly shifting toward a more virtual form of communication [3,4] because of the increased use of social media and social networking.

Online physician-patient communication can improve health care by enhancing patient education [5], improving patient compliance to medication use, bettering adherence to physician recommendations [6,7], and facilitating easier patient follow-up for chronic diseases [8].

Currently, there are no studies in Lebanon that describe physicians' perceptions on the use of virtual communication in patient care. Since social media is becoming a more abundant form of communication in clinical practice, we designed this study with the aim of describing how social media and networking among other virtual communication technologies are regarded and utilized by practicing and training physicians in Lebanon.

Methods

Participants

Our study was a cross-sectional observational study that took place between April 2016 and June 2016 at the American University of Beirut Medical Center (AUBMC), a large tertiary care academic medical center in Beirut.

Practicing and training physicians from all medical specialities at AUBMC with an email address in the AUBMC directory were eligible to participate in our study. Online questionnaires were distributed via an online survey tool (Lime survey).

Responders to the questionnaires remained anonymous and individuals who completed the online questionnaire did not receive any financial compensation. The institutional review board committee at AUBMC approved the study.

Questionnaire

Our research team developed the questionnaire after a thorough literature review and the questions were developed based on the currently available literature and our basic research question. The questionnaire included 22 questions that examined general demographics as well as three scopes of the doctor-patient online interaction: (1) extent of participants' personal use of online apps and social media, such as email, LinkedIn (a business community networking site), Instagram (a photo-sharing Web service), WhatsApp (an instant messaging service), and Facebook and Twitter (social networking websites); (2) participants' opinions on the use of virtual communication in patient care; and (3) participants' views of development of a regulated platform for the use of virtual communication to aid patient care.

Data Collection and Analysis

All responses to the online questionnaire were automatically recorded through the Lime Survey platform and downloaded to SPSS software version 19.0. A descriptive analysis was obtained followed by a bivariant analysis to detect statistical associations between the participant independent variables and their standpoint (with/against/neutral) on the use of social media in the medical setting. A *P* value of .05 or less was considered statistically significant.

Results

Participants

Of the 834 doctors who received the invitation email, 270 participated in the survey but only the 238 physicians who fully completed the survey were included in our data analysis, yielding an overall response rate of 28.5%. The participants were almost equally distributed by gender with males making up 55.0% (131/238) of the responders. The mean age of the participants was 39.4 (SD 13.3) years and the majority of responders were attending physicians (57.1%, 136/238). Most participants were from medical specialties (183/238, 76.9%), which included internal medicine and its various subspecialties: family medicine, radiology, dermatology, psychiatry, pediatrics, pathology, emergency medicine, neurology, and laboratory medicine. The responders' demographics and characteristics are reported in Table 1.

Online App Usage Patterns

All physicians reported any-purpose use of online apps in the last 6 months. All participating physicians used email in the given time frame. The second most common form of any-purpose online app for communication was WhatsApp (230/238, 96.6%) followed by Facebook (177/238, 74.4%), and to a lesser extent LinkedIn (91/238, 38.2%) and Twitter (54/238, 22.7%). The use of email and LinkedIn for professional purposes was higher than their use for personal purposes (230/238, 96.6% vs 167/238, 70.2% and 79/238, 33.2% vs 34/238, 14.3%, respectively). Physicians used the remaining online apps more frequently for personal purposes (Table 2).

Attitudes and Opinions Toward the Use of Online Apps by Physicians

The use of online apps including social media in a professional setting was regarded to aid the communication between different physicians according to 70.2% (167/238) of our participants. Only 42.4% (101/238) of the respondents believed that the use of online apps including social media can be a beneficial tool for patient education. Furthermore, the idea that this virtual form of communication can be used to improve patient health and treatment compliance was only shared by 34.0% (81/238) of our respondents. Approximately half of the participants believed that the use of these online apps and social media can be of use for patients to communicate with one another to share experiences and receive reassurance about their medical condition (Table 3).

Table 1. Physician demographics and characteristics (N=238).

Demographics and characteristics	Participants
Age (years), mean (SD)	39.4 (13.3)
Gender, n (%)	
Male	131 (55)
Female	107 (45)
Marital status, n (%)	
Single	102 (42.9)
Married	133 (55.9)
Divorced	2 (0.8)
Widowed	1 (0.4)
Specialty, n (%)	
Medicine	183 (76.9)
Surgery	55 (23.1)
Years in practice, n (%)	
<5 years	112 (47.1)
5-10 years	29 (12.2)
10-15 years	28 (11.8)
15-20 years	17 (7.1)
>20 years	52 (21.8)
Number of patients seen weekly, n (%)	
<10	15 (6.3)
10-20	52 (21.8)
20-40	80 (33.6)
40-60	44 (18.5)
>60	47 (19.7)
Medical status, n (%)	
Resident/fellow	102 (42.9)
Attending	136 (57.1)

Table 2. Online app use according to purpose.

Online app forms	Any-purpose use, n (%)		Personal purpose, n (%)		Professional purpose, n (%)	
	Yes	No	Yes	No	Yes	No
Email	238 (100)	0 (0)	167 (70.2)	71 (29.8)	230 (96.6)	8 (3.4)
WhatsApp	230 (96.6)	8 (3.4)	222 (93.3)	16 (6.7)	166 (69.7)	72 (30.3)
LinkedIn	91 (38.2)	147 (61.8)	34 (14.3)	204 (85.7)	79 (33.2)	159 (66.8)
Facebook	177 (74.4)	61 (25.6)	172 (72.3)	66 (27.7)	20 (8.4)	218 (91.6)
Twitter	54 (22.7)	184 (77.3)	52 (21.8)	186 (78.2)	19 (8.0)	219 (92.0)

Table 3. Physician standpoint on the potential benefits of virtual communication.

Benefits of virtual communication	Standpoint, n (%)	
	Yes	No
Provides quicker and more efficient communication between physicians	167 (70.2)	71 (29.8)
Decreases nonurgent telephone calls	126 (52.9)	112 (47.1)
Reassures patient about disease	120 (50.4)	118 (49.6)
Allows patients to share similar experiences (eg, on blogs and forums)	112 (47.1)	126 (52.9)
Allows better patient education	101 (42.4)	137 (57.6)
Creates continuous access to health care system	87 (36.6)	151 (63.4)
Helps monitor patients' health and improve treatment compliance	81 (34.0)	157 (66.0)
Allows physicians to handle larger number of patients	40 (16.8)	198 (83.2)

Table 4. Physician standpoint on the potential barriers of virtual communication.

Barrier	Standpoint, n (%)	
	Yes	No
Raises medicolegal issues	186 (78.6)	51 (21.4)
Patients are not able to judge authenticity of information provided online	177 (74.4)	61 (25.6)
Provides false patient reassurance	171 (71.8)	67 (28.2)
Invades physician privacy	169 (71.0)	69 (29.0)
Is unprofessional	127 (53.4)	111 (46.6)
Delays patients from visiting health care professionals	124 (52.1)	114 (47.9)
Effects patient-physician confidentiality	115 (48.3)	123 (51.7)
Increases patient anxiety	110 (46.2)	128 (53.8)
Increases physician workload	98 (41.2)	140 (58.8)
Invades patient privacy	94 (39.5)	114 (60.5)

The barriers to the use of virtual communication as a means of communicating with patients according to our participating physicians are reported in Table 4. The majority of the participants (186/238, 78.6%) felt that this mode of communication can result in medicolegal issues, and 71.0% (169/238) felt that it was a breach of privacy. Most physicians also believed that online sources of information for patients are problematic because patients are not sufficiently qualified to judge the authenticity of information presented to them by social media.

The results demonstrating the effect of participant characteristics on their attitudes toward relying on virtual communication in medical settings is shown in Table 5. Physicians' attitudes were nearly equally distributed between those who were strictly against the use of the online apps and social media (113/238, 47.5%) in their daily profession and those who were with or neutral with such use (125/238, 52.5%). Physicians who were of male gender (P=.003), of older age (P=.02), faculty members (P<.001) and in the surgical specialty (P=.03) were more likely to have positive attitudes toward the use of online apps and social media. A prior positive experience with such use in their interactions with their patients was also shown to be a positive predictor of a positive attitude. Years in practice or the number of patients seen per week did not influence physician attitude (P=.07 and P=.58, respectively).

The most common method adopted by participating physicians in dealing with unwanted online communication with patients was by adjusting their privacy settings on their online apps (143/238, 60.0%), closely followed by ignoring a friend request (105/238, 52.1%). Only 13.4% (32/238) of the participants reported blocking patients as their way to avoid communicating with patients on social media.

When asked about awareness of existing current guidelines on physician-patient communication and only 6.7% (16/238) of the participants answered "yes" and 81.5% (195/238) felt that guidelines are necessary to facilitate and direct this form of communication (Table 6).

Table 5. Influence of participant characteristics on standpoint toward using virtual communication.

Variable	Total (N=238)	Standpoint			*P* value
		With (n=63)	Neutral (n=62)	Against (n=113)	
Age (years), mean (SD)	39.36 (13.27)	43.33 (13.13)	38.19 (12.22)	37.78 (13.55)	.02
Gender, n (%)					.003
Male	131 (55.0)	46 (73.0)	28 (45.2)	57 (50.4)	
Female	107 (45.0)	17 (27.0)	34 (54.8)	56 (49.6)	
Marital status, n (%)					.52
Single	105 (44.1)	24 (38.1)	28 (45.2)	53 (46.9)	
married	133 (55.9)	39 (61.9)	34 (54.8)	60 (53.1)	
Specialty, n (%)					.03
Medicine	183 (76.9)	41 (65.1)	50 (80.6)	92 (81.4)	
Surgery	55 (23.1)	22 (34.9)	12 (19.4)	21 (18.6)	
Years in practice, n (%)					.07
<5 years	112 (47.1)	21 (33.3)	31 (50.0)	60 (53.1)	
5-20 years	74 (31.1)	22 (34.9)	21 (33.9)	31 (27.4)	
>20 years	52 (21.8)	20 (31.7)	10 (16.1)	22 (19.5)	
Medical status, n (%)					<.001
Resident/fellow	102 (42.9)	14 (22.2)	28 (45.2)	60 (53.1)	
Attending	136 (57.1)	49 (77.8)	34 (54.8)	53 (46.9)	
Experience in past, n (%)					<.001
With	70 (29.4)	50 (79.4)	12 (19.4)	8 (7.1)	
Neutral	112 (47.1)	13 (20.6)	43 (69.4)	56 (49.6)	
Against	56 (23.5)	0 (0)	7 (11.3)	49 (43.4)	

Table 6. Methods of avoiding patient communication on online apps and social media.

Method of avoidance	n (%)
Adjust privacy settings	143 (60.1)
Ignore friend requests	124 (52.1)
Ignore emails	105 (44.1)
Block people of social media	32 (13.4)

Discussion

This survey on the perception of the use of virtual communication in the medical setting at an academic center in Lebanon showed that the majority of participants are not opposed to the idea of using online apps and social media and communication in an interdisciplinary manner to communicate with other physicians. However, most felt that using it as a tool to communicate with patients would not result in an improved physician-patient interaction. The main reasons voiced by physicians for holding this standpoint were that they felt this mode of communication could result in increased medicolegal issues, could cause a breach in privacy, is unprofessional, and could cause a delay in patients visiting health care professionals. Trends in the use of social media for patient communication among physicians at our medical center compare to those of physicians in Australia [9] and the United States [10]. According to a study carried out in Australia, only a minimal number of physicians use social media as part of their professional careers [9]. A national survey in the United States showed that almost half of physicians and medical students did not believe that online services and communications could improve patient-physician communication [10].

Our results also demonstrated that older physicians and practicing/attending physicians where more accepting of the idea of using online apps and social media as a form on patient-physician communication. This is contrary to what we would expect. This could be because older physicians are more confident in their ability to maintain personal boundaries between their patients and themselves even through social media and online communication. This study also showed that physicians in the surgical speciality were less opposed to the

use of social media for patient interaction. Surgical physicians may favor this form of communication because it facilitates short-term follow-up postoperatively and provides a more effective form of communicating preoperative preparation instructions [11]. Online postoperative follow-ups are more time efficient and are effective in determining if patients need further personal care [12]. The benefits of online postoperative patient follow-up may not only be beneficial for short-term follow-ups. An ongoing study is being conducted to determine benefits of online long-term follow-up of patients' post-bariatric surgery in attempt to provide continued support to patients who may otherwise be lost to follow-up [13].

Patients with portable devices can communicate with physicians and clinicians through email, text messages, chat, and Web apps, and receive advice and recommendations without visiting the hospital or clinic resulting in better accessibility to health care and faster access to test results and reports [2]. Additionally, social media and online apps as a means of relaying messages and reminders to patients can improve the field of preventive medicine by increasing vaccination rates [14] and improving smoking cessation rates [15]. Virtual communication between health care providers and patients may also result in improved patients' understanding and compliance to physician instructions [6]. Not to mention better adherence to medications in patients with chronic diseases by providing a fast track to prescription refills [2]. However, not only physicians resist this shift in health care practice, but also patients. Patients also seem less willing to use online apps and social media and networking for health care services as compared to non-health care services [16].

Although the potential pitfalls to online physician-patient communications are concerning, the American College of Physicians published a policy statement titled "Online Medical Professionalism: Patient and Public Relationships" [17] in 2013. The policy statement emphasizes that online physician-patient communication should supplement rather than replace the traditional face-to-face encounter. To maintain patient confidentiality, protocols for storage and transfer of patient information must be established and secure networks employed. To preserve both personal and patient privacy, physicians are recommended to avoid "friending" patients on social media and to use professional profiles for communication rather than personal profiles. To minimize medicolegal issues that may arise from online communication, it is suggested that online communication only take place between a physician and a patient with an already established face-to-face relationship and prior discussion and agreements to be set. In fact, several professional opinions have advocated against online communication sharing of medical advice with an unknown patient [18]. Because of the novelty of this form of physician-patient communication, physicians do not have a lot of experience in dealing with online ethical dilemmas. Adhering to available guidelines addressing this issue is fundamental. At our academic medical center, only 6.7% (16/238) of the participants were aware of current existing guidelines on online patient-physician interaction.

Limitations of the study include a low survey response rate and survey distribution being limited to AUBMC, which significantly limits the generalizability of the study. Additionally, this study is prone to self-selection bias, in that individuals who are more comfortable using online communication were more likely to access the online questionnaire. Hence, our study could be underestimating physician disapproval toward the use of social media as a means of communicating with patients.

In conclusion, physicians at an academic medical center in Lebanon were reluctant to use online apps and social media as a form of communication with patients. Our study is the first to evaluate the usage and to describe the views of physicians toward the use of virtual communication via various online apps and social media in their daily interaction with patients. Our results provide novel information to the region, which could help in the development of polices and strategies that will result in better online physician-patient communication and make this form of communication less intimidating to physicians. Institutions need to start embracing this new form of physician-patient communication rather than fight it.

Acknowledgments

The authors appreciate the participation of individuals who completed the survey.

Authors' Contributions

FD designed the questionnaire, supervised data collection, analysis, manuscript drafting, and final editing; SJ: questionnaire editing, data collection, and manuscript drafting; RS: manuscript drafting; YC: participated in initial questionnaire design; and HT: data analysis and manuscript final editing.

Conflicts of Interest

None declared.

References

1. Sarasohn-Kahn J. California Health Care Foundation. 2008 Apr 22. The wisdom of patients: health care meets online social media URL: https://www.chcf.org/publication/the-wisdom-of-patients-health-care-meets-online-social-media/ [accessed 2018-03-13] [WebCite Cache ID 6xtJMWmJQ]

2. Chauhan B, George R, Coffin J. Social media and you: what every physician needs to know. J Med Pract Manage 2012;28(3):206-209. [Medline: 23373164]
3. Merchant RM, Elmer S, Lurie N. Integrating social media into emergency-preparedness efforts. N Engl J Med 2011 Jul 28;365(4):289-291. [doi: 10.1056/NEJMp1103591] [Medline: 21793742]
4. Adams SA. Revisiting the online health information reliability debate in the wake of "web 2.0": an inter-disciplinary literature and website review. Int J Med Inform 2010 Jun;79(6):391-400. [doi: 10.1016/j.ijmedinf.2010.01.006] [Medline: 20188623]
5. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
6. Kang X, Zhao L, Leung F, Luo H, Wang L, Wu J, et al. Delivery of instructions via mobile social media app increases quality of bowel preparation. Clin Gastroenterol Hepatol 2016 Mar;14(3):429-435. [doi: 10.1016/j.cgh.2015.09.038] [Medline: 26492848]
7. Faridi Z, Liberti L, Shuval K, Northrup V, Ali A, Katz DL. Evaluating the impact of mobile telephone technology on type 2 diabetic patients' self-management: the NICHE pilot study. J Eval Clin Pract 2008 Jun;14(3):465-469. [doi: 10.1111/j.1365-2753.2007.00881.x] [Medline: 18373577]
8. Carrasco MP, Salvador CH, Sagredo PG, Márquez-Montes J, González DM, Fragua JA, et al. Impact of patient-general practitioner short-messages-based interaction on the control of hypertension in a follow-up service for low-to-medium risk hypertensive patients: a randomized controlled trial. IEEE Trans Inf Technol Biomed 2008 Nov;12(6):780-791. [doi: 10.1109/TITB.2008.926429] [Medline: 19000959]
9. Brown J, Ryan C, Harris A. How doctors view and use social media: a national survey. J Med Internet Res 2014;16(12):e267 [FREE Full text] [doi: 10.2196/jmir.3589] [Medline: 25470407]
10. Bosslet GT, Torke AM, Hickman SE, Terry CL, Helft PR. The patient-doctor relationship and online social networks: results of a national survey. J Gen Intern Med 2011 Oct;26(10):1168-1174 [FREE Full text] [doi: 10.1007/s11606-011-1761-2] [Medline: 21706268]
11. Sculco PK, McLawhorn AS, Fehring KA, De MI. The future of social media in orthopedic surgery. Curr Rev Musculoskelet Med 2017 Jun;10(2):278-279 [FREE Full text] [doi: 10.1007/s12178-017-9412-9] [Medline: 28456946]
12. Kummerow BK, Oyefule OO, Phillips SE, Baucom RB, Holzman MD, Sharp KW, et al. Postoperative care using a secure online patient portal: changing the (inter)face of general surgery. J Am Coll Surg 2015 Dec;221(6):1057-1066. [doi: 10.1016/j.jamcollsurg.2015.08.429] [Medline: 26453260]
13. Conceição EM, Machado PPP, Vaz AR, Pinto-Bastos A, Ramalho S, Silva C, et al. APOLO-Bari, an internet-based program for longitudinal support of bariatric surgery patients: study protocol for a randomized controlled trial. Trials 2016 Mar 01;17(1):114 [FREE Full text] [doi: 10.1186/s13063-016-1246-z] [Medline: 26927479]
14. Stockwell MS, Kharbanda EO, Martinez RA, Vargas CY, Vawdrey DK, Camargo S. Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: a randomized controlled trial. JAMA 2012 Apr 25;307(16):1702-1708. [doi: 10.1001/jama.2012.502] [Medline: 22535855]
15. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin R, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
16. Jenssen BP, Mitra N, Shah A, Wan F, Grande D. Using digital technology to engage and communicate with patients: a survey of patient attitudes. J Gen Intern Med 2016 Jan;31(1):85-92. [doi: 10.1007/s11606-015-3517-x] [Medline: 26385117]
17. Farnan JM, Snyder SL, Worster BK, Chaudhry HJ, Rhyne JA, Arora VM, American College of Physicians Ethics, Professionalism and Human Rights Committee, American College of Physicians Council of Associates, Federation of State Medical Boards Special Committee on Ethics and Professionalism. Online medical professionalism: patient and public relationships: policy statement from the American College of Physicians and the Federation of State Medical Boards. Ann Intern Med 2013 Apr 16;158(8):620-627. [doi: 10.7326/0003-4819-158-8-201304160-00100] [Medline: 23579867]
18. Chretien KC, Kind T. Social media and clinical care: ethical, professional, and social implications. Circulation 2013 Apr 02;127(13):1413-1421 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.112.128017] [Medline: 23547180]

Abbreviations

AUBMC: American University of Beirut Medical Center

Edited by G Eysenbach; submitted 04.09.17; peer-reviewed by J Brown, M Alkureishi; comments to author 09.12.17; revised version received 20.12.17; accepted 21.12.17; published 06.04.18.

Please cite as:
Daniel F, Jabak S, Sasso R, Chamoun Y, Tamim H
Patient-Physician Communication in the Era of Mobile Phones and Social Media Apps: Cross-Sectional Observational Study on Lebanese Physicians' Perceptions and Attitudes
JMIR Med Inform 2018;6(2):e18
URL: http://medinform.jmir.org/2018/2/e18/
doi:10.2196/medinform.8895
PMID:29625955

Original Paper

Patient Adherence to Scheduled Vital Sign Measurements During Home Telemonitoring: Analysis of the Intervention Arm in a Before and After Trial

Branko Celler[1,2*], BSc, BE(Hons), PhD; Ahmadreza Argha[1*], BE(Hons), ME, PhD; Marlien Varnfield[2*], PhD; Rajiv Jayasena[2*], PhD

[1]Biomedical Systems Research Laboratory, University of New South Wales, Sydney, New South Wales, Australia

[2]Health and Biosecurity Business Unit, eHealth Research Program, Commonwealth Scientific and Industrial Research Organisation, Parkville, VIC, Australia

[*]all authors contributed equally

Corresponding Author:
Branko Celler, BSc, BE(Hons), PhD
Biomedical Systems Research Laboratory
University of New South Wales
Building G7
Anzac Parade
Sydney, New South Wales,
Australia
Phone: 61 2 9385 0746
Fax: 61 2 9385 5993
Email: b.celler@unsw.edu.au

Abstract

Background: In a home telemonitoring trial, patient adherence with scheduled vital signs measurements is an important aspect that has not been thoroughly studied and for which data in the literature are limited. Levels of adherence have been reported as varying from approximately 40% to 90%, and in most cases, the adherence rate usually dropped off steadily over time. This drop is more evident in the first few weeks or months after the start. Higher adherence rates have been reported for simple types of monitoring and for shorter periods of intervention. If patients do not follow the intended procedure, poorer results than expected may be achieved. Hence, analyzing factors that can influence patient adherence is of great importance.

Objective: The goal of the research was to present findings on patient adherence with scheduled vital signs measurements in the recently completed Commonwealth Scientific and Industrial Research Organisation (CSIRO) national trial of home telemonitoring of patients (mean age 70.5 years, SD 9.3 years) with chronic conditions (chronic obstructive pulmonary disease, coronary artery disease, hypertensive diseases, congestive heart failure, diabetes, or asthma) carried out at 5 locations along the east coast of Australia. We investigated the ability of chronically ill patients to carry out a daily schedule of vital signs measurements as part of a chronic disease management care plan over periods exceeding 6 months (302 days, SD 135 days) and explored different levels of adherence for different measurements as a function of age, gender, and supervisory models.

Methods: In this study, 113 patients forming the test arm of a Before and After Control Intervention (BACI) home telemonitoring trial were analyzed. Patients were required to monitor on a daily basis a range of vital signs determined by their chronic condition and comorbidities. Vital signs included noninvasive blood pressure, pulse oximetry, spirometry, electrocardiogram (ECG), blood glucose level, body temperature, and body weight. Adherence was calculated as the number of days during which at least 1 measurement was taken over all days where measurements were scheduled. Different levels of adherence for different measurements, as a function of age, gender, and supervisory models, were analyzed using linear regression and analysis of covariance for a period of 1 year after the intervention.

Results: Patients were monitored on average for 302 (SD 135) days, although some continued beyond 12 months. The overall adherence rate for all measurements was 64.1% (range 59.4% to 68.8%). The adherence rates of patients monitored in hospital settings relative to those monitored in community settings were significantly higher for spirometry (69.3%, range 60.4% to 78.2%, versus 41.0%, range 33.1% to 49.0%, P<.001), body weight (64.5%, range 55.7% to 73.2%, versus 40.5%, range 32.3% to 48.7%, P<.001), and body temperature (66.8%, range 59.7% to 73.9%, versus 55.2%, range 48.4% to 61.9%, P=.03). Adherence with

blood glucose measurements (58.1%, range 46.7% to 69.5%, versus 50.2%, range 42.8% to 57.6%, *P*=.24) was not significantly different overall. Adherence rates for blood pressure (68.5%, range 62.7% to 74.2%, versus 59.7%, range 52.1% to 67.3%, *P*=.04), ECG (65.6%, range 59.7% to 71.5%, versus 56.5%, range 48.7% to 64.4%, *P*=.047), and pulse oximetry (67.0%, range 61.4% to 72.7%, versus 56.4%, range 48.6% to 64.1%, *P*=.02) were significantly higher in males relative to female subjects. No statistical differences were observed between rates of adherence for the younger patient group (70 years and younger) and older patient group (older than 70 years).

Conclusions: Patients with chronic conditions enrolled in the home telemonitoring trial were able to record their vital signs at home at least once every 2 days over prolonged periods of time. Male patients maintained a higher adherence than female patients over time, and patients supervised by hospital-based care coordinators reported higher levels of adherence with their measurement schedule relative to patients supervised in community settings. This was most noticeable for spirometry.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12613000635763; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=364030&isReview=true (Archived by WebCite at http://www.webcitation.org/6xPOU3DpR).

(JMIR Med Inform 2018;6(2):e15) doi:10.2196/medinform.9200

KEYWORDS

patient compliance; vital signs; telehealth; telemonitoring; clinical trial; chronic disease

Introduction

Telehealth, the delivery of health services at a distance, has been extensively studied in various at-home, primary care, and hospital-based settings for more than 20 years [1-5]. Large health care organizations such as the Veterans Administration in the United States or the National Health Service in the United Kingdom have already adopted a range of telehealth solutions [6]. The aging population and increasing burden of chronic disease together with the availability of low-cost monitoring technology have resulted in increasing interest in deploying telehealth services for the management of patients with chronic conditions and an increasing market demand worldwide for telehealth services.

Different aspects of telehealth solutions, such as clinical, service, and economic benefits, have been investigated in several pilot projects [7,8]. These factors need to be evaluated in order to promote wide-scale implementation and reduce costs. However, patient adherence with scheduled vital signs measurements or their use of technology is an important aspect that has not been thoroughly studied. As patients in home telehealth programs may choose to stop their participation [9] or may not follow the intended procedure, poorer results than expected may be achieved [10]. In order to gain the maximum benefits of at-home telehealth services, adherence rates should be high. Hence, analyzing factors that can influence patient adherence is of great importance.

Relatively few randomized controlled trials on at-home telemonitoring of vital signs have been reported [3,11]. This paper reports findings on patient adherence in the recently completed Commonwealth Scientific and Industrial Research Organisation (CSIRO) trial of home monitoring for chronic disease management, carried out at several locations along the east coast of Australia [12]. The trial was designed to explore a wide set of outcomes resulting from the introduction of a telehealth model of service based on at-home telemonitoring of vital signs and administration of a range of clinical questionnaires to patients with a range of chronic conditions supervised in either hospital-based or community-based settings.

The clinical protocols for the trial [12], the data architecture design [13], decision support and statistical trend analysis of vital signs data [14], and the impact of telemonitoring on health care expenditure, hospital admissions, and length of stay [8] have been previously published.

In hospital intensive care units, it is well recognized that serious adverse events can be prevented [15-20] by recognizing early warning signs of clinical and physiological deterioration and responding appropriately. The authors concluded that early recognition of these events presents an opportunity for decreasing mortality. There is thus a clear analogy between the use of early warning systems in intensive care units and emergency departments and the early warning that is available through the longitudinal monitoring of vital signs at home of chronically ill individuals.

However, there is currently limited knowledge about the influences and determinants of home telehealth adherence in frail elderly people and their carers [21,22]. Their ability to adhere to a strict regime of telemonitoring has been identified as an important influence on the success of telemonitoring programs [23], and there is a perception that patients will not adhere to a monitoring protocol and will often abandon the program [24,25].

A systematic review of the uptake and continued use of telehealth for patients with congestive heart failure (CHF) and chronic obstructive pulmonary disease (COPD) [26] reported that almost one-third of patients who were offered telehealth services refused and one-fifth of patients who did accept subsequently withdrew. Other studies on the adherence and effectiveness of at-home telemonitoring [27] report that over 1 year the adherence rate for recording weight was 75% and for the recording of blood pressure, 90%, while another study [28] reported values of 12% to 75% for adherence with daily weighing for CHF patients. According to Maeder et al [6], reported levels of adherence varied from approximately 40% to 90%. In most of these studies, the adherence rate usually

dropped off steadily over time. This drop is more evident in the first few weeks or months after the start. Higher adherence rates have been reported for simple types of monitoring and for shorter periods of intervention [6]. As an illustration, while adherence rates of 89% were reported by Port et al [29] for the first 2 months, this rate dropped to around 50% within 10 months. The adherence rate in a more complex monitoring system after lung transplantation was about 42% [30].

In this paper we undertake a comprehensive analysis of patient adherence with their measurement schedules in the CSIRO National Telehealth Trial, driven by the hypothesis that longitudinal vital signs data and periodic patient-administered clinical questionnaires provide powerful tools for early identification of an exacerbation of a patient's condition and permit the early mobilization of clinical resources to avoid unnecessary hospitalization.

Methods

Research Ethics Committee Approval

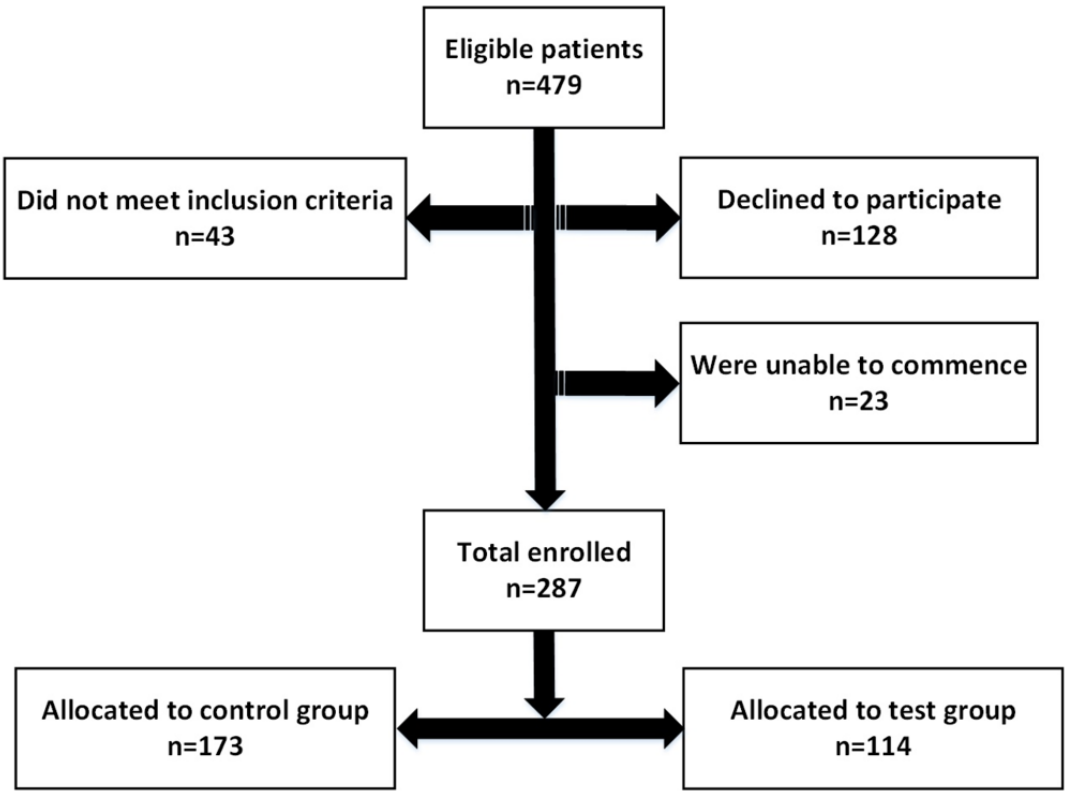
The clinical trial protocol for this study was approved by the CSIRO Human Research Ethics Committee (approval number 13/04, March 25, 2013) as well as 5 other state- and site-based local ethics committees.

Patient Selection

A master register of 1429 eligible patients was formed from hospital lists provided by local health districts and patients known to clinical staff. Local health districts were located in the states of Queensland (QLD), New South Wales (NSW), the Australian Capital Territory (ACT), Victoria (VIC), and Tasmania (TAS). Subjects were eligible to participate in the study if they met inclusion criteria that were comprehensively described in an earlier publication [11] but mentioned briefly here for convenience: age 50 years and older; 2 or more unplanned acute admissions during the last 12 months or 4 or more unplanned acute admissions during the previous 5 years; and a principal diagnosis of COPD, coronary artery disease, hypertensive diseases, CHF, diabetes, or asthma. From the master register, 479 patients were still deemed eligible after individual screening and were contactable. Patients were deemed ineligible if they had been diagnosed with compromised cognitive function, a neuromuscular disease, cancer, or a psychiatric condition. Note that we chose to take a population health approach to patient selection in that all patients who met the inclusion criteria were eligible, irrespective of their chronic condition. As a result, the patient cohort is subsequently treated as a homogenous group in this paper.

As reported in Celler et al [8], of the 479 eligible patients, 128 (26.7%) declined to participate, 41 patients did not meet inclusion criteria at interview, and 23 were unable to commence. Of the remaining 287 subjects who consented, 114 were randomly allocated to the telemonitoring test group and the remaining 173 were allocated to the control group (Figure 1). In addition, it was found that vital sign records for 1 subject were missing. Test patients were supplied with a telemonitoring system and trained on its use on installation, while the control group continued to receive normal care through their primary care physician.

Figure 1. Final cohort of test and control patients.



Organization of Care

A project officer at each test site was responsible for managing operational and research activities for the study, thereby separating patient care from study operations. The PO was also responsible for consenting patients, onsite visits, equipment maintenance, and technical support.

The clinical care coordinators (CCCs) were experienced nursing staff, seconded part-time from each trial site health service provider. In 2 of the 5 trial sites, the CCCs were specialist nurses based in local hospitals and able to draw upon the clinical resources of the hospital. The remaining 3 sites were community sites located within primary care organizations called Medicare Locals. In these sites, the CCCs were community nurses with little access to additional clinical resources other than the patients' own primary care provider.

The CCCs monitored patients' vital signs and clinical questionnaire responses recorded via the telemonitoring unit daily during business hours. In response to exacerbations in their patients' conditions, as evidenced by changes in the patients' vital signs and questionnaire responses, CCCs would initiate and coordinate a timely response to avoid a further exacerbation of their condition and possible hospitalization. Time spent by patients taking their vital signs at home and time spent by CCCs in managing the patients under their care was monitored by comprehensive logs routinely collected by the telemonitoring system and clinician portal. CCCs could also report time spent on a professional portal set up to provide a chat room for CCCs and POs.

Vital Signs Telemonitoring Unit

The TeleMedCare (TMC) Systems Clinical Monitoring Unit (TeleMedCare Pty Ltd), depicted in Figure 2, and associated data hosting and clinical web services were selected for the trial; not all features offered by the device were used in this study. The selection of this telemonitoring system was based on the requirements listed below:

- All vital sign measuring devices are part of the system, minimizing issues that can happen by having separate devices connecting to a central unit
- Entire telemonitoring system together with measuring devices and software are approved by the Australian Therapeutic Goods Administration and the US Food and Drug Administration
- Telemonitoring system has no battery requirements as the unit is mains powered
- No incompatibility and calibration issues given the unit is designed to work together with all its devices
- Must be user friendly and easy to operate
- All these factors including subscription costs fit within the time and budget of the trial

The site POs and CCCs configured the telemonitoring system to reflect clinical best practice for the patient's clinical condition. Patients would be reminded to record some or all of the vital signs shown below, typically in the morning before taking their medications.

- Noninvasive blood pressure (NIBP) using combined oscillometric and auscultatory techniques
- Pulse oximetry to measure arterial blood oxygen saturation
- Single channel ECG, using either the built-in surface electrodes or a custom cable and electrode clamps
- Spirometry, including measurements of vital capacity, peak expiratory flow rate, and volume expired in first second
- Body temperature
- Body weight (SD 100 gm accuracy)
- Glucometer (blood glucose concentration)

In addition to their schedule, patients could take their vital signs at any time. A full suite of clinical questionnaires was also available. These were scheduled and administered by the CCCs and POs.

Monitoring of Vital Signs

Patients were encouraged to record their vital signs daily. CCCs generally viewed every patient's record daily, and time spent on each patient was tracked using the CSIRO Web portal and logs implemented on the clinician portal. On average, CCCs accessed the TMC Clinician Web Portal a little more than once a day and spent on average a little less than 30 minutes per week reviewing each patient's data. There was no significant difference in the time spent by community-based and hospital-based CCCs in the time spent reviewing patient data. CCCs were able to make contact with patients at will to review their progress and discuss the data collected.

Not all patients were required to measure all vital signs. Patients were required to measure vital signs according to their primary disease condition and known comorbidities. Thus 111 patients were monitoring blood pressure, 111 patients were monitoring their peripheral capillary oxygen saturation (SpO_2), 113 patients were recording their ECG, 52 were recording their blood glucometry, 78 were monitoring their respiratory function, 95 were monitoring their body weight, and 103 were recording their body temperature.

Patients' adherence with their scheduled daily measurements was calculated by tracking the total number of scheduled events and then counting the actual number of measurement activities completed. Multiple measurements taken on any 1 day are counted as a single measurement. The ratio of these provides a robust measure of adherence.

Statistical Analysis

Comparisons using the cases available were made between subgroups using the 2-sample *t* test for continuous variables and the Wilcoxon rank-sum test for skewed variables. Baseline characteristics are described using mean and SD for continuous symmetrical variables and mean and 95% CI for skewed data.

Categorical variables are presented as counts and percentages. All statistical tests were 2-tailed, and a *P* value of <.05 was used to indicate statistical significance. Statistical analysis was performed using MATLAB R2016b (MathWorks) and Excel (Microsoft).

Linear Regression

Linear regression was carried out using the *fit* command in the MATLAB statistics toolbox, and 1-way analysis of covariance (ANCOVA) was used to determine whether the slopes of the linear regression lines are different for the average adherence rates associated with blood pressure, ECG, SpO_2, blood glucose, body temperature, and body weight for male and female patients.

Synchronization of Data to Start Date of Telemonitoring

In this study, time is not calendar time but rather time relative to each patient's start date of telemonitoring. This compensates for the fact that individual patients were enrolled and commenced telemonitoring over a period of more than a year (Figure 3) and helps to attenuate seasonal effects.

Figure 2. TeleMedCare clinical monitoring unit.

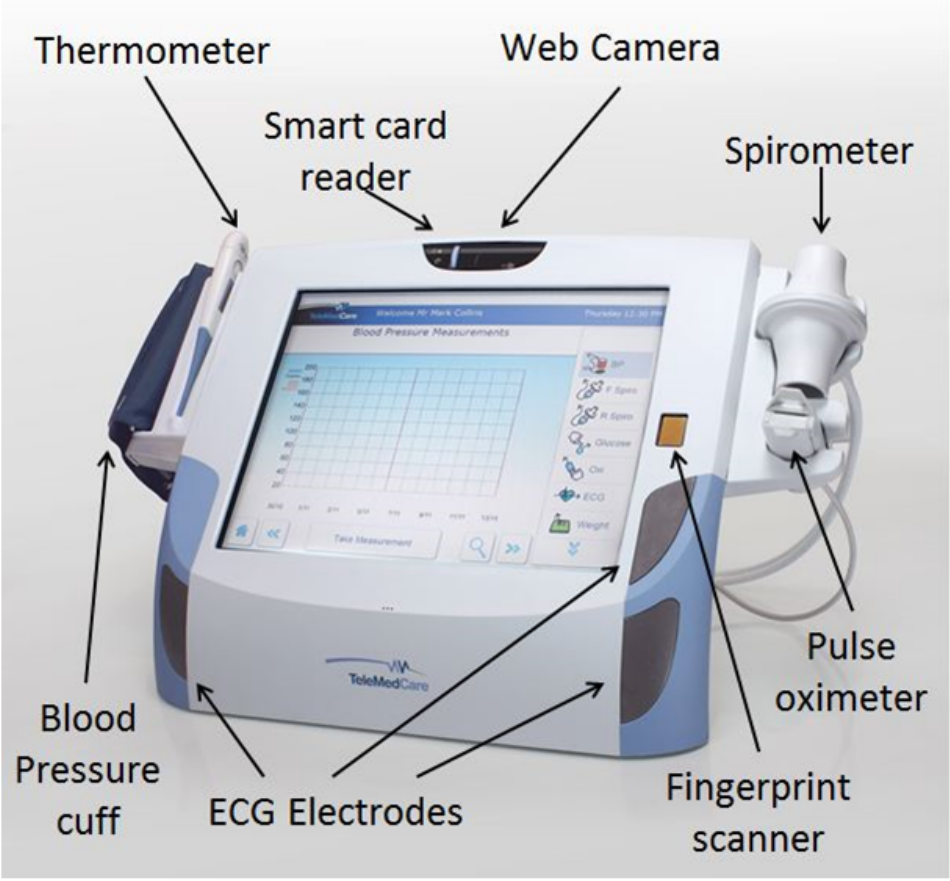
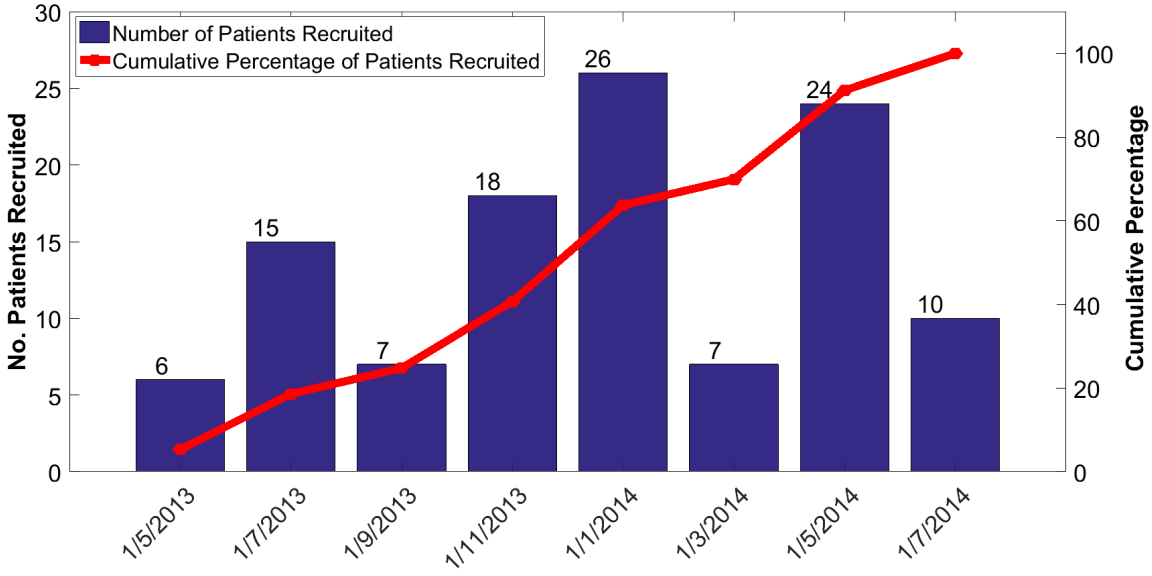


Figure 3. Distribution of commencement dates for monitoring of vital signs.



Results

Demographics of Test Patients

Basic demographics of test patients in the study are shown in Table 1. There were no significant differences in age between test patients in each of the 5 sites and between male and female patients.

A total of 63.7% (72/113) of the test patients were male and 36.3% (41/113) were female. Most test patients had more than 1 condition listed as a primary diagnosis, and for simplicity, primary disease conditions were grouped in the broad categories of cardiovascular disease (N_{Test}=58), respiratory disease (N_{Test}=34), and diabetes (N_{Test}=21). Figure 3 shows the wide distribution of commencement dates for the test patients arising from local operational delays in identifying patients and providing access to the Internet.

Test patients were monitored on average for 302 days, with no significant difference between average monitoring durations for female patients (287 days) and male patients (310 days); 75% of all test patients were monitored for periods exceeding 6 months. The duration of monitoring for patients who did not drop out was primarily determined by the individual start date of the intervention and the end date of the trial.

Patient Adherence With Monitoring Schedules

Overall patient adherence data is shown in Table 2. Test patients successfully completed 120,861 measurements of vital signs over a period of 16 months and on average, patients were recording their vital signs with an adherence rate of 64.78%, a little better than once every 2 days.

Patients under the care of community-based CCCs were overall a little less adherent than patients under the care of hospital-based CCCs (63,553/103,737, 61.26%, versus 56,598/81,821, 69.17%). However, adherence with forced spirometry, body temperature, and body weight measurements was substantially greater for patients under hospital-based supervision than for patients supervised in the community. These differences are further explored in the next section.

Adherence Rates Over Time by Age, Gender, and Mode of Supervision

We now explore differences in adherence rates over time synchronized to the start date of telemonitoring for each patient and between male and female patients and younger and older patients and further analyze the differences observed between hospital-based supervision and community-based supervision. In this analysis, we have only included patients who finished at least 3 months of measurements.

Table 1. Basic demographics of test patients at baseline.

Demographics	Hospital-based		Community-based			Total
	TAS[a]	ACT[b]	VIC[c]	NSW[d]	QLD[e]	
Patients, n	29	16	25	17	26	113
Age, years (SD)	68.8 (9.0)	70.1 (8.2)	68.9 (7.6)	76.7 (9.1)	70.0 (10.8)	70.5 (9.3)
Male patients, n	18	11	18	9	16	72
Age, years (SD)	68.9 (10.0)	69.9 (7.7)	69.5 (7.8)	75.4 (7.4)	69.4 (9.0)	70.1 (8.6)
Female patients, n	11	5	7	8	10	41
Age, years (SD)	68.6 (7.4)	70.6 (10.2)	67.4 (7.4)	78.1 (11.2)	70.9 (13.6)	70.1 (8.6)

[a]TAS: Tasmania.

[b]ACT: Australian Capital Territory.

[c]VIC: Victoria.

[d]NSW: New South Wales.

[e]QLD: Queensland.

Table 2. Patient adherence with measurement.

Vital signs	Scheduled Items, n	Items completed, n	Adherence, %
Blood pressure	31,117	21,890	70.35
Electrocardiogram	33,719	22,405	63.70
Pulse oximetry	31,102	21,363	68.69
Blood glucose	12,579	8501	67.58
Spirometry	20,498	11,493	56.07
Body temperature	29,792	19,158	64.31
Body weight	27,777	16,051	57.79
Total	186,584	120,861	64.78

We note that in the first quarter of monitoring (see Multimedia Appendix 1), the adherence rates of patients monitored in hospital settings relative to those monitored in community settings were significantly higher for spirometry, blood glucose, body weight, and body temperature. Adherence rates for other measurements and differences in adherence across age and gender were not significantly different.

In the second quarter, significantly higher adherence rates for blood glucose, spirometry, body weight, and body temperature continued to be observed for patients supervised in hospital settings relative to those supervised in community settings, and male adherence with NIBP, SpO_2, and ECG began to significantly increase relative to female adherence. Adherence with spirometry measurements for the patient cohort older than 70 years also increased significantly (P=.049) relative to the 70 years and younger patient cohort.

In the third quarter, males continue to be more adherent than females for NIBP and SpO_2 measurements, and hospital-supported patients continued to be more adherent than community-supported patients for spirometry measurements.

In the fourth quarter, those aged over 70 years were more adherent than those aged 70 years and younger for spirometry measurements, and males continue to be more adherent than female patients for NIBP, SpO_2, ECG, and BT. Differences in adherence rates for spirometry between hospital-supported patients (74%, range 64.5% to 83.5%) and community-supported patients (38.6%, 28.6% to 49.1%) become even more pronounced (P=.002).

Linear Regression Analysis of Changes in Adherence Over Time

Linear regression was carried out as described in the Methods section with the results shown in Figures 3-5 and Table 3. When adherence rates initially dropped off in the first 2 months of monitoring, these data were removed and analyzed separately to identify long-term trends. ANCOVA analysis comparing slope of the overall male and female adherence data in Figure 4 confirms that there was a significant (P=.002) difference in the slopes.

A similar analysis was undertaken for subgroups within the total patient cohort to test whether these results were different for those who were monitored within a community environment or within a hospital environment. In Figure 5, a significant difference (P=.02) in slopes for hospital-supervised patients compared with community-supervised patients is observed although the latter appear to increase their adherence over time.

We undertook a similar analysis to test whether these results were different for younger patients (aged 70 years and younger) and older patients (older than 70 years) and for patients with predominantly cardiovascular disease and those with other diseases. Our analysis, however, showed that there were no significant differences between these subgroups.

The spirometry adherence rate (Table 4) was an outlier in this analysis in that no significant difference was seen between male and female adherence rate variation over time. However, as shown in Figure 6, while patients supervised in a hospital setting had far higher adherence rates that indeed increased over time, the adherence rates for community-supported patients started at a much lower level and then decreased over time.

Figure 4. Average adherence rates over time by gender (combined blood pressure, electrocardiogram, peripheral capillary oxygen saturation, blood glucose, body temperature, and body weight). Red lines are linear regression lines (the given equations represent the regression lines).

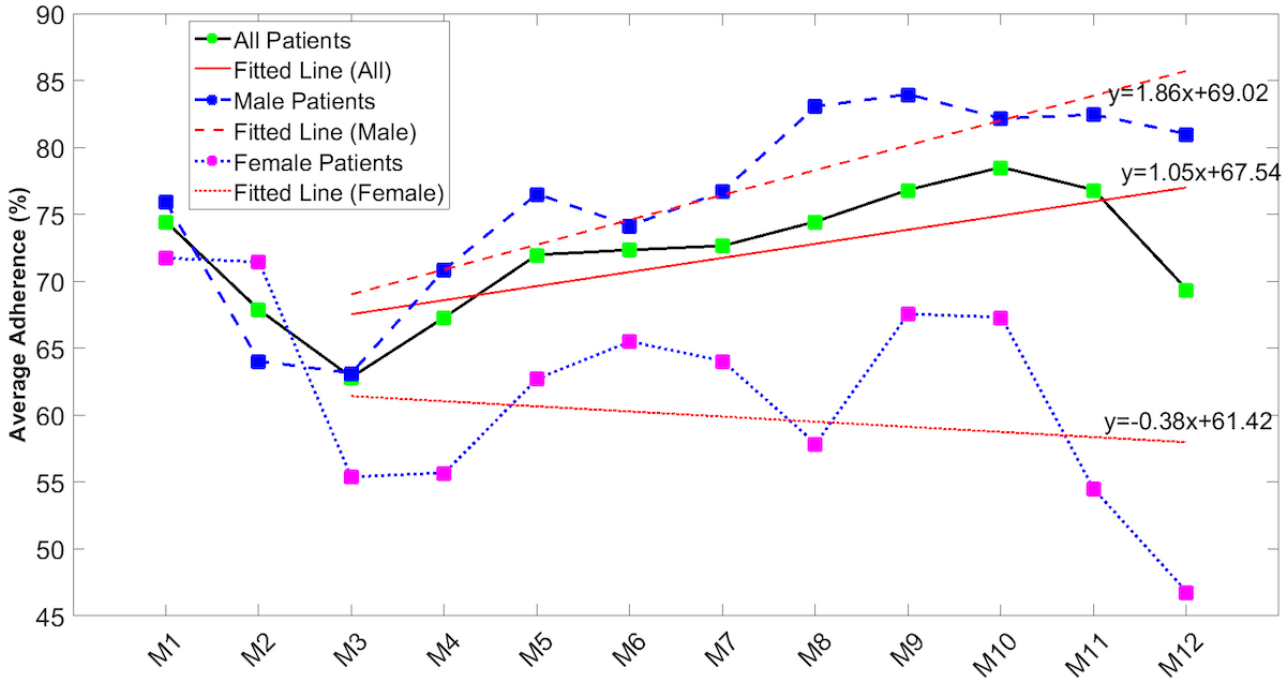


Figure 5. Average adherence rates over time by community-based or hospital-based mode of supervision (blood pressure, electrocardiogram, peripheral capillary oxygen saturation, blood glucose, body temperature, and body weight). Red lines are linear regression lines (the given equations represent the regression equations).

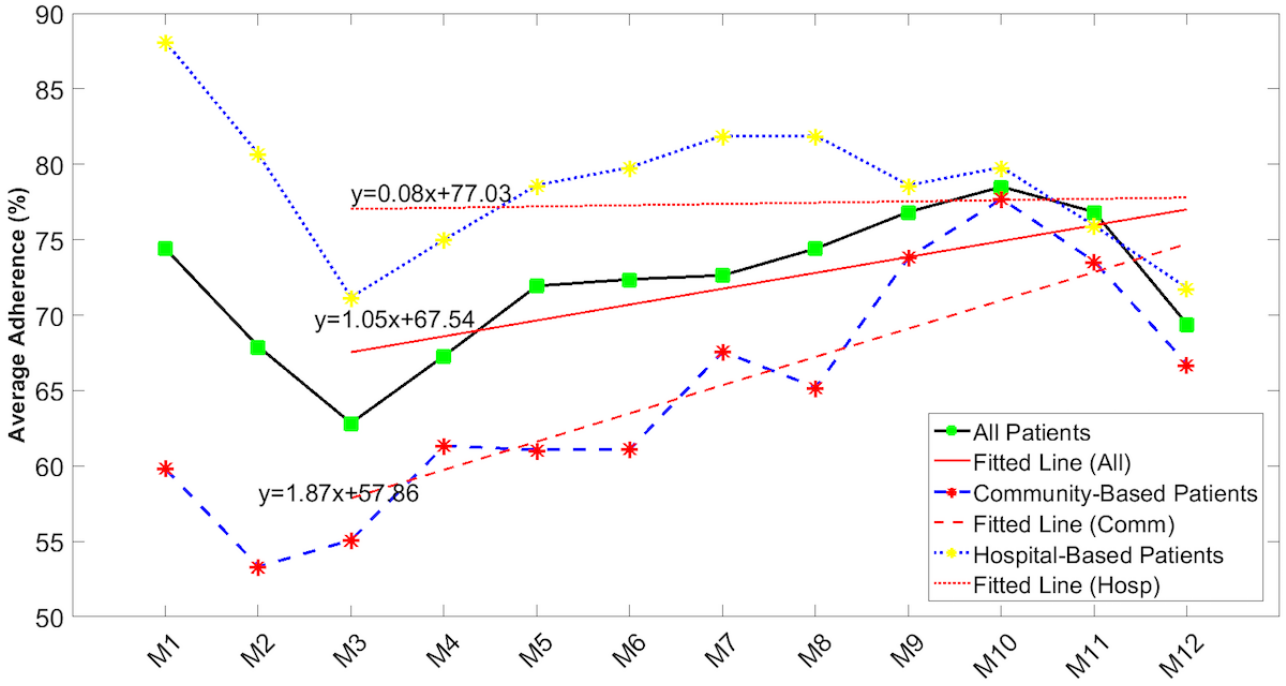


Table 3. Linear regression and analysis of covariance for different subgroups within the total patient cohort.

Patient subgroups	Slope, %/month (95% CI)	Intercept, % (95% CI)	*P* value
Male patients	1.86 (–0.01 to 3.72)	69.02 (59.05 to 78.79)	.02
Female patients	–0.38 (–2.25 to 1.49)	61.42 (51.45 to 71.40)	—[a]
Hospital-based patients	0.08 (–1.32 to 1.49)	77.03 (69.53 to 84.53)	.02
Community-based patients	1.87 (0.47 to 3.28)	57.86 (50.36 to 65.37)	—
Patients 70 years and younger	0.63 (–0.75 to 2.02)	67.90 (60.51 to 70.30)	.25
Patients older than 70 years	1.40 (0.02 to 2.79)	65.21 (57.82 to 72.61)	—
Cardiac patients	1.53 (–0.16 to 3.22)	68.10 (59.08 to 77.12)	.07
Patients with other chronic conditions	–0.01 (–1.70 to 1.68)	66.04 (57.02 to 75.07)	—

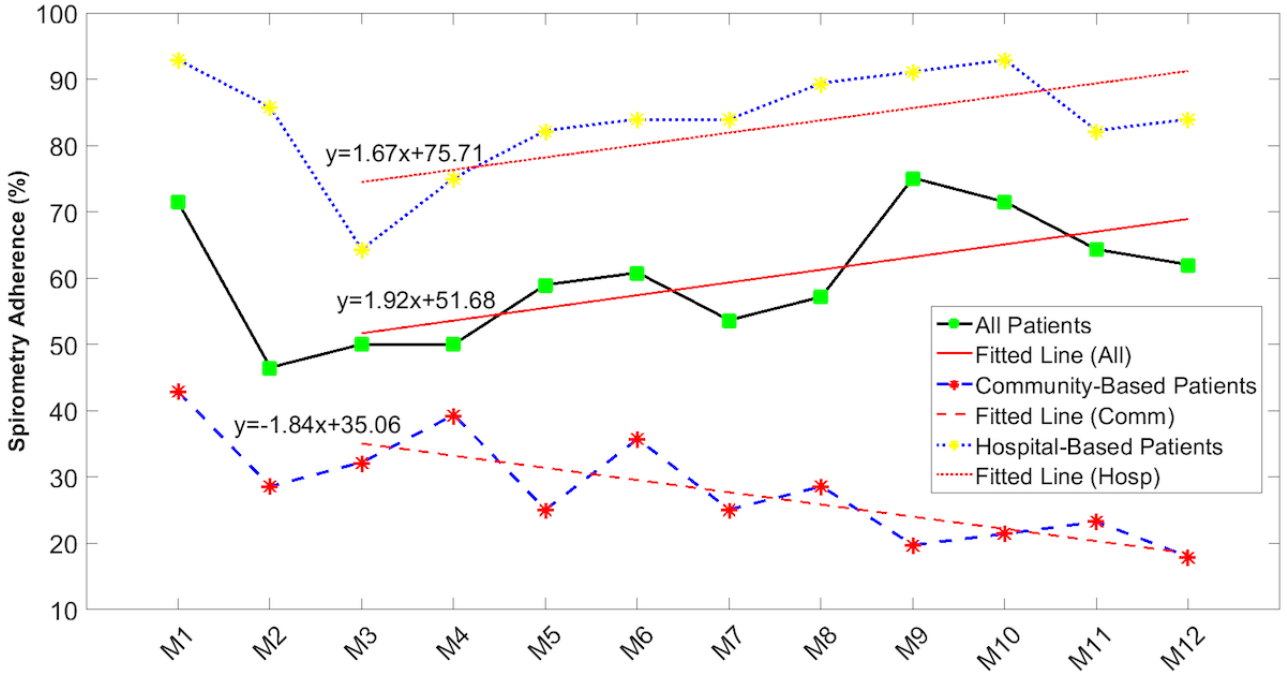
[a]Indicates data is not applicable.

Table 4. Linear regression and analysis of covariance for spirometry adherence.

Patient subgroups	Slope, %/month (95% CI)	Intercept, % (95% CI)	*P* value
Male patients	1.81 (0.05 to 3.57)	55.02 (45.64 to 64.41)	.88
Female patients	1.53 (–2.31 to 5.38)	48.40 (27.86 to 68.93)	—[a]
Hospital-based patients	1.67 (0.19 to 3.14)	75.71 (67.85 to 83.58)	<.001
Community-based patients	–1.84 (–3.00 to –0.68)	35.06 (28.87 to 41.26)	—

[a]Indicates data is not applicable.

Figure 6. Spirometry adherence rates over time by mode of community-based or hospital-based supervision. Red lines are linear regression lines (the given equations represent the regression lines).



Discussion

Principal Findings

We present in this paper some fundamental results on the ability and willingness of chronically ill patients to engage with a comprehensive regime of at-home telemonitoring of their vital signs over extended periods of time, in some cases exceeding a year. In our study there were no significant differences in health literacy, socioeconomic status, or level of previous experience with technology. We recognize that different vital signs are easier and more acceptable than others to carry out, and we anticipated that overall adherence rates would reflect the ease of use and perceived value of the measurement. Overall adherence rates were, however, surprisingly similar, ranging from 56.1% for spirometry to 70.4% for blood pressure measurements. This is consistent with the relative complexity of spirometry measurements, which require 3 forced expiratory efforts and could be quite tiring for frail, chronically ill patients.

Our results straddle levels of adherence reported by Maeder [6] of between 40% to 90% and do show an initial steady drop-off over time, particularly in the first few weeks or months after the start. Unlike Port et al [29], who reported that an initial adherence rate of 89% dropped to around 50% within 10 months, our study shows an increase in adherence following an initial drop. Our study also confirms that higher adherence rates are reported for simple types of monitoring and for shorter periods of intervention [6].

Generally, adherence with measurement of vital signs fell over the first 3 months (Multimedia Appendix 1 and Figures 4 and 5) but recovered over time to levels achieved at the start of monitoring. As a result, there were no significant differences between adherence rates recorded in the first quarter of monitoring and those recorded in the last quarter of monitoring.

Measurement of spirometry by women, however, started from a very low level and continued to fall over time (Figure 6). Age did not appear to be a factor, as there were no significant differences between adherence rates of those aged 70 years and younger and those older than 70 years. Gender, however, was a factor, as male subjects were more adherent than female patients across all measurement modalities. The differences were only significant for the measurement of NIBP, SpO_2, and ECG (Multimedia Appendix 1). These data on gender differences are in good agreement with the literature, which reports reduced adherence by women to telemonitoring of BP [30], antiretroviral therapies [31], management of cystic fibrosis [32], and medications [33].

In 2 of our sites, care coordination was hospital-based and was carried out by highly trained specialist nurses. In the other 3 sites, care coordination was carried out by community nurses often without any additional clinical support. Patients supported by hospital-based care coordinators were significantly more adherent with their measurement of lung function (spirometry), blood glucose, body weight, and body temperature than those supported by community-based care coordinators. Adherence was also higher for patients supported by hospital-based care coordination for NIBP, SpO_2, and ECG, but the differences were not significant.

We cannot fully explain the initial fall of adherence in the first quarter other than to hypothesize that CCCs may be more successful in encouraging patients to continue with their measurements once a longitudinal patient record is achieved and its value noted. The other major conclusion of this study is that care coordination by hospital-based staff leads to higher levels of adherence than for patients managed by community-based staff. This would suggest that hospital-based staff may be more cognizant of the benefits of monitoring vital

signs, more capable of interpreting the longitudinal patient record, and therefore more willing to offer greater encouragement and support to their patients.

According to the systematic review undertaken by Maeder et al [6], the adherence rate usually dropped off steadily over time in most of the studies, and this drop is more evident in the first few weeks or months after the start. Although it is evident from our results that the overall adherence rate dropped off in the first 3 month, it did not continue to fall and indeed recovered over time. We can therefore conclude that chronically ill patients are able to monitor their vital signs at least once every 2 days over prolonged periods of time even for the most difficult measurements such as spirometry and ECG.

Our results also found that the adherence rates in male and community-based subgroups of patients significantly increased over time while the adherence rates in female and hospital-based patients slightly decreased over time. Additionally, the adherence rate variations over time were not significantly different between younger and older subgroups within the total patient cohort or between patients with predominantly cardiovascular disease and those with other diseases.

The telemonitoring system used in this study had some unique features that may have encouraged increased patient adherence with the vital sign measurements. During the recording of a vital sign, the underlying graphical data such as the pressure trace, oscillometric waveforms, and Korotkoff sounds for NIPB measurements were visible to patients and may have promoted greater adherence and recording of better quality traces. Patients were also able to view their own longitudinal records on screen and could respond to increasing and decreasing trends in their data. More research, however, needs to be carried out on the relative importance of user interfaces and the quality of biomedical instrumentation in maintaining the interest and involvement both of patients and the CCCs.

Limitations

Limitations of this study include the reduced numbers of patients who were monitored for the full 12 months, dropping, for example, from 105 patients recording NIBP in the first quarter to 61 in the fourth quarter (as reported in Multimedia Appendix 1). This drop, however, is the result of the fact that patients began their monitoring over a period of more than 6 months and all monitoring ended when the trial ended. As a result, some test patients were monitored for just a little more than 6 months and others for more than 12 months. In addition, the ability of CCCs to support and encourage patients to continue with their telemonitoring of vital signs was not rigorously evaluated and was probably quite variable, although we did confirm that generally patients supported by hospital-based specialist nurses had higher adherence with their measurement schedules.

In a previous paper, we reported that test subjects could be characterized as having the following primary chronic conditions: 50% cardiac disease (mainly CHF), 30% respiratory disease (mainly COPD), and the remaining 20% diabetes [8]. All subjects had some comorbidities. As stated earlier, in choosing a population health approach, we analyze the total patient cohort as a homogeneous group. We recognize, however, that some additional insights could result from analyzing and comparing adherence based on the primary diagnosis.

Conclusions

We have previously reported the following outcomes for this trial: significant drop in expenditure on medical services, reduced admissions to the hospital, and significant reduction in hospital length of stay [8]. Data presented here suggest that levels of adherence to long-term telemonitoring by patients with chronic conditions support the hypothesis that longitudinal recording of vital signs data provides powerful tools for early identification of an exacerbation of a patient's condition and may permit the early mobilization of clinical resources to avoid unnecessary hospitalization [8].

Acknowledgments

This project was funded by the Australian Commonwealth Government National Broadband Network Telehealth Pilot Program with additional financial support from CSIRO and in-kind contributions from health service providers at the 5 trial sites. These health service organizations include Australian Capital Territory (Department of Health), Townsville-MacKay Medicare Local, Launceston Hospital, Nepean Blue Mountains Local Health District, and Djerriwarrh Health Services. Industry and research partners TeleMedCare Pty Ltd, Samsung Australia Pty Ltd, and iiNET Pty Ltd are also acknowledged. Additional support for ongoing data analysis was provided through an Australian Research Council Linkage Grant (LP160101202). All authors made a significant contribution to data analysis and the drafting of the manuscript.

Conflicts of Interest

There were no conflicts of interest during the planning and execution of the project. Six months after its completion, BC, Chief Investigator and Project Director, was appointed to a part-time position at TeleMedCare Pty Ltd as Director of Research.

Multimedia Appendix 1

Adherence rates over time as a function of gender, age, and supervisory setting.

[PDF File (Adobe PDF File), 50KB - medinform_v6i2e15_app1.pdf]

References

1. 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]
2. Bashshur RL, Howell JD, Krupinski EA, Harms KM, Bashshur N, Doarn CR. The empirical foundations of telemedicine interventions in primary care. Telemed J E Health 2016 May;22(5):342-375 [FREE Full text] [doi: 10.1089/tmj.2016.0045] [Medline: 27128779]
3. Bower P, Cartwright M, Hirani SP, Barlow J, Hendy J, Knapp M, et al. A comprehensive evaluation of the impact of telemonitoring in patients with long-term conditions and social care needs: protocol for the whole systems demonstrator cluster randomised trial. BMC Health Serv Res 2011 Aug 05;11:184 [FREE Full text] [doi: 10.1186/1472-6963-11-184] [Medline: 21819569]
4. Brown EM. The Ontario Telemedicine Network: a case report. Telemed J E Health 2013 May;19(5):373-376. [doi: 10.1089/tmj.2012.0299] [Medline: 23301768]
5. Pare G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007 May 01;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
6. Maeder A, Poultney N, Morgan G, Lippiatt R. Patient compliance in home-based self-care telehealth projects. J Telemed Telecare 2015 Dec;21(8):439-442. [doi: 10.1177/1357633X15612382] [Medline: 26556057]
7. Shany T, Hession M, Pryce D, Roberts M, Basilakis J, Redmond S, et al. A small-scale randomised controlled trial of home telemonitoring in patients with severe chronic obstructive pulmonary disease. J Telemed Telecare 2017 Aug;23(7):650-656. [doi: 10.1177/1357633X16659410] [Medline: 27464957]
8. Celler B, Varnfield M, Nepal S, Sparks R, Li J, Jayasena R. Impact of at-home telemonitoring on health services expenditure and hospital admissions in patients with chronic conditions: before and after control intervention analysis. JMIR Med Inform 2017 Sep 08;5(3):e29 [FREE Full text] [doi: 10.2196/medinform.7308] [Medline: 28887294]
9. Merilahti J, Pärkkä J, Antila K, Paavilainen P, Mattila E, Malm E, et al. Compliance and technical feasibility of long-term health monitoring with wearable and ambient technologies. J Telemed Telecare 2009 Aug;15(6):302-309. [doi: 10.1258/jtt.2009.081106] [Medline: 19720768]
10. Botsis T, Hartvigsen G. Current status and future perspectives in telecare for elderly people suffering from chronic diseases. J Telemed Telecare 2008 Jun;14(4):195-203. [doi: 10.1258/jtt.2008.070905] [Medline: 18534954]
11. Celler BG, Sparks R, Nepal S, Alem L, Varnfield M, Li J, et al. Design of a multi-site multi-state clinical trial of home monitoring of chronic disease in the community in Australia. BMC Public Health 2014 Dec 15;14:1270 [FREE Full text] [doi: 10.1186/1471-2458-14-1270] [Medline: 25511206]
12. Nepal S, Jang-Jaccard J, Celler B, Yan B, Alem L. Data architecture for telehealth services research: a case study of home tele-monitoring. 2013 Presented at: Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 9th International Conference Conference; Oct 20, 2013; Austin p. 458-467. [doi: 10.4108/icst.collaboratecom.2013.254220]
13. Sparks R, Celler B, Okugami C, Jayasena R, Varnfield M. Telehealth monitoring of patients in the community. J Intel Syst 2016 Jan 01;25(1):37-53. [doi: 10.1515/jisys-2014-0123]
14. Wilson RM, Harrison BT, Gibberd RW, Hamilton JD. An analysis of the causes of adverse events from the quality in Australian health care study. Med J Aust 1999 May 03;170(9):411-415. [Medline: 10341771]
15. Kause J, Smith G, Prytherch D, Parr M, Flabouris A, Hillman K, Intensive Care Society (UK), Australian and New Zealand Intensive Care Society Clinical Trials Group. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom—the ACADEMIA study. Resuscitation 2004 Sep;62(3):275-282. [doi: 10.1016/j.resuscitation.2004.05.016] [Medline: 15325446]
16. Harrison GA, Jacques TC, Kilborn G, McLaws M. The prevalence of recordings of the signs of critical conditions and emergency responses in hospital wards—the SOCCER study. Resuscitation 2005 May;65(2):149-157. [doi: 10.1016/j.resuscitation.2004.11.017] [Medline: 15866394]
17. Franklin C, Mathew J. Developing strategies to prevent inhospital cardiac arrest: analyzing responses of physicians and nurses in the hours before the event. Crit Care Med 1994 Feb;22(2):244-247. [Medline: 8306682]
18. Schein RM, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. Chest 1990 Dec;98(6):1388-1392. [Medline: 2245680]
19. Resuscitation Council (UK). Prevention of cardiac arrest and decisions about cardiopulmonary resuscitation. URL: https://www.resus.org.uk/resuscitation-guidelines/prevention-of-cardiac-arrest-and-decisions-about-cpr/ [accessed 2017-10-02] [WebCite Cache ID 6tu9NY8hn]
20. Bleyer AJ, Vidya S, Russell GB, Jones CM, Sujata L, Daeihagh P, et al. Longitudinal analysis of one million vital signs in patients in an academic medical center. Resuscitation 2011 Nov;82(11):1387-1392. [doi: 10.1016/j.resuscitation.2011.06.033] [Medline: 21756971]
21. Wade R, Cartwright C, Shaw K. Factors relating to home telehealth acceptance and usage compliance. Risk Manag Healthc Policy 2012;5:25-33 [FREE Full text] [doi: 10.2147/RMHP.S30204] [Medline: 22570580]
22. Broens THF, Huis in't Veld RMHA, Vollenbroek-Hutten MMR, Hermens HJ, van Halteren AT, Nieuwenhuis LJM. Determinants of successful telemedicine implementations: a literature study. J Telemed Telecare 2007;13(6):303-309. [doi: 10.1258/135763307781644951] [Medline: 17785027]

23. Or CKL, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Aug;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
24. Joseph V, West RM, Shickle D, Keen J, Clamp S. Key challenges in the development and implementation of telehealth projects. J Telemed Telecare 2011;17(2):71-77. [doi: 10.1258/jtt.2010.100315] [Medline: 21097563]
25. Gorst SL, Armitage CJ, Brownsell S, Hawley MS. Home telehealth uptake and continued use among heart failure and chronic obstructive pulmonary disease patients: a systematic review. Ann Behav Med 2014 Dec;48(3):323-336 [FREE Full text] [doi: 10.1007/s12160-014-9607-x] [Medline: 24763972]
26. de Lusignan S, Wells S, Johnson P, Meredith K, Leatham E. Compliance and effectiveness of 1 year's home telemonitoring: the report of a pilot study of patients with chronic heart failure. Eur J Heart Fail 2001 Dec;3(6):723-730. [Medline: 11738225]
27. van der Wal MHL, Jaarsma T, van Veldhuisen DJ. Non-compliance in patients with heart failure; how can we manage it? Eur J Heart Fail 2005 Jan;7(1):5-17 [FREE Full text] [doi: 10.1016/j.ejheart.2004.04.007] [Medline: 15642526]
28. Port K, Palm K, Viigimaa M. Daily usage and efficiency of remote home monitoring in hypertensive patients over a one-year period. J Telemed Telecare 2005 Dec;11 Suppl 1:34-36. [doi: 10.1258/1357633054461705] [Medline: 16035987]
29. Kugler C, Gottlieb J, Dierich M, Haverich A, Strueber M, Welte T, et al. Significance of patient self-monitoring for long-term outcomes after lung transplantation. Clin Transplant 2010;24(5):709-716. [doi: 10.1111/j.1399-0012.2009.01197.x] [Medline: 20047613]
30. Kerby TJ, Asche SE, Maciosek MV, O'Connor PJ, Sperl-Hillen JM, Margolis KL. Adherence to blood pressure telemonitoring in a cluster-randomized clinical trial. J Clin Hypertens (Greenwich) 2012 Oct;14(10):668-674. [doi: 10.1111/j.1751-7176.2012.00685.x] [Medline: 23031143]
31. Berg K, Demas P, Howard A, Schoenbaum E, Gourevitch M, Arnsten J. Gender differences in factors associated with adherence to antiretroviral therapy. J Gen Intern Med 2004 Nov;19(11):1111-1117 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30445.x] [Medline: 15566440]
32. Patterson J, Wall M, Berge J, Milla C. Gender differences in treatment adherence among youth with cystic fibrosis: development of a new questionnaire. J Cyst Fibros 2008 Mar;7(2):154-164 [FREE Full text] [doi: 10.1016/j.jcf.2007.07.008] [Medline: 17719857]
33. Manteuffel M, Williams S, Chen W, Verbrugge R, Pittman D, Steinkellner A. Influence of patient sex and gender on medication use, adherence, and prescribing alignment with guidelines. J Womens Health (Larchmt) 2014 Feb;23(2):112-119. [doi: 10.1089/jwh.2012.3972] [Medline: 24206025]

Abbreviations

ACT: Australian Capital Territory
ANCOVA: analysis of covariance
CCC: clinical care coordinator
CHF: congestive heart failure
COPD: chronic obstructive pulmonary disease
CSIRO: Commonwealth Scientific and Industrial Research Organisation
ECG: electrocardiogram
NIBP: noninvasive blood pressure
NSW: New South Wales
QLD: Queensland
SpO_2: peripheral capillary oxygen saturation
TAS: Tasmania
TMC: TeleMedCare
VIC: Victoria

Edited by G Eysenbach; submitted 16.10.17; peer-reviewed by E Krupinski, K Blondon, T Irizarry; comments to author 17.11.17; revised version received 03.12.17; accepted 15.02.18; published 09.04.18.

Please cite as:
Celler B, Argha A, Varnfield M, Jayasena R
Patient Adherence to Scheduled Vital Sign Measurements During Home Telemonitoring: Analysis of the Intervention Arm in a Before and After Trial
JMIR Med Inform 2018;6(2):e15
URL: http://medinform.jmir.org/2018/2/e15/
doi:10.2196/medinform.9200
PMID:29631991

Original Paper

Perspectives of Nurses Toward Telehealth Efficacy and Quality of Health Care: Pilot Study

Ayisha Bashir[1], MS, MBBS; Dhundy R Bastola[2], PhD

[1]Department of Clinical and Translational Science, Creighton University, Omaha, NE, United States

[2]School of Interdisciplinary Informatics, Information Science and Technology, University of Nebraska, Omaha, NE, United States

Corresponding Author:
Ayisha Bashir, MS, MBBS
Department of Clinical and Translational Science
Creighton University
2500 California Plaza
Omaha, NE, 68178
United States
Phone: 1 4027407394
Email: abashir@unomaha.edu

Abstract

Background: Telehealth nursing, or the delivery, management, and coordination of nursing care services provided via telecommunications technology, is one of the methods of delivering health care to patients in the United States. It is important to assess the service quality of the involved health professionals as well as the telehealth nursing process. The focus of this study is the innovative model of telehealth care delivery by nurses for managing patients with chronic disease while they are living in their own residence.

Objective: The primary objective of this pilot study was to examine whether telehealth technology impacts the perceived level of internal service quality delivered by nurses within a telehealth organization. To address this research goal, the notion of telehealth nursing service quality (TNSQ) is empirically tested and validated with a survey instrument.

Methods: Data were collected from nurses belonging to a home care agency based on interview questions inquiring about facilitators and inhibitors to TNSQ. A survey to measure TNSQ based on the SERVQUAL instrument was completed by adjusting descriptions of the original instrument to suit the context. Follow-up interviews were conducted to validate questions on the revised instrument.

Results: The findings of this survey research were positive, based on mean differences between expectations and perceptions of TNSQ. This indicates satisfaction with TNSQ and shows that the quality of the service is higher than what the respondents expect. The Wilcoxon signed-rank test using the *P* value for the test, which is .35, did not show a statistically significant change between the median differences of perception and expectation. The total number of respondents was 13. Results indicate that overall perceived service quality is a positive value (0.05332). This means the perceptions of the level of service are slightly higher than what they expect, indicating there is satisfaction with TNSQ.

Conclusions: The responses to the interview questions and data gathered from the survey showed overall satisfaction with TNSQ. The SERVQUAL instrument was a good framework to assess TNSQ. In a nutshell, the study highlighted how the telehealth process provides daily monitoring of patient health, leading to the benefits of immediate feedback for patients, family, and caregivers as well as convenience of scheduling.

(JMIR Med Inform 2018;6(2):e35) doi:10.2196/medinform.9080

KEYWORDS

telehealth; survey; telemedicine; telenursing

Introduction

Background

Telehealth encompasses all facets of remote health care including clinical services provided using telemedicine, as well as interactions with automated services, systems, or information resources. It is the integrated use of electronic information and telecommunications technology to support remote clinical health care, patient and professional health-related education, public health, and health administration [1]. The terms telehealth and telemedicine have become intertwined in literature; in this study, we use telehealth as a broad term that encompasses clinical and nonclinical services in contrast to telemedicine, which consists of purely clinical services [2]. Given the success of telehealth in specialty health care services, the use of technologies to transmit health information and provide care across a distance is poised to enter mainstream health service delivery [3,4]. Although many studies have assessed the cost-effectiveness and patient acceptance of this new method of service delivery, the perspective of nurses as the service provider has not been previously studied [5,6]. Thus, the aim of this study is to investigate telehealth nursing service quality (TNSQ) from the nurses' perspective. To address this research goal, we developed and empirically tested the notion of TNSQ.

Role of Nurses in Telehealth

One of the evolving roles of nurses as they take on new responsibilities of providing nursing service in home settings is the use of information technology (IT). Telehealth nursing focuses on patients' long-term wellness, self-management, and health [1]. According to the American Telemedicine Association, this IT solution provides nursing care across a distance, empowering the care providers with the ability to monitor, educate, follow-up, collect data, and provide multidisciplinary care including remote interventions, pain management, and family support in an innovative fashion [7]. It has been reported that agencies using telehealth have an average patient-to-nurse ratio of 15:1, as compared with non-telehealth agencies having a ratio of 11:1 [7]. Therefore, telehealth nursing can make a tremendous difference in providing patient care, particularly in rural or underserved areas in states such as Nebraska, where there is generally a shortage of nurses and health care services, as well as resources can be limited [3]. In addition, in rural areas, many patients do not receive timely health care interventions because of the lack of specialist services [2]. Home health agencies with telehealth capability caring for patient populations with chronic diseases can take care of patients in their home setting and therefore fill this gap. This provides convenience and a sense of security to the patient, allowing timely nursing interventions under supervised physician care [2,7].

Importance of the Study

This study is important and significant for several reasons. First, nurses' perspectives toward telehealth service quality (SQ) is researched through a case study and validated with a survey instrument. Second, a positive role of telehealth nursing and related technology toward patient care is highlighted. Third, it establishes how telehealth provides daily monitoring of patient health, which has benefits of immediate feedback for patients, family, and caregivers as well as convenience of scheduling. Quality of service is essential for the telehealth performance and viability of the service; therefore, it is important to assess the SQ of involved health professionals as well as the telehealth nursing process [6,8].

Service Quality

Among different characteristics of service, the quality of service in the health sector shapes the experience of care beyond technical competence. In the conceptual model of IT-SQ, it has been described as a comparison between levels of perceived services and consumer expectations [9], where the author suggests that SQ has a functional component concerned with *how the service is delivered* and a technical component concerned with *what is delivered*. This concept is implemented [10] by suggesting that SQ can be measured as the "gap" between the customers' expectations of service and their perceptions related to the quality of service actually being delivered. In other words, SQ is a comparison of expectations (E) with perceptions (P) and can be represented as SQ=P−E [11]. This concept was further described and operationalized [9], leading to the development of the SERVQUAL instrument for SQ [10,11].

Telehealth Nursing Service Quality

The conception of SQ has more to do with the "manner" in which employees deliver services to their customers, rather than the outcome of the service [11]. In a similar vein, TNSQ is the synthesis of the perceptions of the telehealth nurses from the functional and interactive aspects of telehealth interventions. In addition, the SQ items consisting of tangibles (TA), empathy (EM), assurance (AS), reliability (RL), and responsiveness (RS) are the dimensions measured by the SERVQUAL instrument [10]. Furthermore, these are applicable to the measurement of TNSQ [8,12]. In this study, the SERVQUAL instrument is modified and utilized to measure TNSQ [10] (Table 1). The SERVQUAL dimensions developed [10] and revised for this TNSQ study are described in Table 1.

Traditionally, the quality of nursing care was perceived to relate to the degree to which patients' physical, psychosocial, and extra care needs were met [13]. The consequences of quality care were interpreted as *therapeutic effectiveness* where the therapy provided by nurses was perceived to positively affect patients' healing. In other words, it referred to "how" the core service is delivered [14]. In health care services, expression of compassion and empathy by doctors or nurses is an example of functional quality, which also applies to telehealth nursing [13]. Telehealth nursing aims to propagate and advance the process of nursing quality care by enhancing the attributes of empathy, assurance, attention to detail, interpersonal and communication skills, responsiveness, and reliability, while adding convenience to the process [14,15]. Thus, telehealth nursing quality is described as a multidimensional construct that encompasses attributes such as empathy, assurance, and reliability.

Table 1. Adapting general service quality (SERVQUAL) dimensions to telehealth nursing service quality (TNSQ).

Serial number and dimension		General definition	TNSQ definition
1	Tangibles	The physical evidence of service	Physical facilities, equipment, and appearance
2	Reliability	The consistency of performance and dependability	The consistency of performance and dependability of telehealth nursing services
3	Responsiveness	The willingness and readiness of employees to provide service	The willingness and readiness of nurses to provide service
4	Assurance	The confidence communicated by the service provider	The confidence communicated by the telehealth nurse or provider
5	Empathy	The service provider's efforts to understand the customer's needs and then to individualize the service delivery	The nurse's effort to understand the customer's needs and then to individualize the service delivery

This study follows a research approach in which perspectives of telehealth nurses were elicited through a case study and administration of a survey (Multimedia Appendix 1). Expected results include a better understanding of nurses' perspectives of telehealth interventions and quality of the health care services provided. Here, telehealth nursing quality is defined from the nurses' perspectives as a *comparison of expectations (E) about telehealth nursing services with perceptions (P)* using the formula $TNSQ=P-E$. TNSQ attempts to capture nurses' expectations and perceptions of telehealth nursing services as they impact patients.

Service Quality and the SERVQUAL Instrument

As discussed previously, TNSQ is assessed using a modified version of the SERVQUAL instrument. The SQ encompasses 2 related components, namely, technical and functional quality [11]. The technical (core) component is defined as the "what" performed during the service. This concept is implemented [10] by suggesting that SQ has a functional component concerned with "how the service is delivered" and can be measured as the "gap" between the respondent's expectations of service and their perceptions related to the quality of service being delivered [10,11]. To understand the gap between how the technological service is delivered and "what is delivered," it is important to measure the SQ of different advanced information and communication technologies in various industries [16].

Previously, these SERVQUAL items have been used as a reliable and valid measure of SQ because of the dimensionality, item compositions, validity, and disconfirmation paradigm used in its measurement. The "type of technology" has a direct effect on barrier reduction, attitudes, and quality and quantity of communication [16]. There is a link between potential SQ facets and health care organizations' adoption of health care information technologies and telehealth [17,18]. SERVQUAL and its improved versions are still widely used for measuring SQ in various sectors including health care, justifying SERVQUAL as an appropriate tool for TNSQ [6,12]. In the last 2 decades, the SERVQUAL instrument has been refined by the creators [10] to the useful acronym RATER: reliability, assurance, tangibles, empathy, and responsiveness [10,12]. The rationale behind the development of the general instrument is that although each service is unique in some aspect, there are other aspects that are applicable to all services in general [12,13]. This is also the case for evaluating SQ in telehealth. For example, in Greece, SERVQUAL was used to assess the quality services provided by the hospital, after the hospital was transferred to a new and modern location with up-to-date technology including telehealth services [12]. The study showed that the transfer to the new premises and new telehealth equipment substantially improved all of the 5 dimensions; however, the most improvement was seen in the "tangibles" dimension [12]. Other studies have been conducted in the health care industry using SERVQUAL instrument. The application of SERVQUAL instrument was beneficial in a study conducted to optimize the telemedicine model in Korea, ultimately improving the SQ of the telemedicine service providers by providing a checklist of the critical-to-quality telehealth processes, requirements, and expectations of the involved parties [6].

SQ is generally conceptualized in terms of 5 dimensions: tangibles, reliability, responsiveness, assurance, and empathy [10] (Table 1). One of the benefits of applying SERVQUAL for health care to telehealth nursing is that the 5 dimensions can be assessed overall as well as individually and the perspectives of nurses can be assessed thoroughly [6]. In a nutshell, through this research, nurses' perceptions about their ability to deliver telehealth services to patients are collected.

Telehealth Nursing Service Quality and the SERVQUAL Instrument

SERVQUAL remains a valuable tool to assess quality of health care services and assess the perceptions of the patients and health care providers related to new equipment and new hospital facilities [6,12]. The gap analysis method based on SERVQUAL has been used to calculate patient's perceptions related to hospital SQ [19,20]. Published research is lacking with the instrument being used to measure nurses' perspectives toward telehealth quality. This is what makes this pilot study unique, as past studies have been conducted to determine the quality attributes for telemedicine services from the perspective of the patients and physicians [18,19]. There is a need to further investigate the quality attributes for telemedicine services from the nurses' perspectives, which have not been clearly defined before [5,8]. The TNSQ survey contains questions belonging to the 5 dimensions [10,12] (Multimedia Appendix 1). The gap method of subtracting nurses' expectations from perceptions in each of the dimensions (using the survey) helps in assessing the TNSQ [10]. The dimensions having the negative gap score indicate that there is a barrier or an inhibitor present in that dimension; in the same manner, a positive value indicates that

there is a facilitator in that dimension. This study addresses this gap in our understanding of telehealth care by developing concepts and empirically assessing TNSQ. In addition, the SQ items consisting of empathy, assurance, reliability, and responsiveness are dimensions measured by the SERVQUAL instrument [10]. These are reckoned to be applicable in measuring the TNSQ [6,10]. Further evaluation of the responses helps in indicating the positive factors, barriers, and facilitators of TNSQ and ties the responses to the interview question related to facilitators of telehealth barriers or inhibitors to nursing quality care [6,12].

Methods

Case Study Site

The Visiting Nurse Association (VNA) of Omaha was selected as the case study site because it employs telehealth interventions for patient care and because of its geographical proximity; in addition, the VNA telehealth nursing staff was willing to participate in the study. The organization's website provides information regarding the telehealth home care process. After VNA agreed to participate in the study, preliminary interviews were conducted to understand its telehealth environment. To determine the extent of telehealth service provided by VNA and the characteristics of the participants receiving telehealth interventions (age of the patients and types of diseases being addressed), the deidentified patient data (no name or personal identifying values) were obtained from VNA. This consisted of the list of patients who were under VNA care or recently discharged in February 2016. These were all home care patients who were using telehealth services in their own residence. The median age of 205 patients was 74 years, including a total of 128 females and 77 males. The monitoring period of the patients is calculated from the day the patient starts receiving VNA telehealth services and ends on the date they are discharged from VNA. The median monitoring period of a patient was 25 days. The VNA data showed majority of these patients were suffering from chronic disease, were aged above 70 years, and were residing in a home care setting (for more than 3 weeks). This facilitated the research to determine the role of telehealth technology in helping elderly patients living with chronic disease in their home setting.

The methodology of conducting the TNSQ consisted of the following steps: (1) survey research, (2) preliminary interviews to understand the telehealth environment, (3) use SERVQUAL to measure TNSQ—adapted from the original instrument by adjusting the descriptions to suit the context, and (4) follow-up interviews to validate questions on the revised instrument.

Survey Research for Telehealth Nursing Service Quality

The interview at VNA set the baseline for collection of quantitative analysis data. A questionnaire was designed based on SERVQUAL instrument and its 5 dimensions, namely, tangibles, reliability, responsiveness, assurance, and empathy [10]. Looking toward the domains of TNSQ, and bringing together the responses to the interview questions, the nurses had more concerns related to the equipment and its functionality (which belongs to the tangibles dimension). Further evaluation was added to the tangibles dimension in this study, relating to telehealth technology. These additions resulted in the increase in the number of questions from 21 to 25 (Multimedia Appendix 1). In addition, 3 open-ended questions were designed to gain insight toward the positive and negative perspectives of the nurses related to telehealth interventions. Therefore, the original instrument was revised by adjusting the descriptions to address the context.

Measurement

As described above, the SERVQUAL instrument was used to measure TNSQ. The VNA approved the survey. A digital version of the survey was created using Survey Monkey, a Web based software tool that was suitable for our project. Survey Monkey is an online survey development cloud-based software as a service company, with headquarters in San Mateo California. The VNA's chief executive nursing officer emailed the weblink to all participating nurses, which included all 13 nurses employed by the organization. Participants were queried about their expectations regarding quality of service with the incorporation of telehealth technology. Next, they were asked to rate their perceptions of the actual performance (or delivery) of service in the context of telehealth technology. Perceived levels of SQ with the use of telehealth technology were assessed based on the previous definition of SQ, which is obtained by the linear subtraction of service expectations (E) from perceptions (P) for each scale item, using a 7-point Likert scale, according to the SERVQUAL instrument [10]. The complete SQ instrument, which consists of a 25-item scale and accompanying instructions, is included in Multimedia Appendix 1. In this appendix, the first section lists the questions used to measure the nurses' expectations of service in the context of telehealth SQ; the second section lists the items used in measuring their perceptions of the actual performance of service; and the third section includes the demographic questions and 3 open-ended questions. The modified SERVQUAL instrument consisted of 25 items, which included the 5 original dimensions. The first 7 questions belonged to the tangibles dimension (1-7), questions 8 to 12 belonged to the reliability dimension (8-12), questions 13 to 16 belonged to the responsiveness dimension (13-16), questions 17 to 20 belonged to the assurance dimension (17-20), and the last 5 questions belonged to the empathy dimension (21-25).

General Perceptions of Visiting Nurse Association of Nebraska Telehealth Staff

The primary objective of this study was to examine whether telehealth technology impacts the perceived level of internal SQ delivered by nurses within a telehealth organization. Therefore, an in-depth interview was conducted with VNA staff to elicit and analyze the perspective of the telehealth nursing team. VNA services include companion care, infusion pharmacy, home care, home health technology, hospice, and palliative care. VNA telehealth nursing services include monitoring the patient through assessment and collecting data, heart rate, blood pressure, weight, oxygen saturation, and temperature.

Figure 1. Telehealth system.

For the interview, the VNA telehealth team consisted of 3 members: the central station clinician, the clinical manager, and the chief operating officer. The proceedings of the interview were recorded with permission from all participants. The research questions were investigated during the interview, concentrating on the positive and negative aspects of telehealth. This team of nurses highlighted the process of telehealth patient care and daily monitoring. Every patient admitted to VNA services receives a disease-specific home telehealth kit and an educational session on using the equipment (Figure 1). Other members of the household or caregivers may monitor the health care process. The patients are triaged according to their vitals. The central station clinician is responsible for the initial interpretation of the data and contacts the patient with any health care concerns, including blood pressure changes, weight gain, or oxygen level fluctuations. They work closely with physicians, case managers, and other health care providers. The admitting nurse sets a plan with the patients, so they understand the process. One VNA case manager nurse monitors the patient, although different staff may monitor on weekends. Many patients are stable patients; however, others have urgent needs. The service is available 24 hours a day, 7 days a week. The current volume of patient load is manageable, and the maximum is 220 patients.

The below-mentioned interview questions addressed the factors that the nursing staff considers vital toward improving quality of patient care and whether telehealth played a role in providing this to patients. The staff shared perspectives on telehealth contributing and enhancing quality of care as well as barriers and inhibitors to providing quality nursing care via telehealth as described below:

1. What are the positive facilitators of telehealth contributing or enhancing the quality of care?
2. What are the telehealth barriers or inhibitors to nursing quality care?

The telehealth facilitators contributing and enhancing the quality of care mentioned by the VNA during the interview included quality of life and improved communication with health care providers, patients, and caregivers. Some of the telehealth nursing staff statements recorded during the interview include the following:

> *Telehealth interventions improve patient day to day living and enhanced the quality of life. The telehealth system is positive for the VNA as it emphasizes patient safety. It provides independence as well as a sense of security to the patient and caregivers, due to daily monitoring.*
>
> *Telehealth improved communication with health care providers, patients and caregivers. This allowed clinicians to customize patient care quickly and effectively. As a result, the cardiologist and pulmonologist trust the VNA and this trust are gained over time due to their good work ethics and effective delivery of clinical services to the patients.*
>
> *Telehealth interventions are helpful to patients, caregivers and for physician feedback, which leads to everyone's ease of mind. During discussion with staff it was noted the organization's desire to add electronic health records (EHR) and videoconferencing capability to their services. Their goal for adding videoconferencing is to improve timely communication by adding physician, patient, family and telehealth staff on board at the same time for consultation.* [Telehealth nursing staff]

Results

Data Analysis

This study discovered qualitative and quantitative results. A weblink for the survey was provided via email to all the VNA participant nurses. A total of 13 nurses responded to the survey. The survey was generated using a Web-based software tool, Survey Monkey. The data were downloaded from Survey Monkey and processed using SPSS version 22. Descriptive statistics of expectations, perceptions, and gap scores were calculated; the total number of respondents was 13,

nonparametric and not normally distributed. To compare paired means for continuous data that are not normally distributed, the nonparametric Wilcoxon signed-rank test was performed. The result from these data is explained in the quantitative analysis and gap score analysis section (Tables 2-4).

Qualitative Data Analysis

The VNA telehealth staff responses to the questions asked during the exploratory interview showed how the telehealth methods were helping the nurses to achieve good quality patient care, at the same time, providing peace of mind to the clients. These results were possible because of early interventions and monitoring. The telehealth team considered improved quality of life and peace of mind as the most important factors toward providing good quality patient care. The positive facilitators of telehealth, contributing or enhancing quality of care mentioned by VNA, were related to patient safety and quality of life and improved communication with health care providers, patients, and caregivers. In addition, TNSQ is achieving highest standards through the incorporation of technology and teaching patient's personal health habits. Some of the barriers mentioned in the interview consisted of incompatibility and inconsistency of equipment in some cases. For example, the inability to add diabetic patients to their service was an issue at this time because of glucometer incompatibility in some patient's equipment. The response to the open-ended questions in the survey matched the response to the interview questions related to facilitators and inhibitors, as shown by the dimension mapping (Multimedia Appendix 2).

Quantitative Data Analysis

The survey instrument was approved by the University of Nebraska Medical Center (for the University of Nebraska at Omaha) institutional review board (IRB) for protection of human subjects, as it involved survey and interview of nursing professionals. It was qualified as exempt (IRB # 472-16-EX). The data collected from online survey (quantitative data), focus group case study, and interviews (qualitative data) were used to find existing perception of nurses regarding TNSQ. After approval, the surveys were administered through an online software tool to VNA participants as the VNA office was the case study site. The original SERVQUAL instrument was modified by adding technological aspect of the telehealth interventions to the tangible section of the questions. First, the respondents were asked the expectation of SQ because of the incorporation of telehealth technology. Next, the respondents were asked to provide perceptions of the actual performance (or delivery) of service because of telehealth technology. Then, the perceived levels of SQ because of the incorporation of technology were assessed by the linear subtraction of service expectation (E) from perception (P) for each scale item.

Gap Scores Analysis

The distribution scale values obtained for the average telehealth service expectation, service performance, and SQ scores are presented in Tables 2-4. The gap scores are negative in the tangibles (1-7), responsiveness (13-16), and assurance dimensions (17-20). These values demonstrate that the expectations (E) of telehealth nursing service are higher than the actual perception (P) of the service (Table 3,Multimedia Appendix 1). The gap scores are positive in the reliability (8-12) and empathy (21-25) dimensions, which means that the performance in reliability is more than expectations. This has a positive effect on the level of the nursing services within the telehealth provider system based on the incorporation of telehealth nursing services. The values of the items scored on a scale of 1 to 7 can vary from a possible minimum of 25 (25×1) to a maximum of 175 (25×7). Therefore, the more positive the value, the higher the perceived telehealth service item (Table 2).

Table 2. Descriptive statistics.

Dimension	Minimum	Maximum	Mean (SD)	Skewness (SE)	Kurtosis (SE)
Tangibles	−0.90	0.40	−0.50 (0.50)	1.80 (0.845)	3.19 (1.74)
Reliability	0	0.80	0.395 (0.27)	0.275 (0.845)	2.16 (1.74)
Responsiveness	−0.70	0	−0.12 (0.26)	−1.33 (0.845)	2.16 (1.74)
Assurance	−0.90		−0.11 (0.36)	−2.21 (0.845)	4.92 (1.74)
Empathy	0	0.90	0.39 (0.36)	0.99 (0.845)	−0.31 (1.74)

Table 3. Descriptive statistics.

Statement	Number of statements (N)	Minimum	Maximum	Sum	Mean (SD)
Expectation	25	5.40	7.00	162.50	6.58 (0.415)
Perception	25	5.60	7.00	164.10	6.48 (0.43)
Valid N (listwise)	25	–	–	–	–

Table 4. Hypothesis test summary of Wilcoxon signed-rank test.

Serial number	Null hypothesis	Test	*P* value	Decision
1	The median of differences between expectation (E) and perception (P) equals 0	Related-samples Wilcoxon signed-rank test	.35	Retain the null hypothesis

The scale items and sum of average showing higher positive values indicate higher perceived telehealth service items (Table 3). These values show that the perception of performance of telehealth nursing service is overall more than the expected level of SQ. The research generated 13 pairs of ranked data, which are nonparametric, not normally distributed. To compare paired means for continuous data that are not normally distributed, and to compare paired means for ranked data, the nonparametric Wilcoxon signed-rank test is selected [21]. The Wilcoxon signed-rank test using the *P* value for the test, which is .353 (Table 4), did not show a statistically significant change between the median differences of perceptions (P) and expectations (E). Results indicate that the overall perceived SQ is a positive value (0.05332), which means the perceptions of the level of service is slightly higher than what they expect, indicating there is satisfaction with TNSQ.

The responses to the interview and data gathered from the survey showed overall satisfaction with TNSQ. Some issues of equipment malfunction and inconsistency in the telehealth equipment were pointed out by the respondents of the survey. The results of this case study are helpful in reinforcing the positive role of telehealth in impacting patient care and providing alternative solutions to the management of chronic disease. The VNA experience of implementing telehealth methods in the case of chronic patients suffering from chronic obstructive pulmonary disease (COPD) and heart disease is leading to better quality of life, as well as offering preventive services at a lower cost. Elderly patients suffering from chronic diseases, especially COPD and heart disease, are mostly benefiting from the telehealth TH interventions [3,22].

Overall Perceived Telehealth Nursing Service Quality

Overall SQ is measured by obtaining an average gap score of the SERVQUAL dimensions [10]. As shown in Table 2, the overall average gap score for the SQ of telehealth nursing is 0.05332. In other words, the overall perceived SQ is a positive value (0.05332), which means the perceptions of the level of service are slightly higher than what they expect, indicating there is satisfaction with TNSQ. Overall average gap score for the SQ is calculated as (TA+RL+RN+AS+EM)/5=0.05332.

Demographics and Open-Ended Questions

All the 13 respondents who completed the survey were female, and majority of them were aged 35 years or older. In addition, most of the respondents were seen to have earned a bachelor's degree in nursing. The responses indicate that the nursing staff considered the convenience of scheduling and monitoring patients daily as a benefit of telehealth nursing. This was like the response of the focus group interview question related to the benefits and facilitators of telehealth interventions leading to peace of mind of the patients and the nursing staff. These responses belong to the tangible dimensions, highlighting the benefits of technology, like the findings documented by previous studies [3,23].

The second open-ended question was asked to get the nurses' perspectives related to what they liked most about telehealth. Eight of the responses obtained mapped mainly to the reliability dimension and one each to the tangibles and responsiveness dimensions. This indicated that the nurses like the security of monitoring the patients daily, which adds to the convenience and immediate review of the information, adding validity to the previous open-ended question and interview questions. The final open-ended question was intended to get the perspective of what the nurses did not like about telehealth. A total of 9 participants responded, and majority of their responses mapped to the tangibles dimension and only one to empathy. The responses highlighted the nurses' perspectives related to what they did not like about telehealth, and these responses were on the same pattern as the interview question part b, emphasizing the barriers or negative aspects of TNSQ. The responses mostly belonged to the tangibles dimension related to equipment malfunction, inconsistency of technology, and update issues (Multimedia Appendix 2). There are similar findings documented in previous studies [3,24].

Discussion

Principal Findings

The responses to the interview and data gathered from the survey showed overall satisfaction with TNSQ. Some issues of equipment malfunction and inconsistency in the telehealth equipment were pointed out by the respondents of the survey. The results of this case study are helpful in reinforcing the positive role of telehealth in impacting patient care and providing alternative solutions to the management of chronic disease. The VNA experience of implementing telehealth methods in the case of chronic patients suffering from COPD and heart disease is leading to better quality of life, as well as offering preventive services at a lower cost. Elderly patients suffering from chronic diseases, especially COPD and heart disease, are mostly benefiting from the TH interventions [3,22].

Limitations

This research was an initial pilot study on TNSQ. Due to the limit in time and resources, other telehealth nursing organizations were not contacted; therefore, the sample size was restricted to 13 participants belonging to the same organization VNA.

Implications for Future Research and Practice

Conducting nation-wide nursing telehealth quality service studies would have a much more significant impact on research and perceptions of nursing telehealth quality service. Future research of telehealth interventions and chronic disease management in elderly home care will be important. A bigger sample size would allow comprehensive statistical data analysis to be possible. Other telehealth nursing organizations that are willing to participate in the survey will be contacted. More data will be collected, and CHERRIES checklist for reporting the online survey will be used as in this case study [25]. The user experience of the consumers, patients, and health care providers will be another aspect to follow-up. To understand the positive implementation of telehealth nursing, it is important to dig deeply into the successful models implemented, and learn from the barriers and obstacles that other organizations are facing in relation to the telehealth nursing intervention methods [21]. Any researcher who may consider telehealth nursing for future

research should consider taking this survey to a larger number of telehealth nursing professionals; this will give quantitative validity. Comparisons with other nations such as Norway in Europe, where telehealth implementation is successful, would be of interest for research purposes [26].

Conclusions

SERVQUAL instrument was a good framework to assess TNSQ. The main advantage of SERVQUAL was the ability to estimate not only the overall level of satisfaction but also to identify the dimensions where perceptions transcended expectations (indicating performance excellence in the dimensions of reliability and empathy) and dimensions where the experience falls short of expectations (the dimensions of tangibles, assurance, and responsiveness; Multimedia Appendix 2). The interview responses and data gathered from the survey showed overall satisfaction with TNSQ.

Acknowledgments

The authors would like to thank the VNA of Omaha for their help and support regarding the telehealth nursing process. The research was made possible because of the valuable insight and guidance of Dr Ann L Fruhling and Dr Deepak Khazanchi and the cooperation and support of the faculty and staff of the School of Interdisciplinary Informatics, College of Information Science and Technology, University of Nebraska.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The telehealth nursing service quality (TNSQ) survey instrument.
[PDF File (Adobe PDF File), 49 KB - medinform_v6i2e35_app1.pdf]

Multimedia Appendix 2
Responses to the open-ended questions.
[PDF File (Adobe PDF File), 30 KB - medinform_v6i2e35_app2.pdf]

References

1. Shortliffe EH, Cimino JJ. Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics) 4th ed. London, UK: Springer-Verlag London; 2014:541-560.
2. Doarn CR, Pruitt S, Jacobs J, Harris Y, Bott DM, Riley W, et al. Federal efforts to define and advance telehealth--a work in progress. Telemed J E Health 2014 May;20(5):409-418 [FREE Full text] [doi: 10.1089/tmj.2013.0336] [Medline: 24502793]
3. Black JT, Romano PS, Sadeghi B, Auerbach AD, Ganiats TG, Greenfield S, BEAT-HF Research Group. A remote monitoring and telephone nurse coaching intervention to reduce readmissions among patients with heart failure: study protocol for the Better Effectiveness After Transition - Heart Failure (BEAT-HF) randomized controlled trial. Trials 2014 Apr 13;15:124 [FREE Full text] [doi: 10.1186/1745-6215-15-124] [Medline: 24725308]
4. Hogan TP, Wakefield B, Nazi KM, Houston TK, Weaver FM. Promoting access through complementary eHealth technologies: recommendations for VA's Home Telehealth and personal health record programs. J Gen Intern Med 2011 Nov;26 Suppl 2:628-635 [FREE Full text] [doi: 10.1007/s11606-011-1765-y] [Medline: 21989614]
5. Taunton RL, Bott MJ, Koehn ML, Miller P, Rindner E, Pace K, et al. The NDNQI-Adapted Index of work satisfaction. J Nurs Meas 2004;12(2):101-122. [Medline: 16092709]
6. Yun EK, Chun KM. Critical to quality in telemedicine service management: application of DFSS (Design for Six Sigma) and SERVQUAL). Nurs Econ 2008;26(6):384-388. [Medline: 19330973]
7. Peck A. Changing the face of standard nursing practice through telehealth and telenursing. Nurs Adm Q 2005;29(4):339-343. [Medline: 16260998]
8. Yavas U, Babakus E, Westbrook KW, Grant CC, Deitz GD, Rafalski E. An investigation of service quality—willingness to recommend relationship across patient and hospital characteristics. J Health Manag 2016;18(1):49-69. [doi: 10.1177/0972063415625508]
9. Grönroos C. A service quality model and its marketing implications. Eur J Mark 1984;18(4):36-44. [doi: 10.1108/EUM0000000004784]
10. Parasuraman A, Zeithaml VA, Berry LL. A conceptual model of service quality and its implications for future research. J Mark 1985;49(4):41-50. [doi: 10.2307/1251430]
11. Grönroos C. Innovative marketing strategies and organization structures for service firms. In: Berry LL, Shostack GL, Upah GD, editors. Emerging Perspectives on Services Marketing. Chicago: AMA; 1983:9-21.

12. Fotiadis AK, Vassiliadis CA. The effects of a transfer to new premises on patients' perceptions of service quality in a general hospital in Greece. Total Qual Manag Bus Excell 2013;24(9-10):1022-1034. [doi: 10.1080/14783363.2013.799328]
13. Rodríguez-Martín B, Martínez-Andrés M, Cervera-Monteagudo B, Notario-Pacheco B, Martínez-Vizcaíno V. Perception of quality of care among residents of public nursing-homes in Spain: a grounded theory study. BMC Geriatr 2013 Jun 28;13:65 [FREE Full text] [doi: 10.1186/1471-2318-13-65] [Medline: 23809066]
14. Snooks HA, Williams AM, Griffiths LJ, Peconi J, Rance J, Snelgrove S, et al. Real nursing? The development of telenursing. J Adv Nurs 2008 Mar;61(6):631-640. [doi: 10.1111/j.1365-2648.2007.04546.x] [Medline: 18302604]
15. Eaves YD. A synthesis technique for grounded theory data analysis. J Adv Nurs 2001 Sep;35(5):654-663. [Medline: 11529967]
16. Khazanchi D, Shank M, Reynolds J. The impact of voice mail technology on internal service quality. J Customer Serv Mark Manag 1998 Sep 2;4(4):75-106.
17. Hogan TP, Wakefield B, Nazi KM, Houston TK, Weaver FM. Promoting access through complementary eHealth technologies: recommendations for VA's Home Telehealth and personal health record programs. J Gen Intern Med 2011 Nov;26 Suppl 2:628-635 [FREE Full text] [doi: 10.1007/s11606-011-1765-y] [Medline: 21989614]
18. LeRouge CM, Garfield MJ, Hevner AR. Patient perspectives of telemedicine quality. Patient Prefer Adherence 2015;9:25-40 [FREE Full text] [doi: 10.2147/PPA.S67506] [Medline: 25565781]
19. Darkins A, Ryan P, Kobb R, Foster L, Edmonson E, Wakefield B, et al. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health 2008 Dec;14(10):1118-1126. [doi: 10.1089/tmj.2008.0021] [Medline: 19119835]
20. Choi KS, Lee H, Kim C, Lee S. The service quality dimensions and patient satisfaction relationships in South Korea: comparisons across gender, age and types of service. J Serv Mark 2005;19(3):140-149. [doi: 10.1108/08876040510596812]
21. Sproull NL. Handbook of Research Methods: A Guide for Practitioners and Students in the Social Sciences. Lanham, Maryland: Scarecrow Press; 2003.
22. Radhakrishnan K, Jacelon C, Roche J. Perceptions on the use of telehealth by homecare nurses and patients with heart failure: a mixed method study. Home Health Care Manag Pract 2012;24(4):175-181. [doi: 10.1177/1084822311428335]
23. Silva GS, Farrell S, Shandra E, Viswanathan A, Schwamm LH. The status of telestroke in the United States: a survey of currently active stroke telemedicine programs. Stroke 2012 Aug;43(8):2078-2085 [FREE Full text] [doi: 10.1161/STROKEAHA.111.645861] [Medline: 22700532]
24. Young H, Miyamoto S, Ward D, Dharmar M, Tang-Feldman Y, Berglund L. Sustained effects of a nurse coaching intervention via telehealth to improve health behavior change in diabetes. Telemed J E Health 2014 Sep;20(9):828-834 [FREE Full text] [doi: 10.1089/tmj.2013.0326] [Medline: 25061688]
25. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
26. Nagel DA, Penner JL. Conceptualizing telehealth in nursing practice: advancing a conceptual model to fill a virtual gap. J Holist Nurs 2016 Mar;34(1):91-104. [doi: 10.1177/0898010115580236] [Medline: 25858897]

Abbreviations

AS: assurance
ATA: American Telemedicine Association
COPD: chronic obstructive pulmonary disease
EM: empathy
IT: information technology
RATER: Reliability, Assurance, Tangibles, Empathy, Reliability
RL: reliabilty
RS: responsiveness
SQ: service quality
TA: tangibles
TNSQ: telehealth nursing service quality
VNA: Visiting Nurse Association

Edited by G Eysenbach; submitted 28.09.17; peer-reviewed by S Miyamoto, H Young, P Więch, S Assi; comments to author 15.03.18; revised version received 18.03.18; accepted 05.04.18; published 25.05.18.

Please cite as:
Bashir A, Bastola DR
Perspectives of Nurses Toward Telehealth Efficacy and Quality of Health Care: Pilot Study
JMIR Med Inform 2018;6(2):e35
URL: http://medinform.jmir.org/2018/2/e35/
doi:10.2196/medinform.9080
PMID:29802089

Review

Health Information Technology in Healthcare Quality and Patient Safety: Literature Review

Sue S Feldman[1*], RN, MEd, PhD; Scott Buchalter[2*], MD; Leslie W Hayes[3*], MD

[1]Department of Health Services Administration, The University of Alabama at Birmingham, Birmingham, AL, United States
[2]Pulmonary and Critical Care, The University of Alabama at Birmingham Medical Center, Birmingham, AL, United States
[3]Department of Pediatrics, The University of Alabama at Birmingham Medical Center, Birmingham, AL, United States
[*]all authors contributed equally

Corresponding Author:
Sue S Feldman, RN, MEd, PhD
Department of Health Services Administration
The University of Alabama at Birmingham
1716 9th Avenue
SHPB #590K
Birmingham, AL, 35294
United States
Phone: 1 6616188805
Email: suefeldman1009@gmail.com

Related Article:

This is a corrected version. See correction statement: https://medinform.jmir.org/2019/1/e11320/

Abstract

Background: The area of healthcare quality and patient safety is starting to use health information technology to prevent reportable events, identify them before they become issues, and act on events that are thought to be unavoidable. As healthcare organizations begin to explore the use of health information technology in this realm, it is often unclear where fiscal and human efforts should be focused.

Objective: The purpose of this study was to provide a foundation for understanding where to focus health information technology fiscal and human resources as well as expectations for the use of health information technology in healthcare quality and patient safety.

Methods: A literature review was conducted to identify peer-reviewed publications reporting on the actual use of health information technology in healthcare quality and patient safety. Inductive thematic analysis with open coding was used to categorize a total of 41 studies. Three pre-set categories were used: prevention, identification, and action. Three additional categories were formed through coding: challenges, outcomes, and location.

Results: This study identifies five main categories across seven study settings. A majority of the studies used health IT for identification and prevention of healthcare quality and patient safety issues. In this realm, alerts, clinical decision support, and customized health IT solutions were most often implemented. Implementation, interface design, and culture were most often noted as challenges.

Conclusions: This study provides valuable information as organizations determine where they stand to get the most “bang for their buck” relative to health IT for quality and patient safety. Knowing what implementations are being effectivity used by other organizations helps with fiscal and human resource planning as well as managing expectations relative to cost, scope, and outcomes. The findings from this scan of the literature suggest that having organizational champion leaders that can shepherd implementation, impact culture, and bridge knowledge with developers would be a valuable resource allocation to consider.

(JMIR Med Inform 2018;6(2):e10264) doi:10.2196/10264

KEYWORDS

Health Information Technology; Healthcare Quality; Patient Safety

Introduction

Background

It has long been known and accepted that healthcare in the US is too expensive and the outcomes are less than predictable [1]. The turn of the century brought with it a realization that healthcare, like other industries, could use data to increase our awareness of seemingly uncontrollable costs and unpredictable outcomes. With almost two decades of compiling, analyzing, mashing up data, and trying to make sense of how the data inform multiple layers of healthcare, it is time to look beyond the awareness that the data provide, and instead develop an understanding of how to use the data for predictable and actionable purposes, especially with regard to healthcare quality and patient safety. The literature is mixed on the degree to which health information technology (IT) as a valuable suite of tools, applications, and systems that have contributed to actual savings and efficiencies [1-4]. However, the area of healthcare quality and patient safety lends itself to many of the same business intelligence and predictability advantages that are seen in the credit card industry [5-7].

Much like the Triple Aim of Healthcare, the credit card industry is working toward decreased costs (fraud), increased quality (better transactions), and increased satisfaction (happier merchants and happier cardholders). The credit card industry began using business intelligence to predict behavior that suggested fraud, developed process maps for transaction processing, and offered perks to merchants and cardholders. Just as the credit card industry learned from healthcare, healthcare can borrow from the credit card industry to use healthcare intelligence for prevention, identification, and action related to healthcare quality and patient safety events.

The Institute for Healthcare Improvement (IHI) suggests that reliability around healthcare is a three-part cycle of failure prevention, failure identification, and process redesign and defines reliability as "failure-free operation over time." [8]. Other areas of healthcare have used information systems to provide continuous monitoring with real-time, or near real-time reporting as a means of achieving reliability [9]. As such, it makes sense to think about the role of health IT in reliability as it relates to healthcare quality and patient safety. A review of the literature suggests that healthcare organizations are using health IT for healthcare quality and patient safety and that they have replaced *redesign* in Figure 1 with *action* as shown in Figure 2 [10-12]. Action, in this case, allows for health IT to be implemented after a potential healthcare quality or patient safety event has occurred and does not necessarily require a redesign. Ordering alerts in the electronic health record are an example of action; the event has occurred (the order has been entered) and health IT in the form of an alert is initiated to stop the potentially unsafe order from being filled by the pharmacy.

Having an understanding of this cycle helps to create awareness around where various applications of health IT find their "best fit" in improving the reliability of healthcare quality and patient safety. A distinct advantage of this being a "cycle" is that there is no defined beginning and ending point, but rather an insertion point. This is all to say that the cycle should not be interpreted as starting with prevention and ending with action.

Health Information Technology for Prevention of Quality and Safety Events

Health IT for prevention of quality and safety events involves the use of health IT to *prevent a quality and safety event from even happening*. Automated reminders and alerts are useful in providing essential information that supports safe and effective clinical decisions [13]. Such alerts in the electronic health record (EHR) are a standard mechanism for the use of health IT for prevention of potential missed quality and patient safety events. For example, immunization alerts have led to a 12% increase in well-child and a 22% increase in sick child immunization administration [14] and drug alerts have been associated with a 22% decrease in medication prescription errors [15]. Soft-stops can provide key information about a potential quality or patient safety issue. They may offer choices but usually, require only that the user acknowledge the alert to proceed.

Figure 1. Improving the reliability of healthcare.

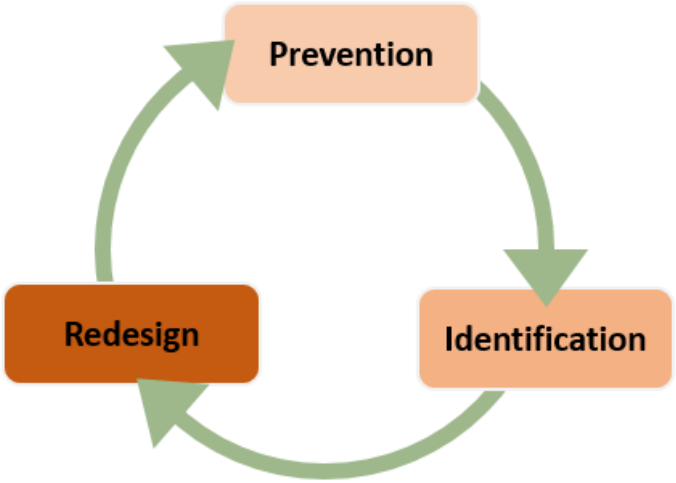
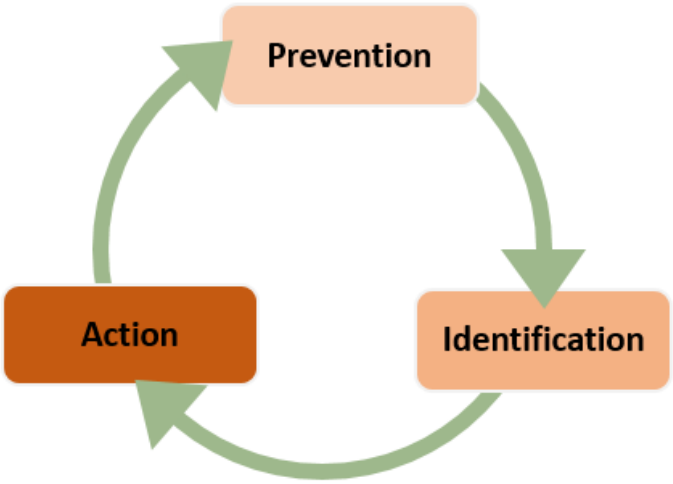


Figure 2. Improving the reliability of healthcare quality and patient safety.



A hard-stop, on the other hand, prevents the user from moving forward with an order or intervention that would be potentially dangerous to a patient. Hard-stops may allow continuation of the process, but only if significant required action is taken by the user, such as a call to or consultation with an expert (such as a pharmacist or a medical specialist). In some cases, soft-stops might be ignored or overridden because of such issues as alert fatigue, poor implementation, or poor interface design [16,17]. Hard-stops, when appropriately designed, have been shown to be more successful in changing an unsafe plan or preventing a potentially dangerous intervention [18,19].

Health Information Technology for Identification of Quality and Safety Events

Health IT for identification of quality and safety events involves health IT that is used to *identify a quality and safety event when it is about to occur*. Health insurance providers increasingly place pressure on healthcare systems to reduce the cost of care delivery and improve patient outcomes. This pressure may exist through tiered reimbursement structures, benefitting those systems which meet or exceed specific benchmarks of performance. Growing pressure from these payers takes the form of non-reimbursement for care determined by the payer to be unnecessary or in excess of "standard care." Health IT can be used to find the EHR populations of patients for whom reimbursement might be lower than expected. One such example to consider is the length of stay for a particular procedure. While the use of health IT can produce reports and dashboards that are helpful for decision-making relative to reimbursement trends and practices for lengths of stay for that diagnosis, it is crucial that thoughtful consideration be given for appreciating any unintended consequences. For example, when reducing the length of stay, unintended readmissions are an important metric to follow.

Health Information Technology for Action in Quality and Safety Events

Health IT for action of quality and safety events involves health IT that is used to a*ct on a quality and safety event once it has already occurred.* That is to say that these are actions that were reported in the literature that were taken *as a result of* an event. Health IT for action differs from health IT for prevention in that the former is a reaction directly correlated to an event reported in the article, whereas the latter is reported in the article as a preemptive measure, in advance of an event.

Because of their standardization, there are several clinical care pathways that lend themselves to clinical decision support. One such clinical care pathway is sepsis. Despite nearly two decades of advances in early sepsis care, sepsis outcomes persist to be poor, and sepsis remains a leading cause of death worldwide and accounts for significant morbidity and mortality [20]. In light of this, there is a growing national push to increase early identification and treatment of sepsis with a goal of improving outcomes. Patients with sepsis are some of the most critically ill patients admitted to hospitals, and survival depends heavily upon timely and early administration of key interventions followed quickly by assessing and acting on results of these interventions [21]. Some examples include administration of IV antibiotics and aggressive IV fluids within one hour [21]. Examples of assessments of interventions include measuring specific physical and laboratory values that provide crucial information about the patient response. All too often, clinicians are faced with an overabundance of data, that while all necessary, may not be relevant to the issue at hand. For example, lab results might be presented in their entirety, when in practice, there are only 3 or 4 tests that will drive decision-making. The difficulty is how to separate the noise (non-essential at that moment) from the signal (essential at that moment). Health IT solutions, such as dashboards and other solutions can be used to ensure that essential data are in a primary viewing position and non-essential data in a secondary viewing position (perhaps on drill down, for example).

This paper will provide foundational knowledge and understanding for organizations of where to focus health IT fiscal and human resources. It will also provide information relative to some of the challenges that can be expected in implementing health IT for quality and patient safety.

Methods

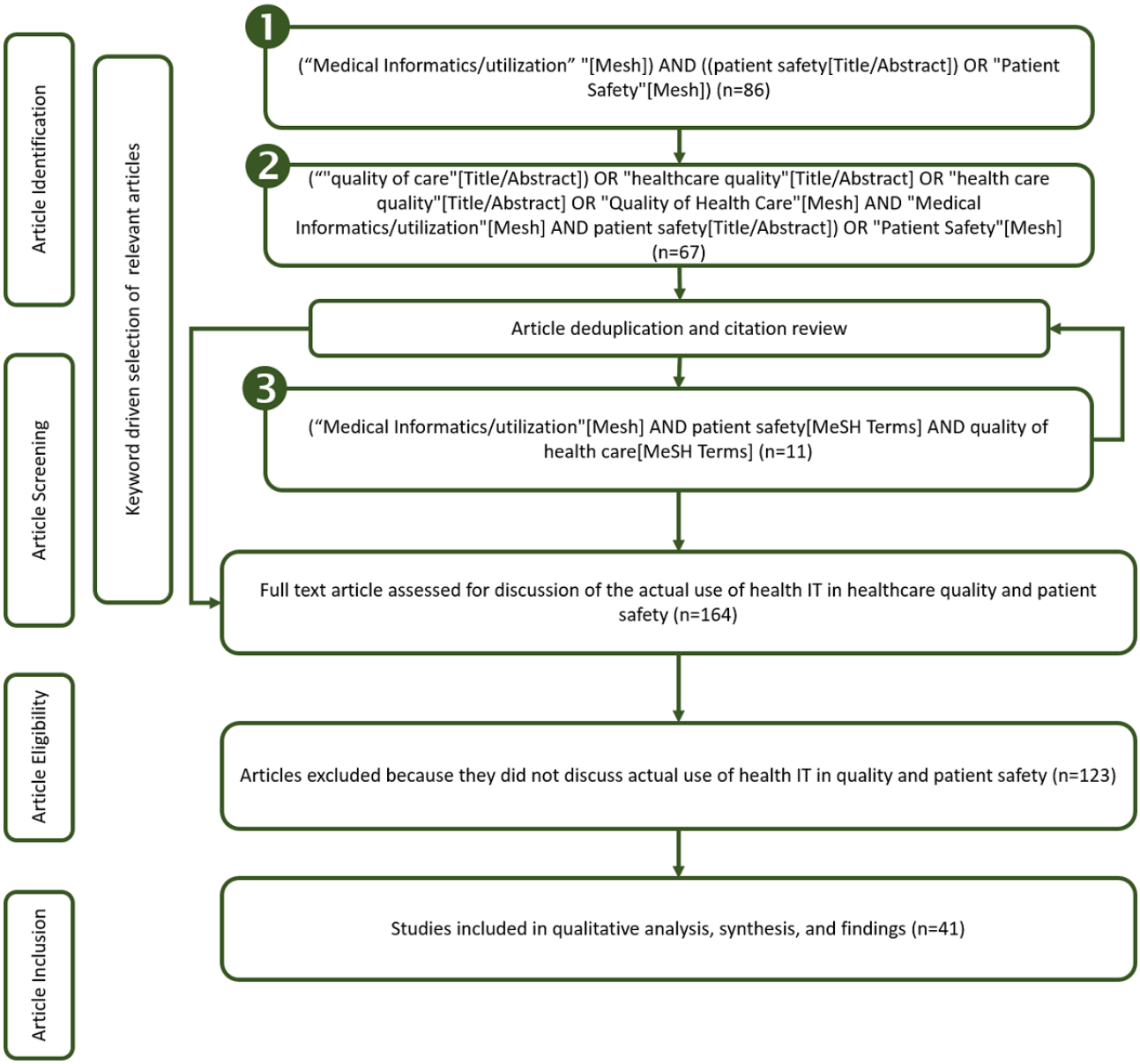
This review of the literature took a structured approach using PubMed and a combination of keywords. Since PubMed indexes peer-reviewed articles from biomedical information, it was felt

that this was the most appropriate and inclusive source. A healthcare-focused librarian, under the direction of all authors, conducted the literature search. The articles for final selection were discussed and decided upon among the authors. The structured approach was guided by the model illustrated in Figure 3.

The process to article inclusion involved three passes to collect publications related to health IT in quality and patient safety for peer-reviewed studies published between 2012-2017, inclusive. The first pass, (shown as "1" in Figure 3), used high-level keywords and returned 86 full-text articles. From the articles gathered, additional keywords were added to the search. After deduplication and citation review, the second pass (shown as "2" in Figure 3) added 67 unique full-text articles. After deduplication and citation review, the third pass (shown as "3" in Figure 3) added 11 unique full-text articles, for a total of 164 unique full text articles. Each article was further analyzed to identify the degree to which the article discussed health IT in healthcare quality and patient safety. To be considered for inclusion, the study needed to report on the *actual* use of health IT in healthcare quality and patient safety. Forty-one studies met these criteria. Those studies with their contributions to the results are shown in the results section of this paper.

Qualitative data analysis software (Atlas.ti 8 for Windows) was used in directed content analysis as a method to categorize and code the 41 studies relative to *how* health IT was used in healthcare quality and patient safety. All 41 documents were uploaded into the document manager in Atlas.ti as Primary Documents (PD). During this process, the article title was used as the PD name. Inductive thematic analysis with open coding was used under the three pre-set categories of prevention, identification, and action [22]. This allowed for capturing descriptions of how health IT was used in each circumstance.

Figure 3. Literature search process.



For example, prevention included descriptions of any use of health IT to prevent quality issues or potential safety events, identification included any descriptions of the use of health IT to identify quality issues or safety events, and action included any descriptions of the use of health IT to act on quality issues or safety events that have occurred. When content was noted that did not fit into the three pre-set categories, an additional category was created. Additional categories were created to capture challenges relative to the use of health IT in quality and patient safety. Since some papers discussed how the use of health IT impacted health outcomes, an additional category was created for outcomes. Lastly, an additional category was created to capture the study settings or location.

The coding structure was agreed upon by all authors, and one author conducted the coding. After all of the studies were coded, two additional passes were made through the data. The first pass was to ensure that all information from the studies that should be coded was actually coded and coded to the correct code (ie, was a passage that described prevention actually coded to prevention?). The second pass was to consider sub-categories for consolidation. Six sub-categories were consolidated.

The purpose of examining co-occurrences is to understand what, if any, relation exists between concepts [22,23]. Within Atlas.ti, a co-occurrence table was run to find codes that co-occur across the literature, the purpose of which was to illuminate the areas most discussed. This table was then exported to Microsoft Excel for further analysis.

Network maps are a means by which analysis can be visualized in relationships to provide a different perspective on the codes, categories, etc., and with that visualization, provide a mechanism for moving codes around [22]. Those presented in the results do not differ from the final coding structure, but instead are used to provide a visual representation.

Results

Overview

Literature reviews can be conducted using a qualitative approach [24,25] with the results displayed in a variety of ways to support models and show connections [22]. As such, this review presents qualitative findings to support the "improving the reliability of healthcare quality and patient safety" model introduced earlier in this paper and shows connections via network mappings in Figure 6 through Figure 7 and co-occurrences in Table 2.

Table 1 provides a listing of the articles and their contribution in this results section to support the model (Figure 2), network maps (Figure 4 through Figure 7), and co-occurrences (Table 2).

From the 41 studies that fit the inclusion criteria, any element in which the authors discussed the use of health IT for healthcare quality and patient safety was identified, even if it did not fit into the three previously determined categories. This process yielded a total of 50 codes across five categories: action (7/41, 17.1%), challenges (12/41, 29.3%), identification (10/41, 24.4%), outcomes (5/41, 12.2%), and prevention (16/41, 39.0%) across seven study settings. Just under a quarter of the studies identified a study setting: anesthesia (2/41, 4.9%), behavioral health (1/41, 2.4%), emergency department (2/41, 4.9%), any intensive care unit (3/41, 7.3%), clinical diagnostic laboratory (1/41, 2.4%), pediatrics (2/41, 4.9), surgery (1/41, 2.4%).

Across all of the articles, there were 63 and 92 descriptions of the use of health IT for identification and prevention of healthcare quality and patient safety issues, respectively. Health IT for action and the challenges associated with health IT for healthcare quality and patient safety was described 41 and 43 times, respectively.

The findings from the literature review are presented by the categories outlined in the previously introduced model for improving the reliability of healthcare quality and patient safety.

Prevention

The first exploration was across the literature that discussed health IT for *prevention* of quality and patient safety issues to see exactly how organizations were reporting health IT use to *prevent a quality and safety event from even happening*. The greatest areas of use were around alerts [30,31,44,56,58], clinical decision support [39,44,47,56], implementation [10,32,37,38,56], interface design [26,34,42,45,56,59], and customized health IT solutions [29,30,32,34,46-50,56,58,59]. Customized health IT solutions were anything that described the use of health IT but lacked any specificity beyond that described in this section. For example, this could be something as simple as checklists or as complex as algorithmic diagnostic trees. To clarify, alerts are a subset of clinical decision support. Since so many of the occurrences specified alerts and clinical decision support separately, these were coded separately. Clinical decision support, by definition, includes alerts, clinical care guidelines, condition-specific orders sets, clinical reports and/or summaries, documentation templates, diagnostic support, and clinical reference support. Implementation and interface design were each described in terms of having been poorly implemented or poorly designed and having implications on utility in healthcare quality and safety.

Identification

The next exploration was across the literature that discussed health IT for *identification* of quality and patient safety issues; in other words, how health IT was used to *identify a quality and safety event when it is about to occur*. In this regard, similar to prevention (but described differently in the included studies), alerts [26,30,31,44,56,58], clinical decision support [30,31,39,44,56,58], implementation [10,32,38,56], and customized health IT solutions [10,30,31,34,46-49,52,56,58] were most prominent. For example, alerts, clinical decision support, and customized health IT solutions were all described in the literature as having been implemented to identify a potential quality or patient safety issue, yet the literature also described how the implementation of these could have been better in terms of providing more training to those on the receiving end of the alerts, clinical decision support, or other customized health IT solutions.

Table 1. Article contribution to results (in alphabetical order). An "X" indicates the area of the results contribution and "—" indicates no contribution.

Citation	Action	Challenges	Identification	Outcomes	Prevention
Ancker et al [10]	X	X	X	—	X
Arabi et al [26]	—	—	X	X	X
Asch et al [27]	—	—	—	—	X
Badrick et al [28]	—	—	—	—	X
Coiera et al [29]	—	X	—	—	X
Colicchio et al [11]	X	—	—	—	—
El Morr et al [30]	—	—	X	—	X
Every et al [31]	—	—	X	—	X
Farzandipour et al [32]	—	X	X	—	X
Gupta and Kaplan [12]	X	—	—	—	—
Hoonakker et al [33]	—	—	X	—	X
Jensen [34]	X	X	X	—	X
Khullar et al [35]	X	—	—	—	—
Kim et al [36]	—	X	—	—	X
Koppel [37]	X	X	—	—	X
Lassere et al [38]	X	X	X	—	X
Levesque et al [39]	X	—	X	X	X
Magrabi et al [40]	X	—	—	—	—
Martin et al [41]	—	X	—	—	—
Mazur et al [42]	X	—	—	—	X
Nakhleh [43]	-—	—	—	—	X
Peters [44]	X	X	X	—	X
Popovici [45]	—	X	—	—	X
Rizzato et al [46]	X	X	X	—	X
Seblega et al [47]	X	—	X	X	X
Shy et al [48]	X	—	X	—	X
Skyttberg et al [49]	—	X	X	X	X
Stanton [50]	X	—	—	—	X
Strickland [51]	—	—	—	—	X
Suresh [52]	—	X	X	—	X
Wang et al [53]	—	X	—	—	X
Weiner [54]	—	—	—	—	X
Whipple et al [55]	X	—	—	—	—
Whitt et al [56]	X	X	X	—	X
Yermak, et al [57]	—	—	—	—	X
Yu et al [58]	—	X	X	—	X

Figure 4. ACTION Network Diagram (G=groundedness, D=density).

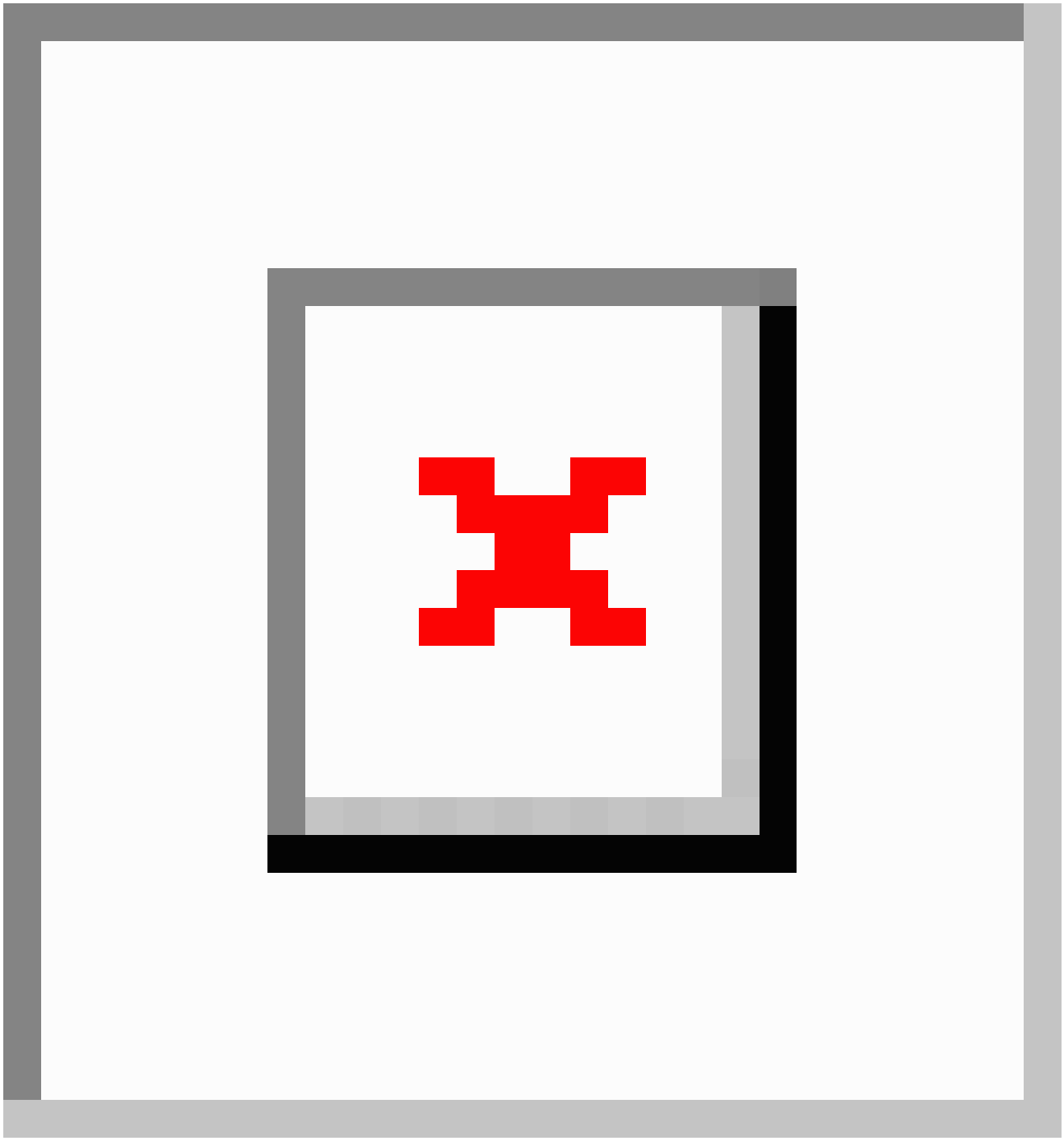


Figure 5. OUTCOMES Network Diagram (G=groundedness, D=density).

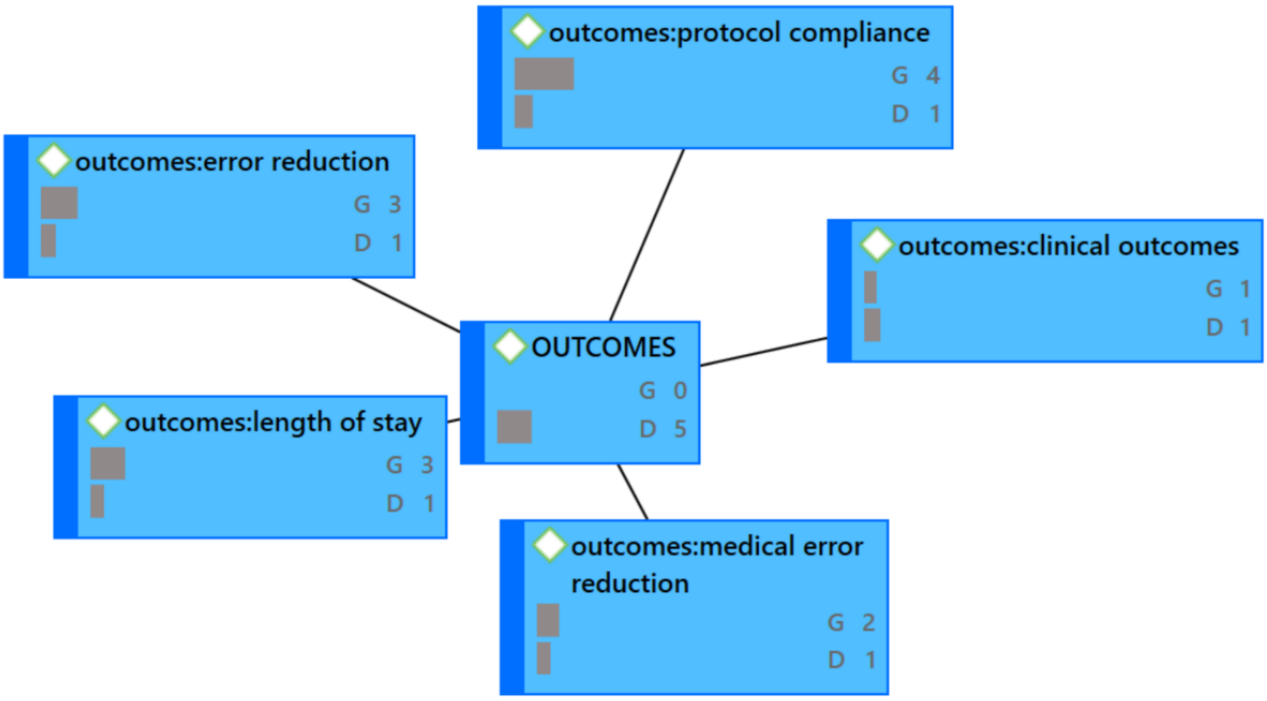


Figure 6. IDENTIFICATION Network Diagram (G=groundedness, D=density).

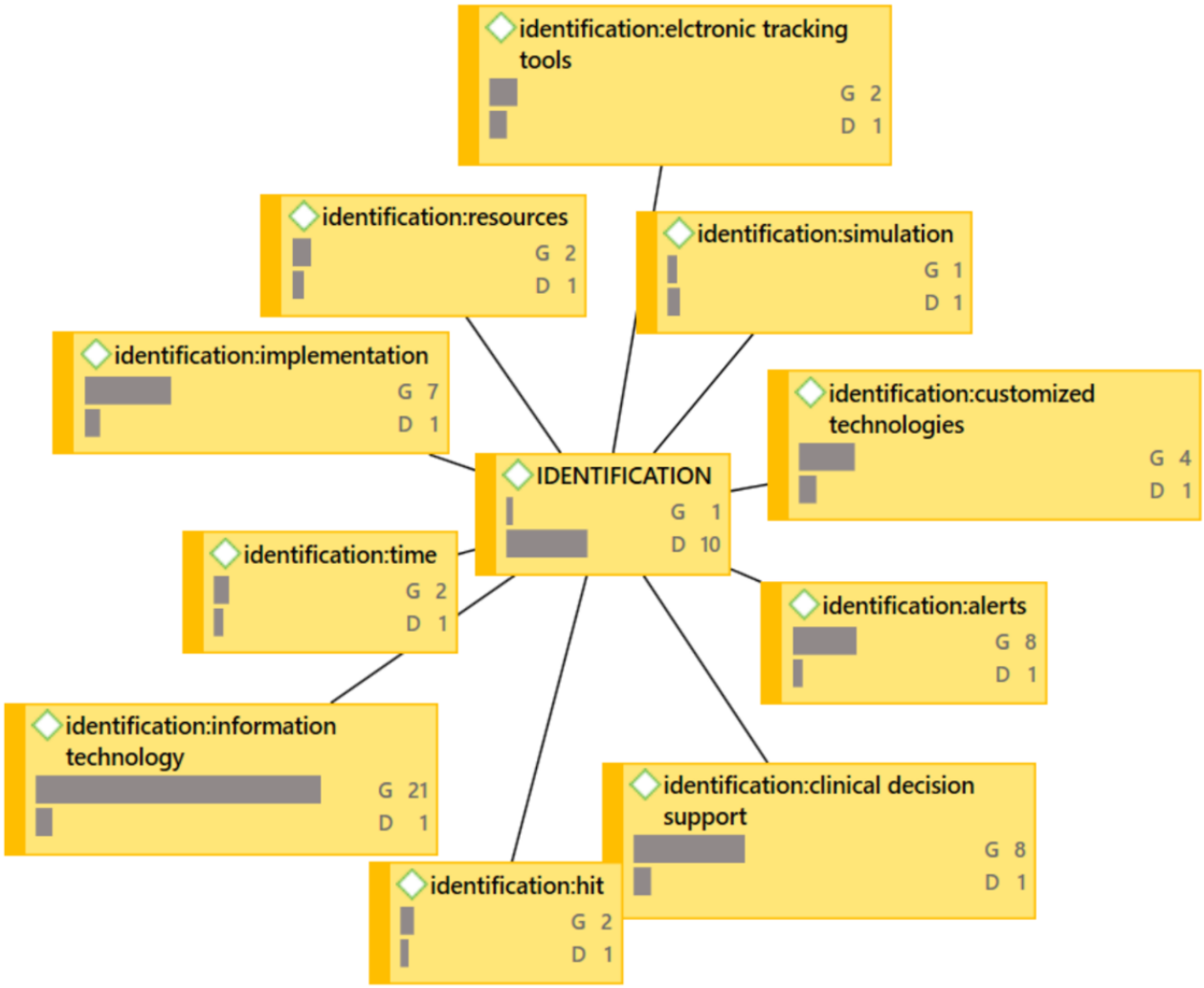


Figure 7. PREVENTION Network Diagram (G=groundedness, D=density).

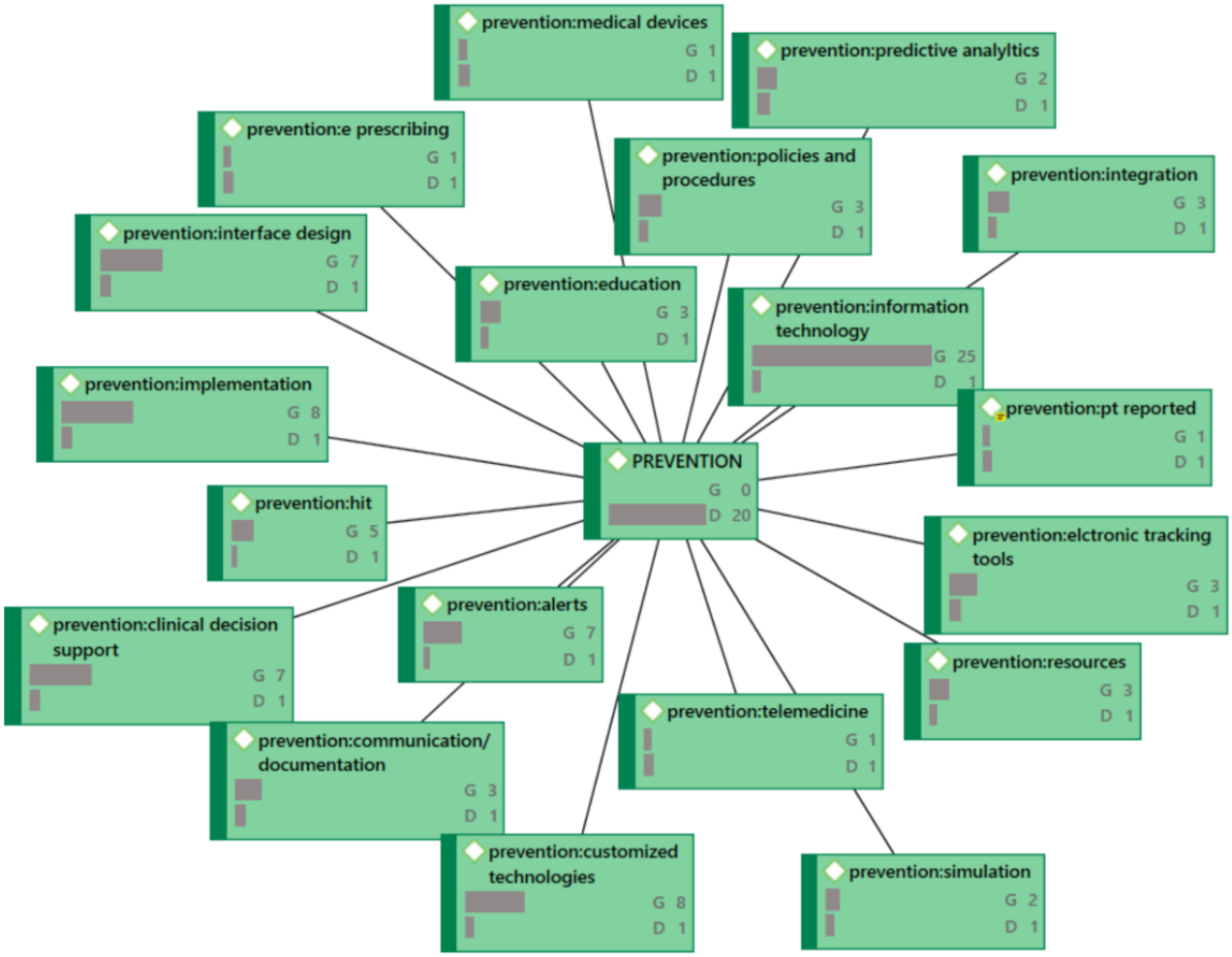


Table 2.

Code	Co-occurrences
Implementation	Prevention, Identification, Action, Challenges
Alerts	Prevention, Identification
Clinical decision support	Prevention, Identification
Interface design	Prevention, Challenges
Culture	Action, Challenges (tattling)
Customized health IT[a] solutions	Prevention, Identification

[a]IT: information technology.

Action

The third exploration was across the literature that discussed health IT for *action on a quality and safety event once it has already occurred.* That is to say that these are actions that were reported in the literature that were taken *as a result of* an event. In regards to action, the major areas were documentation [10,32,37,41,46,56,58], implementation [10,32,37,58], and culture [10,29,41,53,58] relative to the use of health IT.

The findings from the review of the literature show that implementation appeared in prevention, identification, *and* action. Implementation in general has been demonstrated in the literature as a challenge, and that was revealed in this literature review also. Culture was most often referred to as needing to create a culture of quality and patient safety in order for health IT to be embraced. Organizations that started working on culture change before implementation of health IT solutions suggested that health IT for acting on quality and patient safety events was more favorable. Therefore, the analysis was run with challenges which suggests the major areas are: culture, implementation, and interface design.

Co-occurrences

Employing the Improving the reliability of healthcare quality and patient safety model introduced in Figure 2 and adding challenges, six critical co-occurrences emerged (see Table 2).

As described earlier, co-occurrences expose relationships exists between concepts [22,23]. The top co-occurring codes in Table 2 create a macro level view of how health IT was most commonly used for quality and patient safety relative to the "improving the reliability of healthcare quality and patient safety" model introduced in Figure 2. However, it is also important to understand the universe of ways in which organizations used health IT for quality and patient safety; in other words, the art of the possible when using health IT for quality and patient safety. Network maps provide a mechanism by which to visualize the connectedness of all data coded across all 41 articles included in this analysis. These maps, along with some quantitative information increase understanding at this universe level (macro and micro views).

In the network diagrams that follow (which also represent the coded categories and sub-categories), G signifies the level of groundedness of the particular code. Groundedness, in this case, indicates the frequency of the code relative to the code category. D signifies the level of density or connectedness of the particular code. Density, in this case, indicates the number of other codes to which this code is connected. For example, under ACTION, Figure 4, the code action: culture shows G6, D2. ACTION is the code category and action: culture is the code "culture" under the ACTION code category (this coding structure helps to maintain alpha order). This can be read as the following: "Culture was described six times across all 41 papers relative to ACTION and is connected to two code categories total." Because it would make the network diagrams unwieldy, not shown in the exhibits is the specificity around the groundedness or the density. See Figures 4 through Figure 7.

Discussion

Principal Findings

This scan of the literature is intended to inform practice. The information from this study could be useful as organizations determine where they stand to get the most "bang for their buck" relative to health IT for quality and patient safety. Centered around the Improving the Reliability of Healthcare Quality and Safety model introduced in Figure 2 and the macro level uses of health IT for quality and patient safety outlined in Table 2, organizations in the planning stages may want to begin with alerts and clinical decision support, understanding that alerts are a subset of clinical decision support. This information also helps with resource planning. For example, implementation appeared in all three categories of the Improving the Reliability of Healthcare Quality and Safety model. Additionally, culture was shown to be a challenge. Organizational leaders know that changing culture can be a long and intensive process. The findings from this scan of the literature suggest that having organizational champion leaders that can shepherd implementation, impact culture, and bridge knowledge with developers would be a valuable resource allocation to consider.

Health IT must meet quality improvement at the intersection with care delivery. From a clinical perspective, this is experienced on several levels, and the solution depends, in part, on the clinical problem to be addressed. Some typical examples of health IT interventions illuminated in the findings include: (1) reminders and alerts, (2) decision support tools, (3) checklists (including order sets and protocols), and (4) soft- and hard-stops.

As noted, this scan of the literature is provided as a means to inform practice. It does not consider further model modification, and this represents an area of future research in the application of health IT for quality and patient safety.

Limitations

This study is limited in that it used PubMed as a single source for the searching and one coder coded all studies. A more comprehensive and systematic review would include multiple databases and multiple coders. Although all authors reviewed the codes, multiple coders would ensure intercoder reliability, which cannot be assured in this study. Additionally, since all studies reviewed did not include locations, generalizability to all areas of clinical care cannot be certain.

Conclusion

A review of the literature for this study concluded that organizations in the planning stages of using health IT to improve quality and safety may want to begin with reminders and alerts, decision support tools, and checklists.

Acknowledgments

The authors acknowledge and appreciate the careful, detailed, and thoughtful comments by the reviewers whose suggestions strengthened this paper.

Conflicts of Interest

None declared.

References

1. Marmor T, Oberlander J, White J. The Obama administration's options for health care cost control: hope versus reality. Ann Intern Med 2009 Apr 07;150(7):485-489. [Medline: 19258549]
2. Buntin M, Burke M, Hoaglin M, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health affairs 2011;30(3):a. [doi: 10.1377/hlthaff.2011.0178]

3. Goldzweig C, Towfigh A, Maglione M, Shekelle P. Costs and benefits of health information technology: new trends from the literature. Health affairs 2009;28(2). [doi: 10.1377/hlthaff.28.2.w282]
4. Karsh B, Weinger MB, Abbott PA, Wears RL. Health information technology: fallacies and sober realities. J Am Med Inform Assoc 2010;17(6):617-623 [FREE Full text] [doi: 10.1136/jamia.2010.005637]
5. Jamal A, McKenzie K, Clark M. The impact of health information technology on the quality of medical and health care: a systematic review. HIM J 2009;38(3):26-37. [doi: 10.1177/183335830903800305]
6. McCullough JS, Casey M, Moscovice I, Prasad S. The effect of health information technology on quality in U.S. hospitals. Health Aff (Millwood) 2010 Apr;29(4):647-654. [doi: 10.1377/hlthaff.2010.0155]
7. Parente S, McCullough J. Health information technology and patient safetyvidence from panel data. Health Affairs 2009;28(2):e. [doi: 10.1377/hlthaff.28.2.357]
8. Nolan TR, Haraden C, Griffin F. ,. Improving the Reliability of Health Care. IHI Innovation Series white paper. Boston2004 December 16;www.ihi.org 2017.
9. Kothamasu R, Huang S, VerDuin W. System health monitoring and prognostics — a review of current paradigms and practices. Int J Adv Manuf Technol 2006 Jul;28(9-10):1012-1024 [FREE Full text] [doi: 10.1007/s00170-004-2131-6]
10. Ancker JS, Kern LM, Abramson E, Kaushal R. The Triangle Model for evaluating the effect of health information technology on healthcare quality and safety. J Am Med Inform Assoc 2012 Jan;19(1):61-65 [FREE Full text] [doi: 10.1136/amiajnl-2011-000385]
11. Colicchio TK, Facelli JC, Del FG, Scammon DL, Bowes WA, Narus SP. Health information technology adoption: Understanding research protocols and outcome measurements for IT interventions in health care. J Biomed Inform 2016 Dec;63:33-44 [FREE Full text] [doi: 10.1016/j.jbi.2016.07.018]
12. Gupta M, Kaplan HC. Improving Quality Improvement in Neonatal-Perinatal Care. Clin Perinatol 2017 Sep;44(3):xvii-xxix. [doi: 10.1016/j.clp.2017.06.002]
13. Grissinger M. Small Effort, Big Payoff: Automated Maximum Dose Alerts With Hard Stops. Pharmacy and Therapeutics 2016;41(2):A. [Medline: 26908995]
14. Fiks AG, Grundmeier RW, Biggs LM, Localio AR, Alessandrini EA. Impact of clinical alerts within an electronic health record on routine childhood immunization in an urban pediatric population. Pediatrics 2007 Oct;120(4):707-714 [FREE Full text] [doi: 10.1542/peds.2007-0257]
15. Smith DH, Perrin N, Feldstein A, Yang X, Kuang D, Simon SR, et al. The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation. Arch Intern Med 2006 May 22;166(10):1098-1104 [FREE Full text] [doi: 10.1001/archinte.166.10.1098]
16. van DSH, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809]
17. Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care--an interactive sociotechnical analysis. J Am Med Inform Assoc 2007;14(5):542-549 [FREE Full text] [doi: 10.1197/jamia.M2384]
18. Strom BL, Schinnar R, Aberra F, Bilker W, Hennessy S, Leonard CE, et al. Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial. Arch Intern Med 2010 Sep 27;170(17):1578-1583. [doi: 10.1001/archinternmed.2010.324]
19. Eslami S, de KNF, Abu-Hanna A. The impact of computerized physician medication order entry in hospitalized patients--a systematic review. Int J Med Inform 2008 Jun;77(6):365-376. [doi: 10.1016/j.ijmedinf.2007.10.001]
20. Carcillo JA, Kellum JA. Is there a role for plasmapheresis/plasma exchange therapy in septic shock, MODS, and thrombocytopenia-associated multiple organ failure? We still do not know--but perhaps we are closer. Intensive Care Med 2002 Oct;28(10):1373-1375. [doi: 10.1007/s00134-002-1428-x]
21. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, Early Goal-Directed Therapy Collaborative Group. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001 Nov 08;345(19):1368-1377. [doi: 10.1056/NEJMoa010307]
22. Friese S. Qualitative Data Analysis with ATLAS. In: Qualitative Data Analysis with ATLAS. Thousand Oaks: Sage Publications; 2012:978-0857021311.
23. Ryan G, Bernard H. Techniques to identify themes. Field methods 2003;15(1):85-109. [doi: 10.1177/1525822X02239569]
24. Brewerton P, Millward L. Organizational Research Methods: A Guide for Students and Researchers. In: Organizational Research Methods: A Guide for Students and Researchers. Thousand Oaks: Sage Publishing; 2001.
25. Creswell J. Research design: Qualitative, quantitative, and mixed methods approaches. Newbury Park, CA: Sage Publications; 2008.
26. Arabi YM, Pickering BW, Al-Dorzi HM, Alsaawi A, Al-Qahtani SM, Hay AW. Information technology to improve patient safety: A round table discussion from the 5(th) International Patient Safety Forum, Riyadh, Saudi Arabia, April 14-16, 2015. Ann Thorac Med 2016;11(3):219-223 [FREE Full text] [doi: 10.4103/1817-1737.176877]
27. Asch E, Shah SH, Berkowitz S, Mehta S, Eisenberg RL, Jayadevan R, et al. Resident Idea System: A Novel Tool to Engage Trainees in Quality Improvement at the Institutional Level. J Am Coll Radiol 2017 Dec;14(2):256-261. [doi: 10.1016/j.jacr.2016.07.022]

28. Badrick T, Gay S, McCaughey EJ, Georgiou A. External Quality Assessment beyond the analytical phase: an Australian perspective. Biochem Med (Zagreb) 2017 Feb 15;27(1):73-80 [FREE Full text] [doi: 10.11613/BM.2017.009]
29. Coiera E, Magrabi F, Talmon J. Engineering technology resilience through informatics safety science. J Am Med Inform Assoc 2017 Dec 01;24(2):244-245. [doi: 10.1093/jamia/ocw162]
30. El MC, Ginsburg L, Seungree N, Woollard S, Hensen B. editors. IT Integration and Patient Safety: The Case of a Software Tool 2016.
31. Every N, Fihn S, Sales A, Keane A, Ritchie J, Committee Q. Quality Enhancement Research Initiative in ischemic heart disease: a quality initiative from the Department of Veterans Affairs. Medical care 2000;38(6):a.
32. Farzandipour M, Meidani Z, Riazi H, Sadeqi JM. Nursing Information Systems Requirements: A Milestone for Patient Outcome and Patient Safety Improvement. Comput Inform Nurs 2016 Dec;34(12):601-612. [doi: 10.1097/CIN.0000000000000260]
33. Hoonakker PLT, Carayon P, Cartmill RS. The impact of secure messaging on workflow in primary care: Results of a multiple-case, multiple-method study. Int J Med Inform 2017 Apr;100:63-76. [doi: 10.1016/j.ijmedinf.2017.01.004]
34. Jensen S. Patient safety and quality of care: How may clinical simulation contribute? Knowledge Management and E-Learning 2015;7(3):412-424.
35. Khullar D, Jha AK, Jena AB. Reducing Diagnostic Errors--Why Now? N Engl J Med 2015 Dec 24;373(26):2491-2493 [FREE Full text] [doi: 10.1056/NEJMp1508044]
36. Kim H, Chan H, Gupta S. Value-based Adoption of Mobile Internet: An empirical investigation. Decision Support Systems 2007 Feb;43(1):111-126. [doi: 10.1016/j.dss.2005.05.009]
37. Koppel R. The health information technology safety framework: building great structures on vast voids. BMJ Qual Saf 2016 Apr;25(4):218-220. [doi: 10.1136/bmjqs-2015-004746]
38. Lassere MN, Baker S, Parle A, Sara A, Johnson KR. Improving quality of care and long-term health outcomes through continuity of care with the use of an electronic or paper patient-held portable health file (COMMUNICATE): study protocol for a randomized controlled trial. Trials 2015 Jun 04;16:253 [FREE Full text] [doi: 10.1186/s13063-015-0760-8]
39. Levesque E, Hoti E, Azoulay D, Ichai P, Samuel D, Saliba F. The implementation of an Intensive Care Information System allows shortening the ICU length of stay. J Clin Monit Comput 2015 Apr;29(2):263-269. [doi: 10.1007/s10877-014-9592-4]
40. Magrabi F, Ammenwerth E, Hyppönen H, de KN, Nykänen P, Rigby M, et al. Improving Evaluation to Address the Unintended Consequences of Health Information Technology:. a Position Paper from the Working Group on Technology Assessment & Quality Development. Yearb Med Inform 2016 Nov 10(1):61-69 [FREE Full text] [doi: 10.15265/IY-2016-013]
41. Martin BS, Arbore M. Measurement, Standards, and Peer Benchmarking: One Hospital's Journey. Pediatr Clin North Am 2016 Dec;63(2):239-249. [doi: 10.1016/j.pcl.2015.11.004]
42. Mazur LM, Mosaly PR, Moore C, Comitz E, Yu F, Falchook AD, et al. Toward a better understanding of task demands, workload, and performance during physician-computer interactions. J Am Med Inform Assoc 2016 Nov;23(6):1113-1120. [doi: 10.1093/jamia/ocw016]
43. Nakhleh RE. Role of Informatics in Patient Safety and Quality Assurance. Surg Pathol Clin 2015 Jun;8(2):301-307. [doi: 10.1016/j.path.2015.02.011]
44. Peters TE. Transformational Impact of Health Information Technology on the Clinical Practice of Child and Adolescent Psychiatry. Child Adolesc Psychiatr Clin N Am 2017 Jan;26(1):55-66. [doi: 10.1016/j.chc.2016.07.003]
45. Popovici I, Morita PP, Doran D, Lapinsky S, Morra D, Shier A, et al. Technological aspects of hospital communication challenges: an observational study. Int J Qual Health Care 2015 Jun;27(3):183-188. [doi: 10.1093/intqhc/mzv016]
46. Rizzato LDA, Benítez SE, Mayan JC, Smith MI, Baum AJ, Luna DR, et al. Patient Safety at Transitions of Care: Use of a Compulsory Electronic Reconciliation Tool in an Academic Hospital. Stud Health Technol Inform 2015;216:232-236.
47. Seblega BK, Zhang NJ, Wan TT, Unruh LY, Miller A. Health information technology adoption: effects on patient safety and quality of care. IJHTM 2015;15(1):31. [doi: 10.1504/IJHTM.2015.070519]
48. Shy BD, Kim EY, Genes NG, Lowry T, Loo GT, Hwang U, et al. Increased Identification of Emergency Department 72-hour Returns Using Multihospital Health Information Exchange. Acad Emerg Med 2016 Dec;23(5):645-649 [FREE Full text] [doi: 10.1111/acem.12954]
49. Skyttberg N, Vicente J, Chen R, Blomqvist H, Koch S. How to improve vital sign data quality for use in clinical decision support systems? A qualitative study in nine Swedish emergency departments. BMC Med Inform Decis Mak 2016 Jun 4;16(1). [doi: 10.1186/s12911-016-0305-4]
50. Stanton BF. Pediatric Safety, Quality, and Informatics. Pediatric Clinics of North America 2016 Apr;63(2):xv-xvi. [doi: 10.1016/j.pcl.2016.01.002]
51. Strickland N. Quality assurance in radiology: peer review and peer feedback. Clinical Radiology 2015 Nov;70(11):1158-1164. [doi: 10.1016/j.crad.2015.06.091]
52. Suresh S. The Intersection of Safety, Quality, and Informatics: Solving Problems in Pediatrics. Pediatric Clinics of North America 2016 Apr;63(2):xvii-xviii. [doi: 10.1016/j.pcl.2016.01.001]
53. Zhou Q, Wang H, Jin J, Feng X, Huang X, Zhao X, et al. Quality improvements in decreasing medication administration errors made by nursing staff in an academic medical center hospital: a trend analysis during the journey to Joint Commission International accreditation and in the post-accreditation era. TCRM 2015 Mar:393. [doi: 10.2147/TCRM.S79238]

54. Weiner S, Fink JC. Telemedicine to Promote Patient Safety: Use of Phone-Based Interactive Voice-Response System to Reduce Adverse Safety Events in Pre-dialysis CKD. Advances in Chronic Kidney Disease 2017 Jan;24(1):31-38. [doi: 10.1053/j.ackd.2016.12.004]
55. Whipple EC, E. Dixon B, J. McGowan J. Linking health information technology to patient safety and quality outcomes: a bibliometric analysis and review. Informatics for Health and Social Care 2012 Jun;38(1):1-14. [doi: 10.3109/17538157.2012.678451]
56. Whitt KJ, Eden L, Merrill KC, Hughes M. Nursing Student Experiences Regarding Safe Use of Electronic Health Records. CIN: Computers, Informatics, Nursing 2017;35(1):45-53. [doi: 10.1097/CIN.0000000000000291]
57. Yermak D, Cram P, Kwan JL. Five things to know about diagnostic error. Diagnosis (Berl) 2017 Mar;4(1):13-15 [FREE Full text] [doi: 10.1515/dx-2016-0043]
58. Yu X, Jiang J, Liu C, Shen K, Wang Z, Han W, et al. Protocol for a multicentre, multistage, prospective study in China using system-based approaches for consistent improvement in surgical safety. BMJ Open 2017 Jun 15;7(6):e015147 [FREE Full text] [doi: 10.1136/bmjopen-2016-015147]
59. Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. J Am Med Inform Assoc 2017 Dec 01;24(2):246-250. [doi: 10.1093/jamia/ocw154]

Abbreviations

IT: information technology
IHI: Institute for Healthcare Improvement
EHR: Electronic Health Records
PD: Primary Documents

Edited by G Eysenbach; submitted 28.02.18; peer-reviewed by K Blondon, WC Su, K Fuji, Y Gong; comments to author 30.03.18; revised version received 01.04.18; accepted 19.05.18; published 04.06.18.

Please cite as:
Feldman SS, Buchalter S, Hayes LW
Health Information Technology in Healthcare Quality and Patient Safety: Literature Review
JMIR Med Inform 2018;6(2):e10264
URL: http://medinform.jmir.org/2018/2/e10264/
doi:10.2196/10264
PMID:

Original Paper

The Impact of Implementation of a Clinically Integrated Problem-Based Neonatal Electronic Health Record on Documentation Metrics, Provider Satisfaction, and Hospital Reimbursement: A Quality Improvement Project

William Liu[1], MD; Thomas Walsh[2], BSN

[1]Neonatology, Golisano Children's Hospital of Southwest Florida, Lee Health, Fort Myers, FL, United States

[2]Information Systems, Lee Health, Fort Myers, FL, United States

Corresponding Author:
William Liu, MD
Neonatology
Golisano Children's Hospital of Southwest Florida
Lee Health
9981 South HealthPark Drive
Suite NICU, 3rd Floor
Fort Myers, FL, 33908
United States
Phone: 1 2393436684
Fax: 1 2393436280
Email: william.liu.md@leehealth.org

Abstract

Background: A goal of effective electronic health record provider documentation platforms is to provide an efficient, concise, and comprehensive notation system that will effectively reflect the clinical course, including the diagnoses, treatments, and interventions.

Objective: The aim is to fully redesign and standardize the provider documentation process, seeking improvement in documentation based on ongoing All Patient Refined Diagnosis Related Group–based coding records, while maintaining noninferiority comparing provider satisfaction to our existing documentation process. We estimated the fiscal impact of improved documentation based on changes in expected hospital payments.

Methods: Employing a multidisciplinary collaborative approach, we created an integrated clinical platform that captures data entry from the obstetrical suite, delivery room, neonatal intensive care unit (NICU) nursing and respiratory therapy staff. It provided the sole source for hospital provider documentation in the form of a history and physical exam, daily progress notes, and discharge summary. Health maintenance information, follow-up appointments, and running contemporaneous updated hospital course information have selected shared entry and common viewing by the NICU team. The interventions were to (1) improve provider awareness of appropriate documentation through a provider education handout and follow-up group discussion and (2) fully redesign and standardize the provider documentation process building from the native Epic-based software. The measures were (1) hospital coding department review of all NICU admissions and 3M All Patient Refined Diagnosis Related Group–based calculations of severity of illness, risk of mortality, and case mix index scores; (2) balancing measure: provider time utilization case study and survey; and (3) average expected hospital payment based on acuity-based clinical logic algorithm and payer mix.

Results: We compared preintervention (October 2015-October 2016) to postintervention (November 2016-May 2017) time periods and saw: (1) significant improvement in All Patient Refined Diagnosis Related Group–derived severity of illness, risk of mortality, and case mix index (monthly average severity of illness scores increased by 11.1%, *P*=.008; monthly average risk of mortality scores increased by 13.5%, *P*=.007; and monthly average case mix index scores increased by 7.7%, *P*=.009); (2) time study showed increased time to complete history and physical and progress notes and decreased time to complete discharge summary (history and physical exam: time allocation increased by 47%, *P*=.05; progress note: time allocation increased by 91%, *P*<.001; discharge summary: time allocation decreased by 41%, *P*=.03); (3) survey of all providers: overall there was positive provider perception of the new documentation process based on a survey of the provider group; (4) significantly increased hospital

average expected payments: comparing the preintervention and postintervention study periods, there was a US $14,020 per month per patient increase in average expected payment for hospital charges (P<.001). There was no difference in payer mix during this time period.

Conclusions: A problem-based NICU documentation electronic health record more effectively improves documentation without dissatisfaction by the participating providers and improves hospital estimations of All Patient Refined Diagnosis Related Group–based revenue.

(JMIR Med Inform 2018;6(2):e40) doi:10.2196/medinform.9776

KEYWORDS

electronic health record; neonatal intensive care unit; NICU; physician documentation; Epic; SOI; ROM; CMI; APR-DRG; informatics

Introduction

Wide-scale adoption of electronic health records (EHRs) became a national policy mandate in 2009, with allocation of significant health care dollars dependent on meaningful use implementation [1]. This has been justified by projected improvements in patient safety and health care quality [2].

The evidence for the benefits of EHR-based physician documentation is evolving. One challenge to implementation remains physician resistance, related to a myriad of operational and human factor barriers to creating the traditional physician medical note, including a perceived decrease in efficiency and an increased time expenditure [3,4].

The importance of physician documentation and the concept of problem-based documentation was originally championed by Lawrence Weed who recognized the importance of a systematic and comprehensive approach toward documenting the care of the complex intensive care patient with multisystem disease [5]. The translation of this complex process was accomplished primarily by handwritten or transcribed notes until the advent of the EHR. The EHR should efficiently collect, store, and display patient information in a way that will facilitate medical decision making and allow the provider to integrate this information, as reflected in his documentation [6].

All Patient Refined Diagnosis Related Groups (APR-DRGs) are the standard measure of provider medical documentation. Similar to the Centers for Medicare and Medicaid Medicare Severity Diagnosis Related Group inpatient prospective payment system, the APR-DRG provides a quantitative tool to measure accuracy and quality of physician documentation [7].

Our primary project goal was to fully redesign and standardize the provider documentation process, seeking improvement in documentation based on ongoing APR-DRG-based coding records, while maintaining noninferiority comparing provider satisfaction to our existing documentation process. We report the fiscal impact of improved documentation based on changes in expected APR-DRG-based hospital payments.

Methods

Background

In 2014-2015, all physician documentation at Lee Health in Fort Myers, FL, was transitioning from a dictation-based system to a full EHR platform using Epic software (Epic Systems Corporation, Verona, WI, USA). The Epic platform was already integrated into obstetrical and neonatal nursing, respiratory therapy, case management, pharmacy, laboratory, and imaging data entry, and the physician or provider component was a mandatory next step. Dictation-based documentation was the documentation method of choice. Between September 2015 and September 2016, the neonatal group transitioned to a combination of analog dictation with hospital-contracted transcriptionist documentation and a nonstandardized, individualized "out-of-box" Epic-based electronic documentation.

We began to redesign the provider documentation system in October 2015. We actively utilized Epic documentation capabilities as a "learning lab" for continuous improvement and refinement to achieve a final documentation system within Epic. From October 2015 to October of 2016, the providers utilized a shared electronic entry template, and some dictation continued as well. The shared template continued to use a clinical systems-based format for progress notes, and the entry was not problem-based. An ongoing feedback structure allowed providers to review benefits and drawbacks to note entry templates, smart phrases, general structure, and work flow. Real-time refinements based on this feedback allowed for a continuously evolving and improving documentation process.

In November 2016, the newly designed neonatal documentation system went live, and has continued through the study period of May 2017. Golisano Children's Hospital level 2 and 3 neonatal services were provided primarily at HealthPark Medical Center neonatal intensive care unit (NICU), with a smaller level 2 care facility at Cape Coral Hospital Special Care Nursery. We chose not to include any further data in our analysis because the hospital NICU moved to a new facility in mid-May 2017.

Aims and Goals

The aims and improvement goals of this project were to:

1. Improve provider awareness of appropriate documentation through a provider education handout and follow-up group discussion. We provided this education in August to September 2015, and an additional education with revision to the *International Classification of Diseases, Tenth Revision* (*ICD-10*) in January to February 2016.
2. Fully redesign and standardize the provider documentation process building from the native Epic-based software.

3. Create a comprehensive neonatal provider documentation system including the history and physical (H&P), progress note, and discharge summary that utilizes sharing and collaborative maternal and neonatal data entry by clinicians or staff in the obstetrical and neonatal work environments.
4. Improve provider care documentation as reflected by hospital 3M severity of illness (SOI), risk of mortality (ROM), and case mix index (CMI) scores. We did not have any target goals to increase diagnosis documentation, but rather sought to improve accuracy of documentation.
5. Achieve these goals without a negative perception of the new documentation process by the provider. This would be measured by time-based study by one provider and group survey after completion of the study.

Methodology

We utilized a problem-based entry capability built into the hospital-wide vendor software to create a physician data entry structure that sought to efficiently manage the large data streams that occur in the NICU, enhance consistent and comprehensive identification of *ICD-10*-based diagnoses, streamline the use of an ongoing clinical summary form that can benefit all care providers with reduced redundancy of data entry, generation of a facile discharge summary, and provide a problem-based daily progress note that would encourage entry similar to the traditional physician problem-based SOAP (subjective, objective, assessment, and plan) note.

This project is a collaborative effort of the medical director of the NICU (WFL), all the neonatal providers, and a key hospital information system programmer (TW) and the Coding Documentation Improvement staff. The preintervention period was October 2015 to October 2016. Problem-based entry was a key component of the platform being created, and could not be implemented until the full configuration was in place. The postintervention period was November 2016 to May 2017.

Our initial intervention was provider education on recommended documentation optimization. In August to September 2015, the provider staff were given an orientation in optimal documentation (NICU Physician Documentation Guidelines) and encouraged to provide optimal and accurate documentation of patient clinical diagnoses. In October to November 2015, the hospital converted to *ICD-10* and revised versions of these guidelines were provided to the providers in January to February 2016.

Our second intervention was to design the neonatal EHR during the preintervention period: October 2015 to October 2016. “Go live” was November 14, 2016. Figure 1 illustrates the flow process for construction of provider documentation.

The following changes to the neonatal documentation process occurred. Over a 1-year period, through an ongoing clinician and information technology (IT) department collaboration, a more efficient and integrated neonatal EHR was constructed. Objectives for creating the tools focused on creating efficient and low effort data entry, and attempted to eliminate redundancies in documentation. Our perspective was that the quality of note readability facilitates communication of patient medical status.

Ongoing PDCA (plan-do-check-act) cycles refined the documentation structure incrementally, with monthly physician and IT design sessions. The neonatal provider team simultaneously constructed a facsimile of the H&P exam, progress note, and discharge summary templates, using these for real-time documentation, as well as reviewing and improving on variations on documentation strategies. Epic Sandbox simulations were conducted as the project matured to allow for more realistic assessments of effectiveness, as well as identification of limitations and needed revisions.

The following are some examples of the PDCA process. We improved efficiencies in data collection and sharing by initial identification of discrete data fields and implementation (eg, perinatal factors). This included early identification of missing data fields or incomplete data entry in the obstetrical admission process. This was followed by discussion with obstetricians and nursing staff, revision of data entry structure in Epic, and new workflow to capture the necessary data. This allowed for more efficient and consistent data entry for all maternal, antenatal, intrapartum, delivery room factors while the mother was being evaluated, and subsequent facile and efficient electronic transfer of data into the neonatal medical record. This would subsequently allow for transfer to the new neonatal H&P.

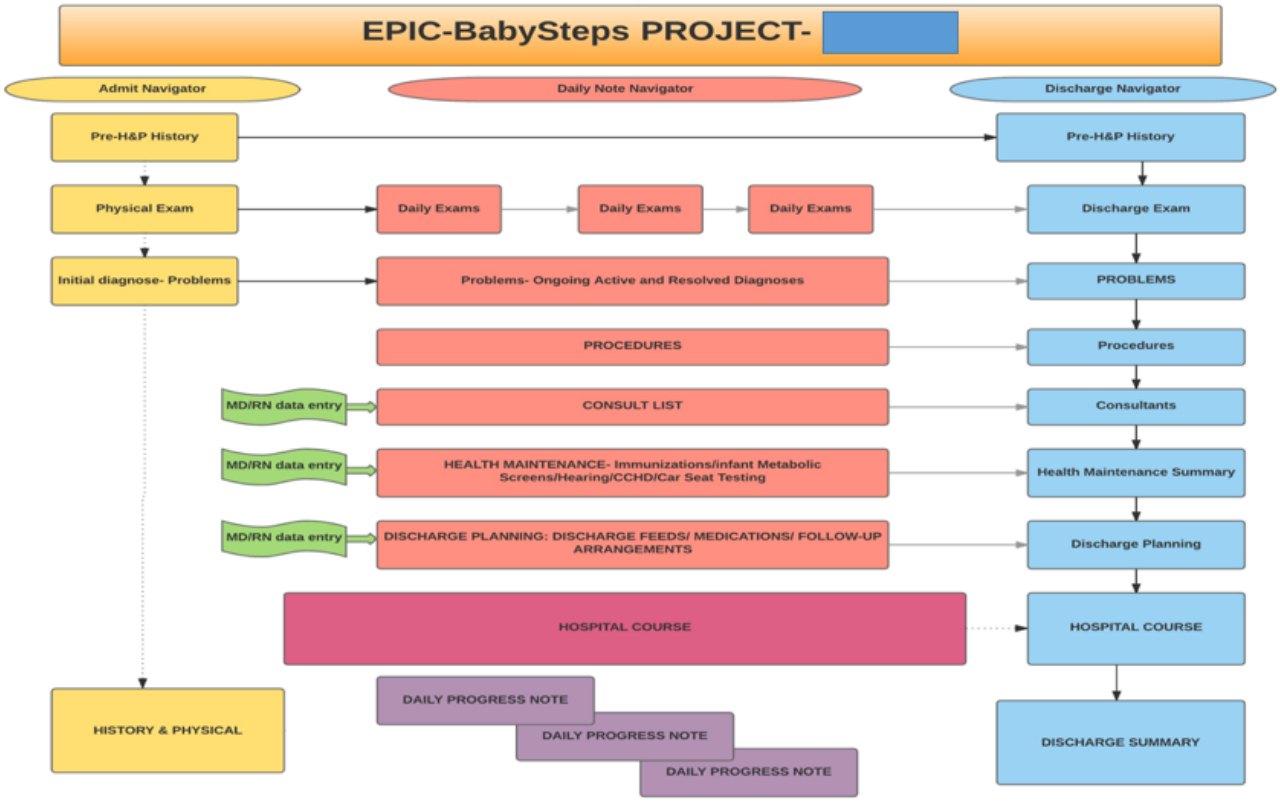
Another example of refinement was that redundant data collection was common to work flow with nursing and provider collection of health maintenance information (infant metabolic screening tests, critical congenital heart disease, pulse oximetry screening results, hearing test, car seat test, immunizations, physician follow-up appointments). As these needs became recognized in ongoing re-evaluations or study of implementation, IT added common entry fields that would be accessible and editable selectively by nursing and medical provider.

Our approach was to view constructing the discharge summary as beginning from birth. The discharge summary template was preconceived and a facsimile was utilized in the ongoing Epic documentation. We introduced the concept of a freestanding hospital course note that was created on admission. The provider entry into this hospital course note was intended to be incremental, with completion at the time of patient discharge. Other elements of clinical care relevant to discharge were preconfigured and completed during the hospitalization. Items such as hearing screen results, infant metabolic screens, and other routine testing were tracked and would be completed as the infant approached their discharge date.

We utilized problem-oriented charting. The transitional workflow encouraged providers to shift from system-oriented documentation to a habit of listing all relevant *ICD-10* diagnoses. We achieved efficient and consistent coding of diagnoses by identifying from the universe of *ICD-10* diagnoses, the most common neonatal diagnoses, which were included in a subset menu of common diagnoses.

The progress note template was created to allow for problem-based and system-bundled format for the progress note, maintaining the “SOAP” template familiar to providers.

Figure 1. Development overview of components of history and physical (H&P), progress note, and discharge summary navigators.



The interim documentation did not utilize the problem-based Epic functionality until the actual program was rolled out in November 2016. This moved compliance from a primarily human vigilance paradigm to a structure and process that facilitated easier documentation compliance. Prior to our "go live" date, IT coordinated several in-service sessions using the hospital computer laboratory for all providers on the final documentation system.

Measures and Analysis

The hospital clinical documentation improvement (CDI) team utilized 3M APR-DRG software, and a hospital-employed staff member reviewed each NICU admission, including all provider-generated neonatal patient documents (H&P, progress note, discharge summary, and consults) in the EHR from admission to discharge. This review was also inclusive of nursing notes, all flow sheets, laboratory values, and mother's EHR. The 3M 360 software has an embedded natural language processor that helps the reviewer to identify potential diagnoses and sequence in order of severity. These diagnoses were manually verified by CDI staff. This encoding process was updated and resequenced until discharge. Utilizing the APR-DRG codes, the CDIS clinical logic algorithm generated a prioritized list of codes and calculated the SOI, ROM, and CMI. These results were reviewed by a CDI team staff member.

When comparing the preintervention and postintervention study periods, we used independent metrics to evaluate for any confounding variables in clinical acuity that might impact on the coding scores. These included monthly average length of stay, NICU total patient days, monthly admissions by birthweight category, and average daily census.

Average expected payment derived from the hospital finance software utilizing the calculated APR-DRG, SOI, ROM, and CMI provided the basis for hospital payment calculations. The hospital has a customized payer mix filter with predetermined payment estimations, that factor in third-party and government payer mix with contracted payments, and calculates an average expected payment for each patient. These algorithms remained consistent during the entire study period.

We employed a time study in which one provider (WFL) tracked his total time to complete clinical documentation, parsed into the time period before initiation of the new documentation system and after: defined as preintervention and postintervention. The preintervention period reflected time duration for completion of clinical documentation utilizing the traditional voice-dictated note, transcribed by a hospital-contracted transcriptionist. The postintervention period was composed of a random collection of EHR documentations using the new documentation process.

Provider perception was measured using a survey methodology. There were 13 providers (six neonatologists and seven neonatal nurse practitioners) who provided continuous documentation during the preintervention and through the postintervention periods; one neonatal nurse practitioner began employment during the study and was excluded from the survey. In June to July 2017, a SurveyMonkey-based survey was conducted for the 12 eligible providers. The survey was comprised of questions addressing subjective assessments of the H&P, progress note, and discharge summary processes, as well as an overall assessment of the change.

Statistics

The clinical data for this study were obtained from the hospital Epic EHR utilizing an interface with Trendstar (The Shams Group) as well as clinical information compiled by an NICU-specific data analyst.

Trended data, using Minitab, was presented using statistical process control charting to visually present the impact of described interventions. Statistical process control uses entered data to describe common causes of variability, generating control limits, and identifying special causes of variation or statistically significant variance. Basic descriptive statistics were used, with *t* test for test of variance for continuous data using Microsoft Excel software or Minitab, with statistically significant variance defined as $P<.05$.

Results

Comparing the time periods October 2015 to October 2016 and November 2016 to May 2017, the monthly average SOI scores increased by 11.1% ($P=.008$; Figure 2), monthly average ROM scores increased by 13.5% ($P=.007$; Figure 3), and monthly average CMI scores increased by 7.7% ($P=.009$; Figure 4).

Figure 2. Severity of illness by pre-post neonatal electronic health record (EHR) implementation.

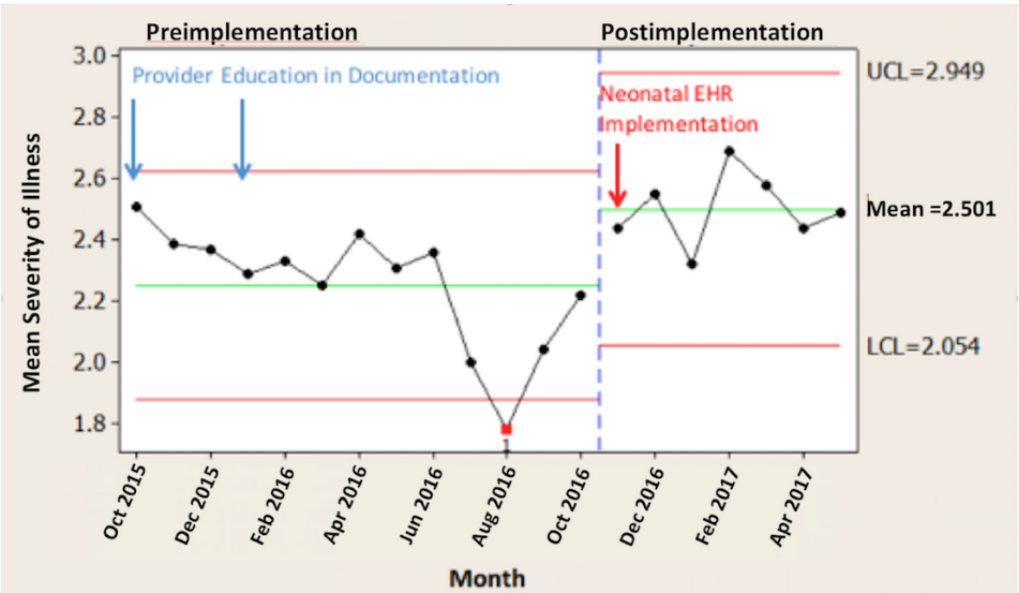


Figure 3. Risk of mortality by pre-post neonatal electronic health record (EHR) implementation.

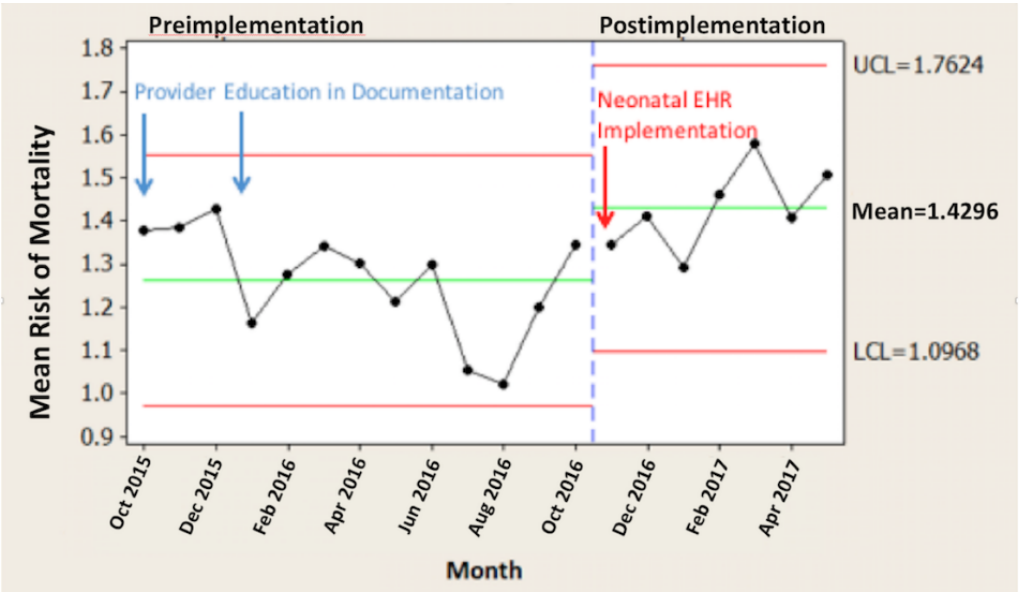


Figure 4. Case mix index by pre-post neonatal electronic health record (EHR) implementation.

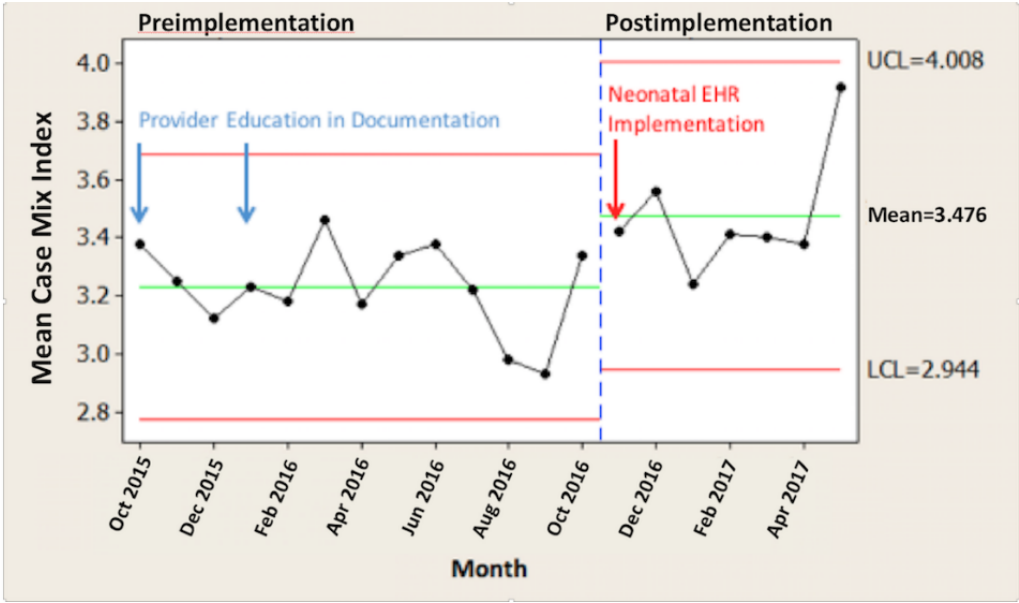


Table 1. Comparison of preintervention and postintervention groups in the neonatal intensive care unit (NICU) with potential confounders.

All values by month	NICU preintervention (Oct 2015-Oct 2016), mean (SD)	NICU postintervention (Nov 2016-May 2017), mean (SD)	*P* value
Severity of illness (SOI)	2.3 (0.2)	2.5 (0.1)	.008
Risk of mortality (ROM)	1.3 (0.1)	1.4 (0.1)	.007
Case mix index (CMI)	3.2 (0.2)	3.5 (0.2)	.009
Acuity indicators			
Average length of stay	19.2 (1.3)	21.1 (3.9)	.12
Total patient days	1329.9 (159.1)	1328.1 (116.6)	.98
≤1000 g	351.0 (58.7)	282.3 (61.2)	.02
1001-1500 g	189.6 (67.8)	237.0 (46.9)	.12
1501-2500 g	442.4 (111.8)	431.7 (84.8)	.83
>2500 g	346.9 (55.8)	377.1 (96.3)	.38
Admissions-total	73.4 (7.8)	68.4 (9.5)	.22
<1000 g	2.9 (2.4)	3.0 (1.2)	.94
1000-1499 g	3.4 (1.5)	4.40 (3.3)	.34
1500-1999 g	9.1 (2.9)	10.0 (2.9)	.51
2000-2499 g	14.9 (3.0)	11.0 (3.9)	.03
≥2500 g	43.2 (5.8)	40.0 (5.4)	.22
ADC[a] (Level 2 and Level 3 combined)	44.1 (4.1)	43.9 (3.5)	.88
Level 2	25.0 (2.2)	25.3 (3.0)	.79
Level 3	19.2 (4.3)	18.6 (2.7)	.75
CCH-SCN[b]	5.7 (1.7)	4.7 (1.3)	.19

[a]ADC: average daily census.

[b]CCH-SCN: Cape Coral Hospital Special Care Nursery.

There was no evidence of change in clinical acuity during the comparison time periods (Table 1). There was a slight increase in the mean length of stay in the postintervention group, but not statistically significant (*P*=.12). There was no difference in three of four weight categories for total patient days. There were significantly more total patient days for babies weighing 1000 grams or less in the preintervention group (preintervention: mean 351.0, SD 58.7; postintervention: mean 282.3, SD 61.2;

P=.02). There was no difference in four of five weight categories for total admissions per month. There were significantly more admissions for babies weighing between 2000 and 2499 grams in the preintervention group (preintervention: mean 14.9, SD 3.0; postintervention: mean 11.0, SD 3.9; *P*=.03). There was no difference in combined, level 3, or level 2 mean daily census.

We assessed the impact on providers through a time expenditure study and a provider survey. One provider tracked his own time expenditure during the course of the project (Figure 5; Table 2). The H&P time allocation increased by 47% (*P*=.05). The progress note time allocation significantly increased by 91% (*P*<.001). The discharge summary time allocation significantly decreased by 41% (*P*=.03).

SurveyMonkey was used to poll all neonatal providers. Of 12 eligible providers, there was a 100% response to the survey, which was obtained in July 2017. Questions asked if the new process, as compared to the old process was... (based on a Likert scale) 1=much worse; 2=somewhat worse; 3=about the same; 4=better; 5=much better. Tables 3 and 4 summarize the results and are reported as percentage answering "somewhat worse" and percentage answering "better" or "much better." No respondents answered "much worse" for any of the questions.

Hospital reimbursement was significantly improved. Comparing the preintervention and postintervention study periods, there was a US $14,020 per month per patient increase in average expected payment for hospital charges (*P*<.001; Figure 6). There was no difference in payer mix during this time period (Table 5).

Figure 5. Time to complete history and physical (H&P) notes, levels 2 and 3 progress notes, and discharge summaries preintervention (dictation) and postintervention (electronic health record).

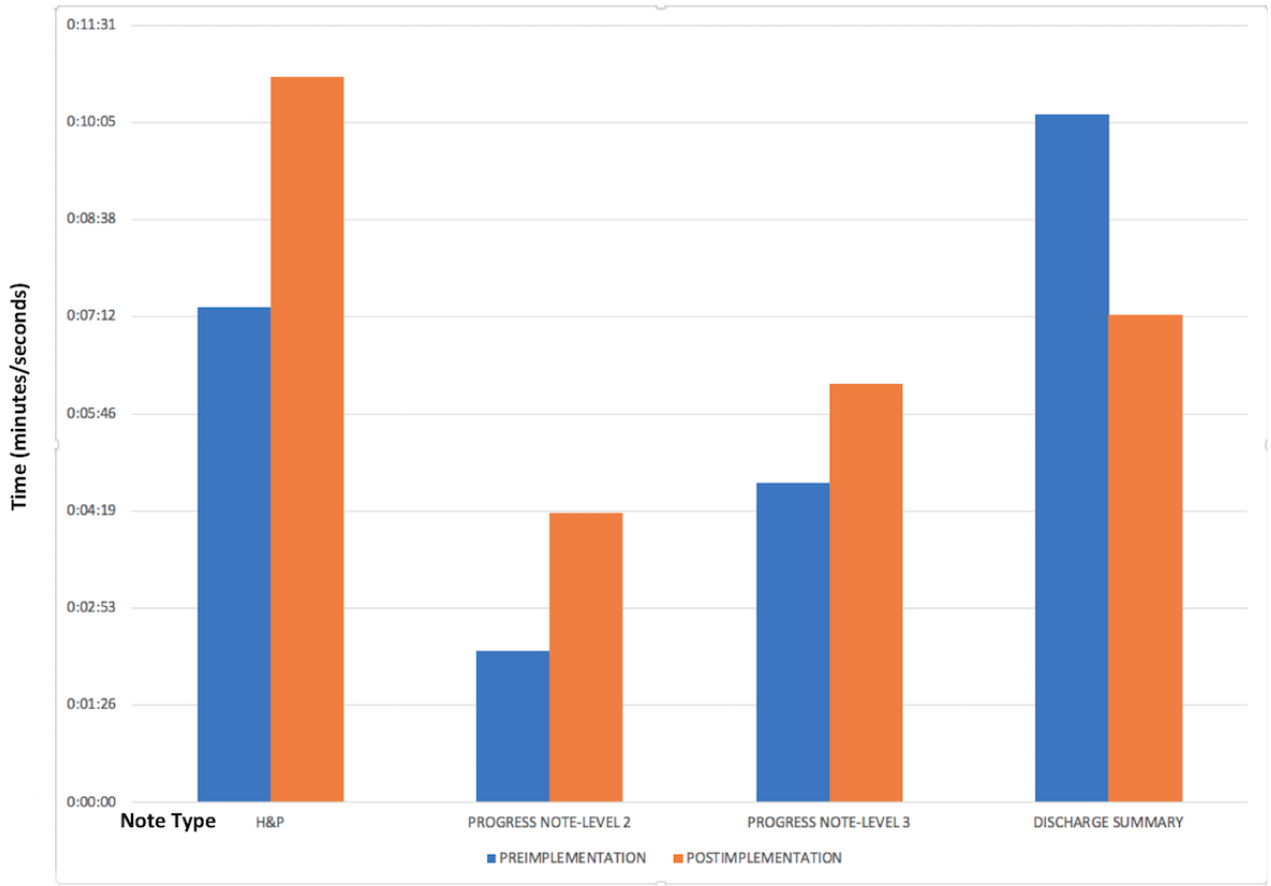


Table 2. Documentation time: preintervention (primarily dictation method) compared to postintervention (electronic health record).

Note type	Preintervention		Postintervention		*P* value
	n	Time (min:sec)	n	Time (min:sec)	
History and physical (H&P)	12	7:20	10	10:45	.05
Progress note	31	2:15	24	4:18	<.001
Discharge summary	15	10:12	31	7:14	.03

Table 3. Survey of providers: summary (n=12).[a]

Survey question	History and physical (H&P), n (%)		Progress note, n (%)		Discharge summary, n (%)		Overall, n (%)	
	Somewhat worse	Better or much better	Somewhat worse	Better or much better	Somewhat worse	Better or much better	Somewhat worse	Better or much better
Ability to be comprehensive in my documentation	0 (0)	11 (92)	0 (0)	10 (83)	1 (8)	10 (83)	1 (3)	31 (86)
Ability to customize my documentation	0 (0)	11 (92)	0 (0)	11 (92)	0 (0)	10 (83)	0 (0)	32 (8)
Time allotment to accomplish this documentation	3 (25)	8 (67)	3 (25)	5 (42)	2 (17)	8 (67)	8 (22)	21 (58)

[a]Worse represents “somewhat worse” (no respondents answered “much worse” for any of the questions); better represents “better” and “much better.”

Table 4. Overall impression of providers (n=12).[a]

Overall impression	Somewhat worse, n (%)	Better or much better, n (%)
The overall documentation experience is...	0 (0)	11 (92)
My overall efficiency with documentation is...	2 (17)	7 (58)
My documentation accuracy (note reflects the true event) and validity (note states what I intended) is...	0 (0)	10 (83)
Documentation has made staff efficiency to collect information from multiple sources...	1 (8)	10 (83)
Documentation system has made the safety of patient care in the NICU[b]...	0 (0)	11 (92)

[a]Worse represents “somewhat worse” (no respondents answered “much worse” for any of the questions); better represents “better” and “much better.”
[b]NICU: neonatal intensive care unit.

Figure 6. Trended average expected payment.

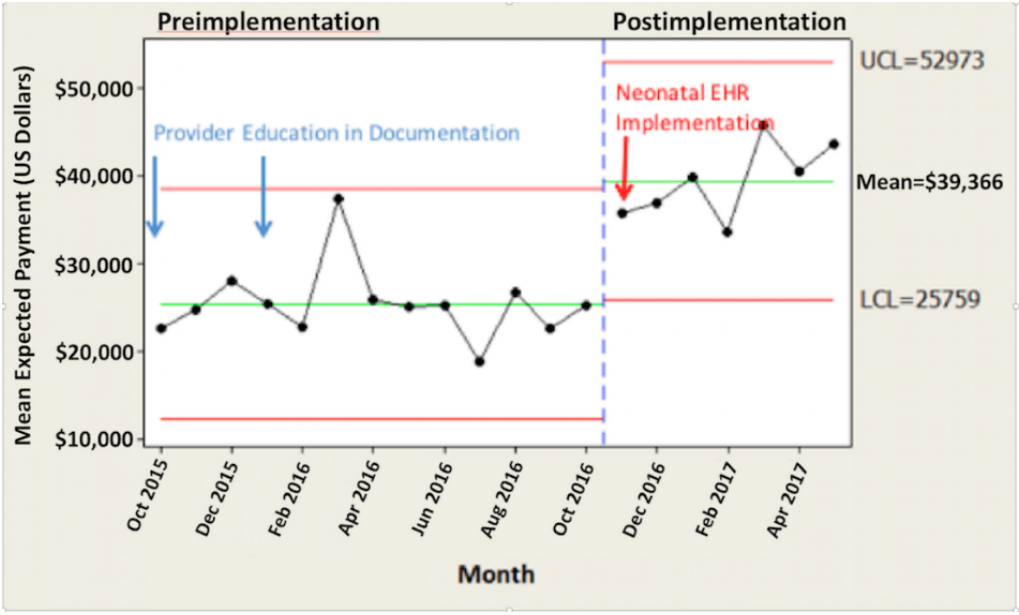


Table 5. Payer mix comparison: preintervention and postintervention periods.

Differential payment estimations	Preintervention	Postintervention	*P* value
Average expected payment (US $), mean (SD)	25,346 (4291)	39,366 (4327)	<.001
Payer mix, %			
Medicaid	71.6	72.2	.94
Third party	26.4	26.3	.99
Self-pay	0.8	0.4	.79

Discussion

A problem-based NICU documentation EHR effectively improves documentation while avoiding dissatisfaction by the participating providers, and improves hospital estimations of APR-DRG-based revenue.

Addressing Challenges to Neonatal Intensive Care Unit Documentation

The provider in the neonatal intensive care setting is challenged with a need to rapidly consolidate and integrate data that is being accumulated from several sources. The challenges to this in the NICU setting have been well described [8,9].

The NICU hospital course reflects a continuum of care with clinically relevant data collection beginning with a body of historical information that resides in the maternal (obstetrical) history: the woman's preexisting medical conditions, her ongoing antepartum care, and the acute management of the delivering mother during the intrapartum care, which overlaps her admission and subsequent delivery. With the delivery of the newborn, the EHR should be populated with accessible information reflecting the newborn's delivery room management and subsequent care in the NICU. This clinical course is composed of *ICD-10* and text-based diagnoses, assessments, and management interventions from not only the neonatal provider and specialist consultants, but also data entry by the NICU nurses, respiratory therapists, pharmacists, occupational therapists, dieticians, social workers, and case managers working with the patient's family as well as a plethora of laboratory testing and imaging study results. The provider routinely assesses, consolidates, and integrates this information into a diagnosis and care plan.

We maintained an awareness of the original goals in a physician note. In 1968, Dr Weed [5] identified the importance of the problem-based note and of organizing clinical data. He outlined the characteristics of effective physician documentation and the need for annotation of active and resolved problem lists. The need to adhere to these principles is especially relevant when dealing with the intensive care patient with multisystem disease.

In October 2015, transcription services were still being used, as well as early participation in electronic entry of notes. The dictation system was a familiar tool that allowed the provider to generate a progress note rapidly, but the information was not digitally accessible. Our early electronic note entry remained essentially a word document without any structured identification of discrete data items. In both cases, the note structure, content, and identification of diagnoses were recorded inconsistently, with a high variation in structure among providers, and did not allow for data tracking or a linkage to a clinical data warehouse. There also was an inefficient collection and sharing of perinatal risk factors and clinical information with other NICU staff members due to the silo-based nature of documentation, with redundant workflows to acquire necessary information. A patient's hospital course information was not easily accessible, and organization and completion of a discharge summary consequently was an inefficient and time-consuming task.

During the study period, we addressed these concerns with the following deliverables: (1) problem-oriented charting for documentation—Epic's suite of documentation tools can be customized to meet complex patient workflows while also facilitating discrete data capture; (2) standard documentation templates for major neonatal diagnoses; (3) standard documentation templates for H&P, daily progress note, and discharge; (4) a suite of neonatal documentation smart links that encourage problem-oriented charting and optimize physician workflow efficiency; (4) the problem list at admission front loaded with gestational age-based problem recommendations giving the physician a quick and easy way to add multiple problem selections; (5) a problem preference list inclusive of the most common neonatal diagnoses; (6) a neonatal handoff to facilitate to-do list management and ease of provider handoffs of patient information; and (7) a structured course of care to manage pertinent historical data and to facilitate the expedited production of the discharge summary.

Documentation Metrics

Illness classification systems function as a predictive model for federal resource allocation and to help track clinical outcomes. The APR-DRGs were developed in 1990 by 3M Health Information Systems jointly with the National Association of Children's Hospitals and Related Institutions as the most comprehensive pediatric logic of any severity illness classification system, and were most recently updated to *ICD-10-CM*. By design, the APR-DRG system reflects completeness, accuracy, and specificity of documentation [7]. The EHR has been used to improve physician progress note documentation with documented improvement in *ICD-9* codes [10], and a provider-targeted educational intervention has been demonstrated to improve documentation as reflected in DRG-based coding metrics [11]. Although we were unable to see any improvement in our patient SOI, ROM, or CMI scores with our educational intervention, converting to a problem-based software platform within our hospital EHR resulted in significant improvement in documentation, without any demonstrable

change in clinical acuity between the study periods. The increase in total patient days and a larger number of admissions of newborns weighing less than 1000 grams in the preintervention period would only skew toward less difference between the groups, suggesting a conservative appraisal of our improved documentation. Our findings suggest that strategies that are dependent on human diligence are much less effective than EHR-based process efficiencies.

Provider Time Expenditure and Satisfaction

Physicians generally perceive that the EHR improves documentation, although many concerns are expressed [12]. Before initiation of this project, there was preexisting provider bias based on prior EHR experiences. These included a concern that there would be excessive time used for data entry and note production, and decreased productivity. An important goal was to decrease the time burden in generating a discharge summary (the NICU discharge summary often involves a 1-3 month length of stay requiring an extensive investment in time to review the hospital course, collect appropriate information, and then transcribe the summary). Our PDCA approach utilized continuous feedback from the end- user, allowing us to refine our templates and note design to better address these concerns.

In estimation of time expenditure, we recognized that many confounders existed in a time study involving more than one provider. After implementation of the EHR, each provider had their own unique approach toward document preparation and note entry. This degree of provider-specific variation did not lend itself well to comparisons among different providers. A single provider case study allowed for consistent measurement, adequate sampling, and provided qualitative insight into changes in time expenditure before and after implementation. These results cannot be taken as anything more than the impact on one individual provider. It did appear that in this one sample, the time required for producing the H&P and progress notes was increased, whereas the time for generating a comprehensive discharge summary was significantly decreased. We obtained overall insight by surveying the entire provider group.

Survey methodology is limited by design; however, in the context of our project, allowed for a systematic way to understand the provider experience. Our results are highly valid, reflecting the entire provider population that experienced the documentation process. In the context of a survey design, we were looking only for noninferiority or a perceptual equivalence, which is a more cautious approach to interpretation of our results. Complementing our case study of time allotment, the survey did not demonstrate a provider perception of increased time allotment for documentation. In fact, no participants chose the “much worse” option on the Likert scales, and there was generally a favorable response on all questions, including the overall documentation process, documentation efficiency, accuracy and validity, ability for staff to obtain and report clinical data, and overall perception of improved NICU safety. Our provider group adopted and adapted well to the new documentation process, without any evidence that the new process was worse, suggesting a better process than the one it replaced. We believe that early and ongoing provider engagement in the development process played a large role in ensuring greater provider satisfaction.

Reimbursement

Projected EHR costs have remained an ongoing barrier to implementation, despite the emerging body of evidence that implementation of EHRs may lead to improved health outcomes with decreased medical errors and improved disease management [13,14].

There has been some evidence that there is a positive return on investment in adopting EHR in an ambulatory setting [15]. A challenge to this is that although adoption of EHRs may digitalize and standardize many of the clinical processes, they also impact on revenue-cycle functions. Poor awareness of this may have a negative impact on an organizations cash flow. Accurate and comprehensive capture of the clinical encounter is an important feature of an effective EHR, and reflects an effective merging of clinical and revenue-cycle operations with information technology [16].

Our documentation process focused not only on accurate and facile documentation of appropriate *ICD-10* diagnosis codes linked to concurrent supporting clinical elements, but on development, design, and entry interfaces based on a tight collaboration between the information systems programmer and the clinicians.

Clinical quality has become a major driver since the 2001 Institute of Medicine’s landmark publication “Crossing the Quality Chasm: A New Health System for the 21st Century” [17]. Concurrently, health care financing continues to move from traditional fee-for-service models to a more pay-for-performance, outcome-based reimbursement. Along these lines, hospital-based EHRs must provide the ability to effectively and efficiently document and track defined clinical interventions and outcomes on an individual and aggregated basis.

There is no doubt that the emergence of the EHR is transforming the way health care is delivered. The implementation of this change is taking place in the face of perceived provider dissatisfaction, decreased productivity, and uncertainty at the corporate level of the return on investment. The next step in health care reform involves improving value (as a function of outcomes divided by cost), especially as we are driven to improve clinical outcomes in the face of increasing fiscal accountability [18].

Of interest, is that this project was primarily focused on physician efficiency and documentation quality. The financials were never a primary goal. Our findings suggest that with a focus on high-quality care delivery, appropriate reimbursement gains will follow.

Conclusions

Our project has demonstrated the clinical and fiscal effectiveness of a collaborative effort to create a more effective documentation system. There is clear noninferiority to our prior documentation process with respect to overall efficiency and a suggestion of an improved overall experience, as well as improved patient safety based on provider perception. We demonstrated that

improved clinical documentation may also lead to improved hospital revenues, and clearly extends the dialog on the role of providers in addressing value-based care.

Acknowledgments

This work is primarily the responsibility of the authors. Our thanks to the participants in this project for their ongoing input and suggestions: Mohamed Faisal, MD; Beatriz de Jongh, MD; William Binder, MD; Mai Phung, MD; Amit Mukhia, MD; Marilyn Farley, ARNP; Lisa Ciambrello, ARNP; Jodi Kurtz, ARNP; April Felton, ARNP; Elaine Pao, ARNP; Cindy Lucas, ARNP; and to David Kanter, MD, for his input on medical coding, to Lee Health Decision Support department and to Kristi Repetto, and the Lee Health CDI department for their chart review.

Conflicts of Interest

None declared.

References

1. Sebelius K. Federal Register. 2009 Oct 09. Department of Health and Human Services: Health Information Technology for Economic and Clinical Health Act URL: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/enforcementrule/enfifr.pdf [accessed 2017-12-13] [WebCite Cache ID 6vgcxuGSL]
2. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [Medline: 16702590]
3. Meinert DB. Resistance to electronic medical records (EMRs): a barrier to improved quality of care. IISIT 2005;2:494-504. [doi: 10.28945/846]
4. Ajami S, Bagheri-Tadi T. Barriers for adopting electronic health records (EHRs) by physicians. Acta Inform Med 2013;21(2):129-134 [FREE Full text] [doi: 10.5455/aim.2013.21.129-134] [Medline: 24058254]
5. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 14;278(11):593-600. [doi: 10.1056/NEJM196803142781105] [Medline: 5637758]
6. Kuhn T, Basch P, Barr M, Yackel T, Medical Informatics Committee of the American College of Physicians. Clinical documentation in the 21st century: executive summary of a policy position paper from the American College of Physicians. Ann Intern Med 2015 Feb 17;162(4):301-303. [doi: 10.7326/M14-2128] [Medline: 25581028]
7. Averill RF, Goldfield N, Hughes J, Bonazelli J, McCullough E, Steinbeck B, et al. All Patient Refined Diagnosis Related Groups (APR-DRGs): Methodology Overview. Wallingford, CT: 3M Health Information Systems; 2003. URL: https://www.hcup-us.ahrq.gov/db/nation/nis/APR-DRGsV20MethodologyOverviewandBibliography.pdf [accessed 2017-12-13] [WebCite Cache ID 6vgebZcOk]
8. Palma JP, Brown PJ, Lehmann CU, Longhurst CA. Neonatal informatics: optimizing clinical data entry and display. Neoreviews 2012 Feb;13(2):81-85 [FREE Full text] [Medline: 22557935]
9. Ellsworth MA, Lang TR, Pickering BW, Herasevich V. Clinical data needs in the neonatal intensive care unit electronic medical record. BMC Med Inform Decis Mak 2014 Oct 24;14:92 [FREE Full text] [doi: 10.1186/1472-6947-14-92] [Medline: 25341847]
10. Grogan EL, Speroff T, Deppen SA, Roumie CL, Elasy TA, Dittus RS, et al. Improving documentation of patient acuity level using a progress note template. J Am Coll Surg 2004 Sep;199(3):468-475. [doi: 10.1016/j.jamcollsurg.2004.05.254] [Medline: 15325618]
11. Rosenbaum BP, Lorenz RR, Luther RB, Knowles-Ward L, Kelly DL, Weil RJ. Improving and measuring inpatient documentation of medical care within the MS-DRG system: education, monitoring, and normalized case mix index. Perspect Health Inf Manag 2014;11:1c [FREE Full text] [Medline: 25214820]
12. Chisolm DJ, Purnell TS, Cohen DM, McAlearney AS. Clinician perceptions of an electronic medical record during the first year of implementaton in emergency services. Pediatr Emerg Care 2010 Feb;26(2):107-110 [FREE Full text] [doi: 10.1097/PEC.0b013e3181ce2f99] [Medline: 20093997]
13. Hayes TO. American Action Forum. 2015 Aug 06. Are electronic medical records worth the cost of implementation? URL: https://www.americanactionforum.org/research/are-electronic-medical-records-worth-the-costs-of-implementation/ [accessed 2017-12-13] [WebCite Cache ID 6vgI47ZK1]
14. Kian LA, Stewart MW, Bagby C, Robertson J. Justifying the cost of a computer-based patient record. Healthc Financ Manage 1995 Jul;49(7):58-60, 62, 64. [Medline: 10143674]
15. Grieger DL, Cohen SH, Krusch DA. A pilot study to document the return on investment for implementing an ambulatory electronic health record at an academic medical center. J Am Coll Surg 2007 Jul;205(1):89-96. [doi: 10.1016/j.jamcollsurg.2007.02.074] [Medline: 17617337]

16. Goldsmith J, McKesson E. Modern Healthcare. 2016 Jul 15. Commentary: finding the hidden ROI in EHR implementation URL: http://www.modernhealthcare.com/article/20160715/NEWS/160719932 [accessed 2017-12-13] [WebCite Cache ID 6vgOgw0QJ]
17. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press (US); 2001.
18. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]

Abbreviations

APR-DRG: All Patient Refined Diagnosis Related Group
CDI: clinical documentation improvement
CMI: case mix index
EHR: electronic health record
H&P: history and physical
ICD-10: International Classification of Diseases, Tenth Revision
IT: information technology
NICU: neonatal intensive care unit
PDCA: plan-do-check-act
ROM: risk of mortality
SOAP: subjective, objective, assessment, and plan
SOI: severity of illness

Edited by G Eysenbach; submitted 08.01.18; peer-reviewed by C McGregor, P Woei Bing, I Mircheva; comments to author 28.02.18; revised version received 02.03.18; accepted 07.03.18; published 20.06.18.

Please cite as:
Liu W, Walsh T
The Impact of Implementation of a Clinically Integrated Problem-Based Neonatal Electronic Health Record on Documentation Metrics, Provider Satisfaction, and Hospital Reimbursement: A Quality Improvement Project
JMIR Med Inform 2018;6(2):e40
URL: http://medinform.jmir.org/2018/2/e40/
doi:10.2196/medinform.9776
PMID:29925495

Original Paper

Implementing an Open Source Electronic Health Record System in Kenyan Health Care Facilities: Case Study

Naomi Muinga[1], MSc; Steve Magare[1], MSc; Jonathan Monda[1], MBChB, MSc; Onesmus Kamau[2], MSc; Stuart Houston[3], BSc (Comp Sci); Hamish Fraser[4], MBChB, MSc; John Powell[5], MBBChir, PhD, FFPH; Mike English[1,6], MD, MB BChir, FMedSci; Chris Paton[7], BMBS, MBA

[1]KEMRI/Wellcome Trust Research Programme, Nairobi, Kenya

[2]e-Health and Systems Development Unit, Ministry of Health, Nairobi, Kenya

[3]Vimak Company Limited, Nairobi, Kenya

[4]Brown Center for Biomedical Informatics, Brown University, Providence, RI, United States

[5]Nuffield Department of of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[6]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

[7]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Naomi Muinga, MSc
KEMRI/Wellcome Trust Research Programme
PO Box 43640
Nairobi, 00100
Kenya
Phone: 254 730162000
Email: nmuinga@kemri-wellcome.org

Abstract

Background: The Kenyan government, working with international partners and local organizations, has developed an eHealth strategy, specified standards, and guidelines for electronic health record adoption in public hospitals and implemented two major health information technology projects: District Health Information Software Version 2, for collating national health care indicators and a rollout of the KenyaEMR and International Quality Care Health Management Information Systems, for managing 600 HIV clinics across the country. Following these projects, a modified version of the Open Medical Record System electronic health record was specified and developed to fulfill the clinical and administrative requirements of health care facilities operated by devolved counties in Kenya and to automate the process of collating health care indicators and entering them into the District Health Information Software Version 2 system.

Objective: We aimed to present a descriptive case study of the implementation of an open source electronic health record system in public health care facilities in Kenya.

Methods: We conducted a landscape review of existing literature concerning eHealth policies and electronic health record development in Kenya. Following initial discussions with the Ministry of Health, the World Health Organization, and implementing partners, we conducted a series of visits to implementing sites to conduct semistructured individual interviews and group discussions with stakeholders to produce a historical case study of the implementation.

Results: This case study describes how consultants based in Kenya, working with developers in India and project stakeholders, implemented the new system into several public hospitals in a county in rural Kenya. The implementation process included upgrading the hospital information technology infrastructure, training users, and attempting to garner administrative and clinical buy-in for adoption of the system. The initial deployment was ultimately scaled back due to a complex mix of sociotechnical and administrative issues. Learning from these early challenges, the system is now being redesigned and prepared for deployment in 6 new counties across Kenya.

Conclusions: Implementing electronic health record systems is a challenging process in high-income settings. In low-income settings, such as Kenya, open source software may offer some respite from the high costs of software licensing, but the familiar challenges of clinical and administration buy-in, the need to adequately train users, and the need for the provision of ongoing technical support are common across the North-South divide. Strategies such as creating local support teams, using local development resources, ensuring end user buy-in, and rolling out in smaller facilities before larger hospitals are being incorporated

into the project. These are positive developments to help maintain momentum as the project continues. Further integration with existing open source communities could help ongoing development and implementations of the project. We hope this case study will provide some lessons and guidance for other challenging implementations of electronic health record systems as they continue across Africa.

(JMIR Med Inform 2018;6(2):e22) doi:10.2196/medinform.8403

KEYWORDS

electronic health records; software; medical records; Kenya; open source

Introduction

Background

A major driver of the increased use of electronic health record (EHR) systems in recent years has been the belief that these systems can support the provision of high-quality care [1,2]. Features such as a clinical decision support system can play a role in reducing medical errors by providing point-of-care information to support decision making by alerting a doctor to drug interactions when they create an electronic prescription [3]. More recently, EHRs have been proposed as the digital infrastructure to support learning health systems that enable continuous improvement through a cycle of EHR data analysis and quality improvement interventions [4-6].

In high-income countries, EHR adoption has been fostered by government incentive schemes such as the Health Information Technology for Economic and Clinical Health Act of 2009 in the United States through which health care providers have been compensated for the costs of information technology (IT) systems if they were able to demonstrate that the systems were used to improve care or increase efficiencies—so-called "Meaningful Use" [7,8].

Low-income countries, despite facing the challenges of resource constraints, inadequate data collection systems, the lack of incentives to collect health information, and inadequately trained personnel [9], have seen the increased use of EHR systems through aid-funded projects linked to specific diseases [1,10]. For example, in Kenya, EHRs have been used within projects that mainly support HIV care leading to well-developed systems for this disease area. For the management of both HIV and tuberculosis (TB), the result of digitization has been better record-keeping, patient management, follow-up, and stock control [11-14]. Although these implementations were largely successful, challenges encountered included limited interoperability with other systems and a lack of direct use by clinicians—systems are often used by clerks who enter data on behalf of the clinical team [15].

In the light of the perceived success of these disease-focused clinic-based systems, the Kenyan Ministry of Health (MOH) has begun to adapt one of the main systems (Open Medical Record System, OpenMRS) for use in public health facilities. This case study describes the current eHealth policies and guidance of the Kenya MOH and identifies the lessons learned from the initial development and implementation of this new OpenMRS-based system called Afya Electronic Health Management System (AfyaEHMS).

Government eHealth Policy, Projects, and Guidance

Health Management Information Systems, Centers for Disease Control and Prevention, and National AIDS and STI Control Programme Electronic Medical Records Reviews (2007-2009)

Several assessments of systems used to manage patient and health data in Kenya (reporting systems and EHRs) were carried out between November 2007 and July 2009 by the Health Management Information Systems department (HMIS in MOH), the US Centers for Disease Control and Prevention (CDC), and the National AIDS and STIs Control Programme (NASCOP) [15]. The narrative synthesis of the findings of the 3 reviews highlighted a number of challenges encountered in previous EHR implementations. Specific challenges identified included varying data security functionality, unreliable vendor support, sustainability issues, variable reporting functionality, limited feedback for patient care, and limited ability to exchange information between systems [16]. Key benefits identified included HIV care systems that were highly developed and that were efficiently handling antiretroviral therapy care data.

From these assessments by HMIS, CDC, and NASCOP, recommendations were made regarding the way forward toward the scale up and harmonization of data systems for health services to improve patient care, facility and resource management, and policy development and evaluation [15].

Electronic Medical Records Standards and Guidelines Report (2010)

The recommendations from the HMIS, CDC, and NASCOP reviews then formed the basis of an "Electronic Medical Records Standards and Guidelines" (ESG) document for Kenya [17] in 2010. The aim of this document was to ensure quality of software, compatibility of data sharing, ease of maintenance, and common understanding among the workforce. The ESG document was designed to offer guidelines to the minimum standard for generic electronic systems in the health care setting for electronic medical record (EMR) system developers, implementers, and those contemplating the use of EMR systems. The guidelines covered the sections mentioned in Table 1.

Table 1. Sections covered in the Electronic Medical Records Standards and Guidelines (ESG) document. EMR: electronic medical record.

Section	Description	Target
EMR development	• Outlines prerequisite processes of EMR development • Identifies basic functional requirements for EMRs • Identifies software attributes needed to ensure quality data and system security	EMR developers
EMR interoperability	• Recommends that EMR systems can transmit and receive a minimum dataset via Health Level 7 messaging • Recommends that systems have the capability to transmit aggregate data to District Health Information Software Version 2 via Statistical Data and Metadata eXchange for the Health Domain, in short SDMX.HD, messaging	EMR developers; program managers
EMR implementation	• Outlines conditions to be met for successful EMR implementation	EMR implementers; program managers

Kenya Electronic Medical Records Review Toward Standardization (2011)

In 2011, a review of 17 EMR systems implemented in Kenya was carried out to assess the progress made toward standardization comparing the recommendations of the ESG document against the actual state of EMR use in health care facilities selected for review [16]. The review scored systems according to 7 functional areas: system details and standards; basic demographic and clinical health information; order entry and prescription; clinical decision support; health information and reporting; security and confidentiality; and exchange of electronic information. The results showed a wide variation of the capabilities of the different systems, variation in the adoption of functionalities of the same systems in different facilities, and variation in the overall adoption and use of systems across different facilities [16].

Of the systems reviewed, the EMR systems with the highest weighted scores over the 7 functional areas were OpenMRS AMPATH, IQ Care, and C-PAD at 95.2%, 90.3%, and 77.1%, respectively; these were systems used for HIV patient care [18].

Kenya National eHealth Strategy (2011-2017)

A Kenya National eHealth Strategy was developed in 2010, with an aim to harness information and communication technologies (ICT) for improved health care delivery by supporting informed policy, improving access to clinical evidence for care providers, fostering interoperability, and creating linkages between service providers and researchers [19]. The strategy outlines 5 key areas: telemedicine, health information systems, information for citizens, mHealth, and e-learning. The health information systems pillar was prioritized and divided into 5 functional domains: patient centric information; pharmacy and medical supply chain information; financial information; health workforce management; and training and regulation.

The strategy identified 6 principles that are key factors in its implementation: strong leadership and governance through a proposed National eHealth Steering committee; formation of partnerships for shared information and services among stakeholders; leveraging available resources (human, financial, and technical); safeguarding privacy and security; harmonization of disparate expertise (health and technological); phased implementation of prioritized initiatives; and ensuring redundancy in mission-critical aspects of eHealth systems.

District Health Information Software Version 2 (2011)

Kenya has adopted, at a national level, the District Health Information Software Version 2 (DHIS2) for aggregating health data across different levels of the health system. The DHIS2 system was implemented as a response to challenges with the previous Microsoft Excel file-based system. These included an inability to fully analyze the data collected due to the way the data were aggregated, a lack of error-checking capabilities, incomplete data, and limited capacity in the use of information for decision making [20]. DHIS2 offers several advantages: it is free and open source (licensed under the new Berkeley Software Distribution license), it allows for data collection and use at different levels of the health system, it has a Web-based interface allowing for access using several devices, and it has a good network of support from worldwide users [21]. Data are entered into the system by health records officers who are responsible for data management at the facility or county level. DHIS2 was implemented through the support of development partners and consultants from the University of Oslo, Norway, after extensive stakeholder consultations [22].

KenyaEMR (Open Medical Record System; 2012-2013)

KenyaEMR [23] is a tailored distribution of Open Medical Record System (OpenMRS), an open source EHR system that has been widely used in several African countries to support the management of HIV/AIDS patients (and more recently other diseases such as TB and noncommunicable diseases). OpenMRS was developed to provide a core system and range of plug-in modules from which clinical health information systems could be created to allow flexibility to include or exclude particular modules depending on the needs of the health care facilities where the software was to be installed [24]. The KenyaEMR system was designed to meet the requirements laid out in the ESG report and has now been implemented in over 300 facilities in 4 geographic regions in Kenya, with support from the International Training and Education Center for Health (ITECH Kenya) of the University of Washington [25]. ITECH Kenya also supports the use of the system through extensive capacity building through facility-based champion mentors.

Table 2. Summary of reports and projects deployed. HMIS: Health Management Information Systems; CDC: Centers for Disease Control and Prevention; NASCOP: National AIDS and STI Control Programme; EMR: electronic medical record; DHIS2: District Health Information Software Version 2; IQCare: International Quality Care; AfyaEHMS: Afya Electronic Health Management System.

Reports and projects	2007-2009	2010	2011	2012-2013	2014-2017
Reports	HMIS, CDC, and NASCOP EMR Evaluations	EMR Standards and Guidelines Report	EMR Review Toward Standardization; Kenya National eHealth Strategy (2011-2017)		
Deployments			DHIS2 Rollout	KenyaEMR Rollout; IQCare Rollout	AfyaEHMS Rollout

International Quality Care (2012-2013)

International Quality Care (IQCare) [26] is a freely available, Windows-based EHR application system that offers a variety of features for managing clinical care for primarily HIV or AIDS patients and has been deployed in over 300 facilities in Kenya. The system also has a supply chain management feature for management of drugs and other consumables. IQCare is implemented in Kenya through the support of the Palladium Group (formerly Futures Group) and is donor-funded through AIDS Relief. Palladium is an international consulting firm that works in various industries to provide customized solutions. In Kenya, they work closely with the MOH in a range of health areas including HIV and AIDS and more generally providing strategic information capacity building.

Afya Electronic Health Management System (2014-Present)

The challenges reported in the reviews and assessments coupled with the need to have a comprehensive picture of patient care from the lowest level of the health system to referral facilities led the MOH, supported by the World Health Organization (WHO), to commission the development of a county electronic health record (CEHR) system now called AfyaEHMS. Other partners supporting the project were Department for International Development (DFID) and United States Agency for International Development-funded projects AfyaInfo [27] and APHIAplus Northern Arid Lands [28]. It was envisioned that the system would be implemented in 2 counties: Turkana County (located in the Northern more sparsely populated areas of Kenya) which had relatively few existing implementations in public health facilities and in theory allowing for a faster county-wide scale up and Machakos County (located in the more central, semiarid but more developed part of Kenya) which already had a system in place but would provide a good comparison to the Turkana implementation. Table 2 shows a timeline of these reports and projects.

Methods

Overview

This case study has been developed in 2 phases over a period of 2 years. In the first phase, a research team from Kenya Medical Research Institute (KEMRI)/Wellcome Trust Research Programme (ME, JM, NM) supported by the University of Oxford (CP) was commissioned by the MoH and WHO to report on the initial plans and progress of the AfyaEHMS project. The second stage of this research followed the conclusion of the initial commissioned work and was undertaken as part of a new wider study investigating the use of open source software in public hospitals in Kenya bringing in coinvestigators from Leeds University (HF) and the Department of General Practice at the University of Oxford (JP). This new remit allowed the team to look at the AfyaEHMS project in the context of the wider health IT landscape in public hospitals in Kenya and to follow up the project as it proceeded to other counties beyond the initial implementation.

Site Visits

To review the progress of implementation of the new system the team undertook 5 site visits to the county facilities (county referral hospital and health centers) over a period of 12 months to conduct semistructured and informal interviews with clinicians and IT staff. The first site visit to the county hospital was carried out to familiarize JM and NM with the site before system installation. Two site visits were done during installation of Version 1 of the system and 1 visit 6 weeks after the initial installation (JM). The last visit was carried out by NM after the developer had installed Version 2 of the system. During the site visits, informal interviews were carried out with the members of staff and field notes recorded as discussions took place. The team also attended 5 project meetings with the Kenyan implementation consultants and the MOH and 2 Skype calls with the developers and system users.

Follow-Up

Following the conclusion of the initial consultancy, the research team conducted a series of informal discussions with MoH officials (eHealth Unit) and the implementing consultants (Vimak). The initial rollout of the AfyaEHMS project was scaled back and a new version developed and implemented in new counties across Kenya. The research team also had discussions with the new system developer on the progress of AfyaEHMS.

Results

System Specification and Requirements Gathering

An MOH working group primarily concerned with carrying out monitoring and evaluation activities at the hospitals was charged with implementation of the CEHR. It was envisioned that the system would have a health information exchange (HIE) component to enable interoperability and sharing of data between the various modules of the EHR, within the hospital, between hospitals in the county, and into other health

information repositories such as the national health information system (DHIS2) and the human resources information system. Information collected by the EHR would include management information such as financial and human resources, and individual electronic medical records including pharmacy and laboratory information. The system was envisaged to function with health workers entering primary data as part of their work or near real time through data clerks. Figure 1 shows the proposed EHR at facility level.

Working with WHO, MOH, and various stakeholders, the implementation consultants defined the EHR requirements and produced a specifications document outlining each component of the system (in-patients, laboratory, billing, etc) at the start of January 2013. The consultants met with hospital teams and, using structured forms, defined various use cases to be included in the new EHR. A use case definition included the use case description, definition of actors, triggers, conditions, normal and alternative flows, frequency of use, exceptions, dependent use cases, special requirements, assumptions, and other notes. The use cases were used to define modules that would be expected in the system including registration, outpatient, referrals, pharmacy, laboratory, inpatient, mother and child health (MCH), specialized clinics, billing, financial information management, human resources, logistics, HIE, and the community health system. A summary of the use case definition is outlined in Table 3.

Figure 1. Proposed electronic health record at facility level (source: Kenyan Ministry of Health). LMIS: Logistics Management Information Systems; IFMIS: Integrated Financial Management Information Systems; DHIS: District Health Information Software; HRIS: Human Resources Information System; HIE: Health Information Exchange.

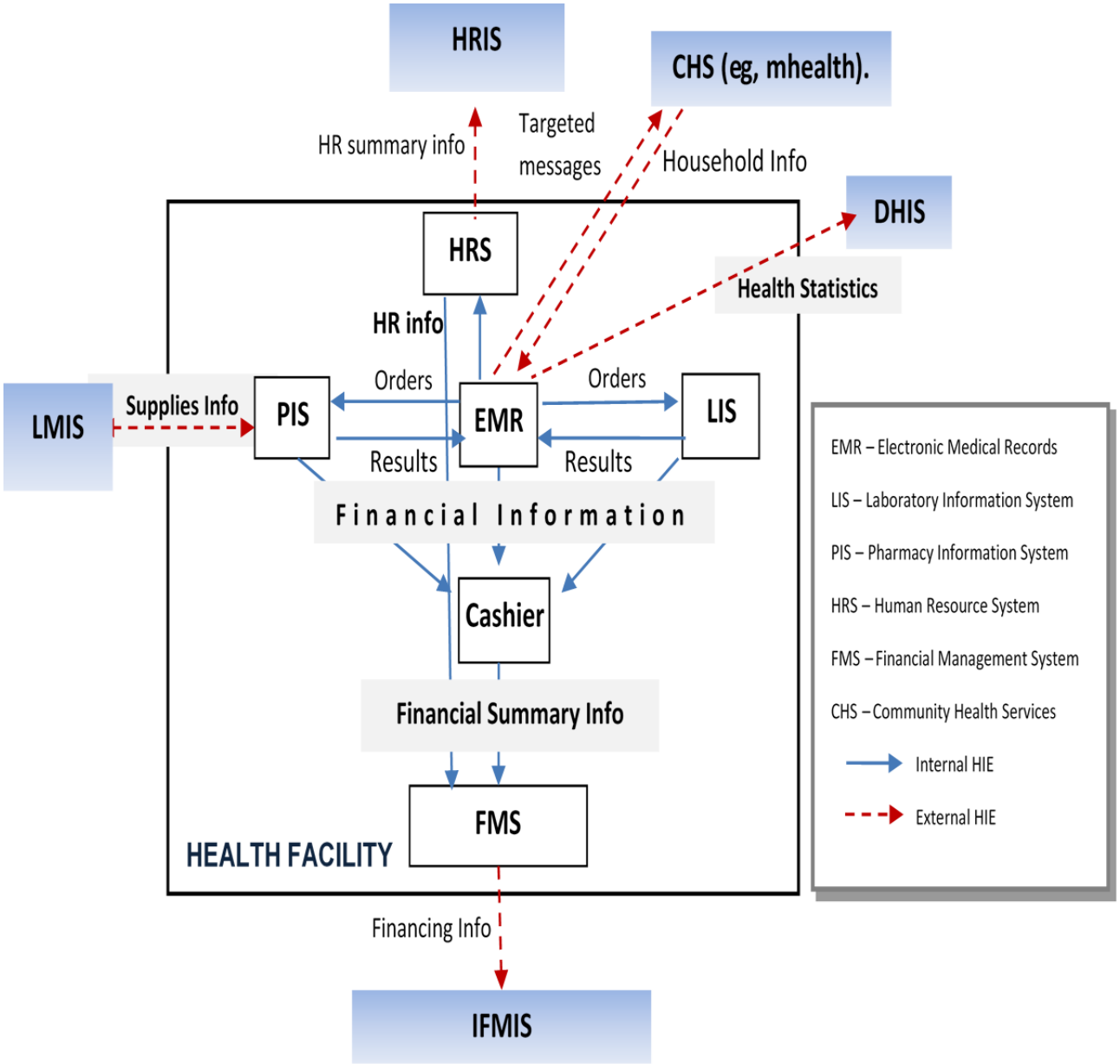


Table 3. Electronic health record (EHR) use cases. Source: EHR validated use cases.

Use case ID	Primary actor (system users)	Use cases	Description
UC-1	Clerk, patient	Registration	Register patient into system to link the patient to other modules or facilities. Used for inpatients and outpatients
UC-2	Clinician, patient	Outpatient	Record clinical details of evaluation of patients
UC-3	Clinician	Referrals	Refer patients for tests, diagnosis, or treatment to internal department or external facility (specialist)
UC-4	Pharmacist, patient	Pharmacy	Receives prescription and dispenses drugs to patient; receive order from inpatient ward or other pharmacy within facility and manage bulk order
UC-5	Lab technician	Laboratory	Lab results or diagnosis recorded
UC-6	Clinician, patient	Inpatient	Admit patient to the ward for further management, discharge patients, patient referral to theater and handling of deaths
UC-7	Clinician, patient	Mother child health	Manage maternity, antenatal care, and immunization services
UC-8	Clinician, patient	Specialized clinics	Record clinical details of evaluation of patient
UC-9	Clerk, patient	Billing	Record charges for health service to patient, produce receipts
UC-10	Accountant or clerk	Financial information management	Record revenues and expenses for the facility
UC-11	HR office or administrator	Human resources	Record cadre workloads in facility
UC-12	Stores officer	Logistics	Receive or dispatch items into or out of store
UC-13	Various	Health information exchange	Return to point of service (PoS) unique patient ID from County Master Patient Registry; retrieve clinical data from another PoS application; push clinical data to electronic medical record for updating orders or prescriptions
UC-14	Community health worker	Community health system	Report vital event data (births, deaths) to County Civil Registration System using mHealth solutions

An assessment of the readiness of the target counties was also carried out to allow for proper planning and support of system rollout. In addition to the readiness assessment, a site visit to Machakos County was also conducted for the consultants and developers to understand the working of typical health facilities within a county. Consultations on the implementation of the system were also undertaken at the county level to engage the county leaders.

System Selection and Development

The OpenMRS system was selected as the base platform for the implementation. A team of developers based in India were contracted to develop new system modules owing to prior experience in customizing OpenMRS for use in Indian hospital settings (in Kenya, OpenMRS had previously usually been implemented in small facilities such as HIV clinics).

The new EHR system would use the OpenMRS core architecture plus standard modules for patient management and clinical documentation. These modules would be augmented by an integrated suite of 10 modules for hospital management, including clinical, management, and administrative systems, customized specifically for workflow process within a district hospital system and integrated with DHIS2 using the WHO Statistical Data and Metadata eXchange for the Health Domain, in short, SDMX.HD standard.

Implementation in Machakos County

The first county to implement the new system, now called AfyaEHMS, was Machakos County, which is the focus of this paper. The system was later also rolled out in Baringo County to primary care-level facilities (4 health centers and 1 dispensary) at the same time. Three of the facilities took up the system but gradually stopped using it due to issues with ongoing support for the system, and they were not able to wait until a newer version of the system was ready.

Machakos County has a population of slightly over a million people and has health facilities that can be grouped as district or mission hospitals, referral and provincial hospitals, health centers, dispensaries, private hospitals, private clinics, maternity hospitals and nursing homes, and special treatment centers with the referral hospital providing the highest level of care in the county and also serving as a referral facility for neighboring counties. Health service delivery in Kenya is a devolved function run by 47 counties. The health delivery system is classified into 4 levels of care with different facilities falling into the levels according to the services they provide [29] as summarized in Table 4. Initially, the system was to be implemented in 6 public facilities (1 county hospital and 5 primary care facilities) with a view to expanding to other facilities as resources became available. The county hospital had an existing IT system in place but was motivated to install an MOH-backed system to try to lower costs, improve system performance, and increase access to technical support.

Table 4. Levels of care defined by the Kenya Health Policy 2014-2030.

Level of Care	Facilities
Level 1: Community	Community: Village/households/families/individuals
Level 2: Primary care facilities	Dispensaries or clinics and health centers
Level 3: County hospitals	Primary care hospitals; secondary care hospitals
Level 4: National referral hospitals	Tertiary care hospitals

To summarize the implementation of the AfyaEHMS project, we use a framework presented by Jawhari et al [30]. Their synthesis of key messages appearing in literature present a framework that can be used to summarize the benefits and barriers to EMR implementation in developing countries as systems, people, process, and products. Systems relates to infrastructure available such as power and a reliable network, people relates to factors to do with users such as their training and attitudes, process relates to how the system is implemented, for example, the change management process and time of deployment, while product relates to the system itself and how it interoperates with other applications. Table 5 summarizes the implementation of versions 1 and 2 of AfyaEHMS project.

Way Forward

Following the experiences during version 1 and 2 system implementations, the project implementers identified challenges and proposed solutions as outlined in Table 6.

The wide scope of the project was a major challenge during system development and implementation of Versions 1 and 2. The scope of the system was thus scaled back from a mix of health centers and county hospitals to cover only primary care facilities (Level 2) for the time-being. This allowed for faster rollout to more sites. Once this was done and at a stable level, then scale up to larger hospitals would be considered. More counties have since been targeted for rollout; currently 5 counties (Baringo, Kilifi, Bungoma, Garissa, Turkana) are on board with more being targeted for rollout with over 70 health centers having been installed to date. Health centers provide a wide range of predominantly outpatient services, such as basic curative and preventive services for adults and children, as well as reproductive health services minor surgical services and are staffed by midwives or nurses, clinical officers, and occasionally by doctors. They augment their service coverage with outreach services and refer severe and complicated conditions to the appropriate level, such as the district hospital [31].

Scaling back the system implementation to only health centers allowed the developer to focus their efforts on system modules other than the finance module. The finance module was an important module to large hospitals that collect revenue but not to health centers where care is free and was a barrier to full system implementation in Versions 1 and 2. Modules that are in use at the health centers include: patient registration, a clinical module for general outpatient services, pharmacy, laboratory, and a maternity module to cover antenatal services and the MCH clinic. Currently, the EHR does not cover the comprehensive care clinics (CCC), that give HIV care, but discussions are underway to find ways of integrating with existing systems and including the CCC functionality in a future version. Other key developments have been the implementation of a reporting module that generates a file that can be uploaded to DHIS2 (the national reporting system). There are plans to introduce internet to the facilities, and this will facilitate automatic reporting of data to DHIS2.

A Kenyan software development company has been engaged to develop the new system which should allow for faster system development and quicker resolution of emerging issues. A plan is in place to have a support team that will be responsible for the system handover over a longer period (6 months) allowing them to provide better system support to the health centers and thus ensuring system sustainability. To further facilitate faster resolution of issues, the project manager uses WhatsApp groups to support implementations within the counties whose membership includes system users and facility incharges.

Early stakeholder engagement with new counties helped to foster a feeling of ownership which was a major barrier to system adoption during the previous installation. The new county administration teams have in turn been supportive of the system implementation by availing resources (monetary and staffing) when necessary. Additionally, during implementation, the project implementation team now trains Trainers of Trainers; a team comprising a national team member, health workers (eg, health records information officer, pharmacist, lab technologist, or nurse) who have worked within the county, and members with IT training. These teams undergo a 3-day training supported by funds from the county and WHO. The eHealth unit at the MOH also sends a member to be present each time an implementation is taking place.

Previously, there were health centers that did not have electricity for up to 2 weeks making system implementation and use impossible. For this implementation round, the use of solar power has been considered for some sites while in others, generators are in use; this is done at the start of the implementation at the site. Depending on the setup, if a generator was available then that would be used as backup in case of power outage, if it was not in a working condition, then efforts were made to fix it.

To counter the challenge of laptops posing a security concern due to theft, the project now employs the use of zero clients and a server. Zero clients are all-in-one computer terminals that occupy relatively less space and are easier to rollout and maintain. The network has also been setup using a Local Area Network as opposed to a wireless network, which was not reliable previously.

Table 5. System implementation—Versions 1 and 2.

Determinants	Version 1		Version 2 (demonstration by clerks)	
	Description	Challenges	Description	Challenges
Systems	• Hardware: 15 laptops preloaded with Ubuntu Linux version 14.0 procured to be used in addition to the already existing hardware • Networking: wired and wireless connections • System setup to use laptops as client computers to access a central server allowing for portability • Information technology (IT) staff (2) in charge of expanding the computer network and general troubleshooting of hardware issues • Software support: provided by developer based in India	• Workstations insufficient: approximately 30 to 35 computers needed to cover all the departments • Laptops raised concerns of theft leading to delay in deployment of equipment in some sites • Inadequacies in infrastructure such as weak or missing Wi-Fi signal and poor 3G network made connecting to the internet difficult • Lack of electric power in a site leading to delay in deployment • Resolution of software issues were perceived to take too long	• Network improved ensuring accessibility from any computer connected to the network • Additional staffing in IT department (3 staff and 4 interns)	• Only the developer team could make software modifications to the system
People	• September 2014: training on system use completed at 4 (1 level-5 hospital and 3 level-3 facilities) out of 6 target facilities concurrently • Training completed at site of work • IT staff trained on system installation on the server	• Low levels of computer literacy • Reported high user workload • Limited support staff • Lack of user buy-in	• Some staff members trained on system use though this was not done for all staff • The data clerks were also trained and expected to train other users such as nurses on system use	• A major barrier to training all the staff was that the schedules for the different staff would not allow for all of them to be gathered at one place • Lack of user buy-in to the project as the development team and end-users were in different countries and had only limited time for communication and training
Process	• Use of data clerks to enter data from physical patient files to counter shortage of staff and busy work schedules		• Shifted responsibility of data accuracy and integrity to clerks, a role normally assigned to nurses and clinicians in order to verify the data	• Commissioning of a major project resulted in a shift of attention and resources hence not feasible to give the required attention and resources to the Afya Electronic Health Management System (in short AfyaEHMS) deployment
Products	• Modules: patient registration, outpatient, inpatient, laboratory, pharmacy, health records and hospital inventory	• Request for additional functionality (more comprehensive symptom lists, an option to enter free text) • Need for a more user-friendly International Statistical Classification of Diseases and Related Health Problems 10th Revision code list for diagnosis • Need to reduce number of steps required to achieve tasks (eg, pharmacy and inventory modules) • Patient identification number generated by the system was too long • Finance module not as comprehensive as the preexisting system	• Modules updated to incorporate requested changes	• Comprehensive testing needed to ascertain whether all changes requested were captured

Table 6. Challenges and proposed solutions.

Challenge	Proposed solution
Poor support and use of external developers	Need for a longer-term support solution
	Need for local developers to get involved sooner rather than later in the project
Poor support from county management	Engage with all stakeholders from an early stage to foster system ownership and ensure they are consulted during development and implementation
Wide project scope	Scale down system to cover smaller health facilities
Infrastructure issues	Better hardware solutions needed to ensure easier overall maintenance.

Discussion

Principal Findings

This case study describes a novel idea: to develop and deploy an EHR using existing open source software for use in public health facilities in Kenya. The project implementation was faced with some of the common problems with EHR roll-outs in both low-income settings, where EHRs have generally been used in smaller clinics and in high-income settings, where EHR implementations have been attempted with varying degrees of success in larger hospitals.

In common with other low-resource eHealth projects, the lack of power, inadequate hardware, and networking were a major challenge to system setup during the deployment of Version 1 and 2. In earlier projects, multiple power sources have been used to ensure the availability of power and system availability if one of the sources fails [32,33]. For this project, the implementing team addressed the power and hardware issues by adding extra local human resources for troubleshooting and fixing issues as they arose.

There is growing consensus in the international eHealth literature that overcoming challenges that are due to human factors such as computer literacy and attitudes can be a major step toward successful EHR implementation in both developed and developing countries [30,32,34,35]. We found that issues due to human factors caused significant problems with the implementations we studied with concerns about user acceptance of the new system. The users felt that the system belonged to *outsiders*, and this affected the system ownership. To overcome this, using system design strategies that are more inclusive, such as codesign or participatory design, at an early stage can be employed to help ensure system buy in from potential users [35,36]. In a similar case study implementing an EMR system at a large hospital, management of different users' expectations was noted as an important aspect of the successful implementation [35]. Different stakeholders have different interests and abilities to influence the process, which needs to be managed and planned for at an early stage of system implementation [37]. This coupled with managing the scope of the system using, for example, Agile software development principles [38] could help in gradually developing a system until it is fully operational while keeping relevant stakeholders on-board.

Hospitals are complex organizations [39] and, as such, implementing any new technology requires careful planning and management. The literature shows that eHealth implementers should take into account the existing workflows and organizational culture to come up with a change management plan that takes into account the different actors and their views [40]. Large hospitals operate with highly hierarchical structures and varying levels of availability of staff and these factors need to be considered to ensure successful implementation. Scaling back the implementation to the primary care facilities, which are less complex, has enabled the implementers to better deploy a better working system with plans to build on it once system operations stabilize.

Historically, data clerks or scribes have been used to enter clinical data into EHRs both in low-income [41,42] and high-income countries [43] in order to overcome the challenge of high clinician workload while deploying an EHR. The HIV clinics that use EHRs in Kenya have used data clerks through external support. However, for developing countries that are resource constrained, the use of data clerks on a long-term basis needs to be explored to establish its sustainability. The use of structured forms has been shown to improve the quality of documentation [44,45], a step toward improved quality of care. It has also been associated with increased generation of useful data [30] in comparison to the use of unstructured forms that rely on free-text input.

From an early stage in this case study, the implementers envisioned that system support would be offered through a help desk, where general system issues are addressed, and a community of practice (COP) where users could share experiences and learn from one another. COPs are often used in EHR implementations to provide an avenue to share innovations, help foster higher system utilization through mentor support, allow new staff members to quickly find clinical staff that are more familiar with the system, and provide an avenue to develop standardized templates for use in practice. They can also allow users greater influence in issues such as coordinating support with the vendor to optimize feature requests and training [46]. While some COPs may be self-organizing, the AfyaEHMS COP would have benefited greatly from a facilitator or coordinator. A dedicated facilitator helps the community to focus on its domain, maintain relationships, and develop its practice [47].

Use of open source software may offer some respite from the high costs of proprietary software, which is a well-documented barrier to adoption of EHRs [48]. Open source software is also often associated with online supporting communities that are constantly improving the software. The Esaude community is

an example of a local community focused on the development and implementation of a Mozambican specific configuration of the OpenMRS medical record software and the integration into a national eHealth architecture [49]. Members of the community collaborate and participate in the global OpenMRS community where they learn from the collaboration model and receive mentorship for learning and developing the software. Tapping into these communities may help reduce over reliance on one individual or software vendor for system updates, which are needed as the software evolves to suit the changing needs of the users and contributes to the principle of operational self-sufficiency noted by Surana et al [50] as being key to implementing any information and communication technology project. Such a community would bring on-board as many interested parties as possible that can continue to contribute to the project. Additionally, partnering with higher learning institutions may be a useful way to get more technical input into the project by engaging students to rapidly develop sections or modules of the system that might need improvement through boot camps or as part of their coursework through projects. An example of where this has been implemented is Rwanda, where a training program for local computer science graduates is being run to enable them to contribute to the implementation of the national EMR system [51].

Conclusions

Implementing EHR systems is a challenging process in high-income settings. In low-income settings, such as Kenya, open source software may offer some respite from the high costs of software licensing, but the familiar challenges of clinical and administration buy-in, the need to adequately train users, and the need for the provision of ongoing technical support are common across the North-South divide. We hope this case study will provide some lessons and guidance for other challenging implementations of EHR systems as they continue across Africa.

Acknowledgments

We would like to thank the WHO and the Ministry of Health for supporting this work. This work is published with the permission of the Director of KEMRI.

This project was funded by the Health Systems Research Initiative (MR/N005600/1) jointly supported by the Department for International Development (DFID), the Economic and Social Research Council (ESRC), the Medical Research Council (MRC), and the Wellcome Trust (WT). Funds from The Wellcome Trust (#076827, #097170) awarded to ME as Senior Fellowships, Oxford University—Wellcome Trust Institutional Strategic Support Fund awarded to ME, together with additional funds from a Wellcome Trust core grant awarded to the KEMRI-Wellcome Trust Research Programme (#092654) also supported this work with additional funds from WHO. HF was a Marie Skłodowska-Curie Fellow, funded from the European Union's Horizon 2020 research and innovation program under grant agreement No. 661289: "Global eHealth" and was a recipient of funds from CDC Global Aids program and PEPFAR for the development of OpenMRS. The funders had no role in drafting or the decision for submitting this manuscript.

Authors' Contributions

NM and CP prepared the initial draft, and all authors contributed to writing of this manuscript and provided their approval of the final manuscript.

Conflicts of Interest

HF is a cofounder of the OpenMRS EHR project and unpaid member of the OpenMRS leadership team.

References

1. Fraser HS, Biondich P, Moodley D, Choi S, Mamlin BW, Szolovits P. Implementing electronic medical record systems in developing countries. Inform Prim Care 2005;13(2):83-95. [Medline: 15992493]
2. Fraser HS, Blaya J. Implementing medical information systems in developing countries, what works and what doesn't. AMIA Annu Symp Proc 2010;2010:232-236 [FREE Full text] [Medline: 21346975]
3. Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. J Am Med Inform Assoc 2001;8(4):299-308 [FREE Full text] [Medline: 11418536]
4. English M, Irimu G, Agweyu A, Gathara D, Oliwa J, Ayieko P, et al. Building learning health systems to accelerate research and improve outcomes of clinical care in low- and middle-income countries. PLoS Med 2016 Apr;13(4):e1001991 [FREE Full text] [doi: 10.1371/journal.pmed.1001991] [Medline: 27070913]
5. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med 2010 Nov 10;2(57):57cm29. [doi: 10.1126/scitranslmed.3001456] [Medline: 21068440]
6. Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. Ann Intern Med 2012 Aug 07;157(3):207-210. [doi: 10.7326/0003-4819-157-3-201208070-00012] [Medline: 22868839]
7. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 4;362(5):382-385. [doi: 10.1056/NEJMp0912825] [Medline: 20042745]

8. Jha AK. Meaningful use of electronic health records: the road ahead. J Am Med Assoc 2010 Oct 20;304(15):1709-1710. [doi: 10.1001/jama.2010.1497] [Medline: 20959581]
9. Azubuike MC, Ehiri JE. Health information systems in developing countries: benefits, problems, and prospects. J R Soc Promot Health 1999 Sep;119(3):180-184. [Medline: 10518358]
10. Allen C, Jazayeri D, Miranda J, Biondich PG, Mamlin BW, Wolfe BA, et al. Experience in implementing the OpenMRS medical record system to support HIV treatment in Rwanda. Stud Health Technol Inform 2007;129(Pt 1):382-386. [Medline: 17911744]
11. Blaya JA, Shin SS, Yagui M, Contreras C, Cegielski P, Yale G, et al. Reducing communication delays and improving quality of care with a tuberculosis laboratory information system in resource poor environments: a cluster randomized controlled trial. PLoS One 2014;9(4):e90110 [FREE Full text] [doi: 10.1371/journal.pone.0090110] [Medline: 24721980]
12. Oluoch T, Katana A, Kwaro D, Santas X, Langat P, Mwalili S, et al. Effect of a clinical decision support system on early action on immunological treatment failure in patients with HIV in Kenya: a cluster randomised controlled trial. Lancet HIV 2016 Feb;3(2):e76-e84 [FREE Full text] [doi: 10.1016/S2352-3018(15)00242-8] [Medline: 26847229]
13. Siedner MJ, Santorino D, Lankowski AJ, Kanyesigye M, Bwana MB, Haberer JE, et al. A combination SMS and transportation reimbursement intervention to improve HIV care following abnormal CD4 test results in rural Uganda: a prospective observational cohort study. BMC Med 2015 Jul 06;13(160):160 [FREE Full text] [doi: 10.1186/s12916-015-0397-1] [Medline: 26149722]
14. Were MC, Nyandiko WM, Huang KT, Slaven JE, Shen C, Tierney WM, et al. Computer-generated reminders and quality of pediatric HIV care in a resource-limited setting. Pediatrics 2013 Mar;131(3):e789-e796 [FREE Full text] [doi: 10.1542/peds.2012-2072] [Medline: 23439898]
15. Ministry of Public Health and Sanitation. Nairobi, Kenya: NASCOP; 2009 Sep 19. EMR system assessments harmonization report URL: http://guidelines.health.go.ke/#/category/15/129/meta [accessed 2018-03-20] [WebCite Cache ID 6y3eaXjqC]
16. Ministry of Medical Services. Guidelines, Standards and Policies Portal.: Ministry of Public Health and Sanitation; 2011 Dec 01. Report on the review of EMR systems towards standardization URL: http://guidelines.health.go.ke/#/category/15/124/meta [accessed 2018-03-20] [WebCite Cache ID 6y3eCz6a8]
17. Guidelines, Standards and Policies Portal. 2010 Dec 20. Standards and guidelines for electronic medical record (emr) systems in Kenya URL: http://guidelines.health.go.ke/ [accessed 2018-03-20] [WebCite Cache ID 6y3eu45uR]
18. Kang'a S, Puttkammer N, Wanyee S, Kimanga D, Madrano J, Muthee V, et al. A national standards-based assessment on functionality of electronic medical records systems used in Kenyan public-Sector health facilities. Int J Med Inform 2017 Jan;97:68-75. [doi: 10.1016/j.ijmedinf.2016.09.013] [Medline: 27919397]
19. World Health Organization. 2011 Apr 01. Kenya national e-Health strategy - 2011-2017 URL: http://www.who.int/goe/policies/countries/ken/en/ [accessed 2018-03-20] [WebCite Cache ID 6y3fW1Gb2]
20. Manya A, Braa J, Øverland LH, Nzioka C. National roll out of District Health Information Software (DHIS 2) in Kenya, 2011 central server and cloud based infrastructure. Presented at: IST-Africa 2012 Conference Proceedings; May 9-11, 2012; Tanzania.
21. DHIS 2 Overview. 2012. URL: https://www.dhis2.org/overview [WebCite Cache ID 6wMnlvOOa]
22. Manya A, Nielsen P. The use of social learning systems in implementing a web-based routine health information system in Kenya. Presented at: 13th International Conference on Social Implications of Computers in Developing Countries; May 20-22, 2015; Negombo, Sri Lanka p. 552-560.
23. OpenMRS. 2016. KenyaEMR distribution URL: https://wiki.openmrs.org/display/docs/KenyaEMR+Distribution [accessed 2016-10-21] [WebCite Cache ID 6qEZHKc5m]
24. Mamlin BW, Biondich PG, Wolfe BA, Fraser H, Jazayeri D, Allen C, et al. Cooking up an open source EMR for developing countries: OpenMRS - a recipe for successful collaboration. AMIA Annu Symp Proc 2006:529-533 [FREE Full text] [Medline: 17238397]
25. International Training and Education Center for Health (I-TECH). Kenya: Where we work URL: https://www.go2itech.org/where-we-work/kenya/ [accessed 2018-01-10] [WebCite Cache ID 6wMZnvbo1]
26. CodePlex Archive. 2015. FGIQCare: IQCare Version 3.6 URL: https://fgiqcare.codeplex.com [accessed 2018-01-11] [WebCite Cache ID 6wNzF61aq]
27. ABT Associates. 2016. Afyainfo: Bringing Good Health to Kenyans through Better Information Systems URL: http://abtassociates.com/Projects/2012/AIDSTAR2.aspx [accessed 2018-01-11] [WebCite Cache ID 6wNxzpmKG]
28. Federal Grants. APHIAplus Northern Arid Lands (APHIAplus NAL) URL: http://www.federalgrants.com/APHIAplus-Northern-Arid-Lands-APHIAplus-NAL-32977.html [accessed 2017-05-05] [WebCite Cache ID 6qEZqN7Nk]
29. Ministry of Health Kenya. Kenya health policy 2014 to 2030 URL: http://publications.universalhealth2030.org/ref/d6e32af10e5c515876d34f801774aa9a [accessed 2018-03-20] [WebCite Cache ID 6y3k0LuU9]
30. Jawhari B, Ludwick D, Keenan L, Zakus D, Hayward R. Benefits and challenges of EMR implementations in low resource settings: a state-of-the-art review. BMC Med Inform Decis Mak 2016 Sep 06;16:116 [FREE Full text] [doi: 10.1186/s12911-016-0354-8] [Medline: 27600269]

31. Muga R, Kizito P, Mbayah M, Gakuruh T. Dhsprogram. Overview of the health system in Kenya, in Kenya service provision assessment (KSPA 2004) survey URL: https://dhsprogram.com/pubs/pdf/spa8/02chapter2.pdf [accessed 2018-03-20] [WebCite Cache ID 6y3kFHBkt]
32. Hannan TJ, Rotich JK, Odero WW, Menya D, Esamai F, Einterz RM, et al. The Mosoriot medical record system: design and initial implementation of an outpatient electronic record system in rural Kenya. Int J Med Inform 2000 Oct;60(1):21-28. [doi: 10.1016/S1386-5056(00)00068-X]
33. Hannan TJ, Tierney WM, Rotich JK, Odero WW, Smith F, Mamlin JJ, et al. The MOSORIOT medical record system (MMRS) phase I to phase II implementation: an outpatient computer-based medical record system in rural Kenya. Stud Health Technol Inform 2001;84(Pt 1):619-622. [Medline: 11604811]
34. Paré G, Raymond L, de Guinea A O, Poba-Nzaou P, Trudel M, Marsan J, et al. Barriers to organizational adoption of EMR systems in family physician practices: a mixed-methods study in Canada. Int J Med Inform 2014 Aug;83(8):548-558. [doi: 10.1016/j.ijmedinf.2014.06.003] [Medline: 24969270]
35. Scholl J, Syed-Abdul S, Ahmed LA. A case study of an EMR system at a large hospital in India: challenges and strategies for successful adoption. J Biomed Inform 2011 Dec;44(6):958-967 [FREE Full text] [doi: 10.1016/j.jbi.2011.07.008] [Medline: 21846508]
36. Waller A, Franklin V, Pagliari C, Greene S. Participatory design of a text message scheduling system to support young people with diabetes. Health Informatics J 2006 Dec;12(4):304-318. [doi: 10.1177/1460458206070023] [Medline: 17093001]
37. Boonstra A, Boddy D, Bell S. Stakeholder management in IOS projects: analysis of an attempt to implement an electronic patient file. Eur J Inf Syst 2008;17(2):100-111. [doi: 10.1057/ejis.2008.2]
38. Gothelf J, Seiden J. In: Eric R, editor. Lean UX: Applying Lean Principles to Improve User Experience. California: O'Reilly Media; 2013.
39. Begun JW, Zimmerman B, Dooley K. Health care organizations as complex adaptive systems. In: Mick SM, Wyttenbach M, editors. Advances in Health Care Organization Theory. San Francisco: Jossey-Bass; 2003:253-288.
40. Blavin F, Ramos C, Shah A, Devers K. Urban Institute. 2013. Lessons from the literature on electronic health record implementation URL: https://www.healthit.gov/sites/default/files/hit_lessons_learned_lit_review_final_08-01-2013.pdf [accessed 2018-03-20] [WebCite Cache ID 6y3m2bU67]
41. Tierney WM, Achieng M, Baker E, Bell A, Biondich P, Braitstein P, Tanzania-Uganda Openmrs Consortium. Experience implementing electronic health records in three East African countries. Stud Health Technol 2010;160(Pt 1):371-375. [Medline: 20841711]
42. Forster M, Bailey C, Brinkhof MW, Graber C, Boulle A, Spohr M, ART-LINC collaboration of International Epidemiological Databases to Evaluate AIDS. Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings. Bull World Health Organ 2008 Dec;86(12):939-947 [FREE Full text] [Medline: 19142294]
43. Greiver M, Barnsley J, Aliarzadeh B, Krueger P, Moineddin R, Butt DA, North Toronto Research Network (NorTReN). Using a data entry clerk to improve data quality in primary care electronic medical records: a pilot study. Inform Prim Care 2011;19(4):241-250 [FREE Full text] [Medline: 22828579]
44. Mwakyusa S, Wamae A, Wasunna A, Were F, Esamai F, Ogutu B, et al. Implementation of a structured paediatric admission record for district hospitals in Kenya--results of a pilot study. BMC Int Health Hum Rights 2006 Jul 20;6:9 [FREE Full text] [doi: 10.1186/1472-698X-6-9] [Medline: 16857044]
45. Tuti T, Bitok M, Malla L, Paton C, Muinga N, Gathara D, et al. Improving documentation of clinical care within a clinical information network: an essential initial step in efforts to understand and improve care in Kenyan hospitals. Br Med J Glob Health 2016;1(1):e000028 [FREE Full text] [doi: 10.1136/bmjgh-2016-000028] [Medline: 27398232]
46. Smith J. Communities of practice: leadership in practice. B C Med J 2010;52(2):65 [FREE Full text]
47. Wenger E, McDermott R, Snyder WM. Cultivating Communities of Practice: A Guide to Managing Knowledge. Boston: Harvard Business Review Press; 2002:304.
48. Kruse CS, Kothman K, Anerobi K, Abanaka L. Adoption factors of the electronic health record: a systematic review. JMIR Med Inform 2016;4(2):e19 [FREE Full text] [doi: 10.2196/medinform.5525] [Medline: 27251559]
49. OpenMRS. 2016. eSaúde: Building a Local OpenMRS Community to Support a National EMR Implementation in Mozambique URL: http://openmrs.org/2016/08/mozambique-case-study/ [accessed 2018-01-11] [WebCite Cache ID 6wNxX0po1]
50. Surana S, Patra R, Nedevschi S, Brewer E. Deploying a rural wireless telemedicine system: experiences in sustainability. IEEE Comput Soc 2008 Jun;41(6):48-56 [FREE Full text] [doi: 10.1109/MC.2008.184]
51. Seymour RP, Tang A, DeRiggi J, Munyaburanga C, Cuckovitch R, Nyirishema P, et al. Training software developers for electronic medical records in Rwanda. Stud Health Technol Inform 2010;160(Pt 1):585-589. [Medline: 20841754]

Abbreviations

AfyaEHMS: Afya Electronic Health Management System
CCC: comprehensive care clinics

CDC: Centers for Disease Control and Prevention
COP: community of practice
DHIS2: District Health Information System Version 2
EHR: electronic health record
EMR: electronic medical record
ESG: EMR Standards and Guidelines
HIE: health information exchange
HMIS: Health Management Information Systems
IQCare: International Quality Care
IT: information technology
ITECH: International Training and Education Center for Health
KEMRI: Kenya Medical Research Institute
MOH: Ministry of Health
NASCOP: National AIDS and STI Control Programme
OpenMRS: Open Medical Record System
SDMX.HD: Statistical Data and Metadata eXchange for the Health Domain
TB: tuberculosis
WHO: World Health Organization

Edited by G Eysenbach; submitted 17.08.17; peer-reviewed by T Hannan, MC Trudel; comments to author 22.11.17; revised version received 16.01.18; accepted 31.01.18; published 18.04.18.

Please cite as:
Muinga N, Magare S, Monda J, Kamau O, Houston S, Fraser H, Powell J, English M, Paton C
Implementing an Open Source Electronic Health Record System in Kenyan Health Care Facilities: Case Study
JMIR Med Inform 2018;6(2):e22
URL: http://medinform.jmir.org/2018/2/e22/
doi:10.2196/medinform.8403
PMID:29669709

Original Paper

A Neuroimaging Web Services Interface as a Cyber Physical System for Medical Imaging and Data Management in Brain Research: Design Study

Gabriel Lizarraga[1*], MA; Chunfei Li[1*], MA; Mercedes Cabrerizo[1*], MA, PhD; Warren Barker[2*], MA; David A Loewenstein[3*], PhD; Ranjan Duara[2*], MD; Malek Adjouadi[1*], MA, PhD

[1]Center for Advanced Technology and Education, Computer Science, Florida International University, Miami, FL, United States

[2]Wien Center for Alzheimer's Disease & Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL, United States

[3]Center on Aging and Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, FL, United States

[*]all authors contributed equally

Corresponding Author:
Gabriel Lizarraga, MA
Center for Advanced Technology and Education
Computer Science
Florida International University
10555 West Flagler Street
EC 2220
Miami, FL, 33174
United States
Phone: 1 3053484106
Email: gliza002@fiu.edu

Abstract

Background: Structural and functional brain images are essential imaging modalities for medical experts to study brain anatomy. These images are typically visually inspected by experts. To analyze images without any bias, they must be first converted to numeric values. Many software packages are available to process the images, but they are complex and difficult to use. The software packages are also hardware intensive. The results obtained after processing vary depending on the native operating system used and its associated software libraries; data processed in one system cannot typically be combined with data on another system.

Objective: The aim of this study was to fulfill the neuroimaging community's need for a common platform to store, process, explore, and visualize their neuroimaging data and results using Neuroimaging Web Services Interface: a series of processing pipelines designed as a cyber physical system for neuroimaging and clinical data in brain research.

Methods: Neuroimaging Web Services Interface accepts magnetic resonance imaging, positron emission tomography, diffusion tensor imaging, and functional magnetic resonance imaging. These images are processed using existing and custom software packages. The output is then stored as image files, tabulated files, and MySQL tables. The system, made up of a series of interconnected servers, is password-protected and is securely accessible through a Web interface and allows (1) visualization of results and (2) downloading of tabulated data.

Results: All results were obtained using our processing servers in order to maintain data validity and consistency. The design is responsive and scalable. The processing pipeline started from a FreeSurfer reconstruction of Structural magnetic resonance imaging images. The FreeSurfer and regional standardized uptake value ratio calculations were validated using Alzheimer's Disease Neuroimaging Initiative input images, and the results were posted at the Laboratory of Neuro Imaging data archive. Notable leading researchers in the field of Alzheimer's Disease and epilepsy have used the interface to access and process the data and visualize the results. Tabulated results with unique visualization mechanisms help guide more informed diagnosis and expert rating, providing a truly unique multimodal imaging platform that combines magnetic resonance imaging, positron emission tomography, diffusion tensor imaging, and resting state functional magnetic resonance imaging. A quality control component was reinforced through expert visual rating involving at least 2 experts.

Conclusions: To our knowledge, there is no validated Web-based system offering all the services that Neuroimaging Web Services Interface offers. The intent of Neuroimaging Web Services Interface is to create a tool for clinicians and researchers

with keen interest on multimodal neuroimaging. More importantly, Neuroimaging Web Services Interface significantly augments the Alzheimer's Disease Neuroimaging Initiative data, especially since our data contain a large cohort of Hispanic normal controls and Alzheimer's Disease patients. The obtained results could be scrutinized visually or through the tabulated forms, informing researchers on subtle changes that characterize the different stages of the disease.

(JMIR Med Inform 2018;6(2):e26) doi:10.2196/medinform.9063

KEYWORDS

neuroimaging; multimodal imaging; magnetic resonance imaging; image processing; positron-emission tomography; diffusion tensor imaging; information storage and retrieval; diagnostic imaging

Introduction

Background

Noninvasive brain imaging modalities contribute considerably to the understanding of brain structure and functionality [1]. Magnetic resonance imaging (MRI), positron emission tomography (PET), diffusion tensor imaging (DTI), and functional magnetic resonance imaging (fMRI) scans, among other modalities, allow clinicians and experts to advance their research and take informed decisions on the diagnosis and the planning of clinical or therapeutic interventions that could follow. The images obtained by these scans must first be preprocessed in order to convert them into numeric values that can be objectively assessed and analyzed. Hospitals and other research institutions can capture, store, and view brain scans on their own picture archiving and communication system; but performing additional processing is often computationally taxing, requiring specialized software, a suitable hardware or computing infrastructure, and image processing expertise, which our Neuroimaging Web Services Interface (NWSI) is designed to offer.

In addition to the need of individual investigators to test and validate results, there is a larger neuroscience community in academia and medical settings that can benefit from this integrated processing and visualization platform. Data sharing, which remains limited due to the different institutional and privacy constraints, should be encouraged within the scientific community to increase the value of research. The governing council of the Organization for Human Brain Mapping, which is the primary international organization dedicated to neuroimaging research, highlighted in 2001 certain challenges in the field of databases in neuroimaging, most of which we still face today. Such challenges include (1) management of large volume and variety of forms in which the data are presented, (2) methods for the processing of brain images, (3) accessibility of data, and (4) the lack of access to neuroimaging results to investigators [2,3].

Among the most established neuroimaging databases is the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [4], which currently contains data from over 1900 subjects, encompassing over 4000 MRI and PET scans, as well as clinical, cerebrospinal fluid, genetic, and biochemical biomarkers, which have been made available to researchers worldwide, who have made over 14 million downloads. Many other databases with more specialized audiences exist and have been cataloged in Neuroscience Information Framework [5]. These include the Minimal Interval Resonance Imaging in Alzheimer's Disease database [6], the OpenfMRI database [7], NeuroVault [8], The Virtual Brain [9], Neuroimaging Data Model [10], the Vanderbilt University Institute for Imaging Science Center for Computational Imaging XNAT (Extensible Neuroimaging Archive Toolkit-based repository) [11].

Study Aims

NWSI serves as an automated, responsive, and scalable neuroimaging database solution. This new design serves also as a cyber physical system in that it offers users access to neuroimaging algorithms through the Internet and provides the needed computational resources with all the required processing, storage capabilities, security and operational maintenance. It comprises a Web interface and a set of replica Linux servers that perform specific tasks. Interacting with the system requires minimal computing knowledge, equivalent to what is expected from social media or similar type of Web interface. This interface would also serve the research community for applying their new data mining and deep learning concepts on a multimodal imaging platform [12].

NWSI is equipped with various useful tools such as Brain Extraction Tool (BET), brain image registration, image format conversion, data processing, and visualization mechanisms that help with rating and diagnosis. The current implementation includes (1) automatic quantification of volumes from anatomical MRIs, (2) 18F-Florbetapir and 18F-Florbetaben for Alzheimer disease (AD), (3) Fluorodeoxyglucose (FDG) PET analysis for epilepsy, and (4) DTI image processing for both AD and epilepsy. All the data results are collected in files and into a MySQL database and can be exported into both tabulated files and image files. The accumulated data can be used in future pipelines as input to multimodal and longitudinal studies.

NWSI utilizes an embedded, modified version of the Papaya viewer (a JavaScript medical research image viewer) developed by University of Texas Health Science Center. The viewer allows interactive display of cerebral regions, diffusion images, and PET data. All images are coregistered to the anatomical MRI as part of the pipeline; they can be displayed on the same viewer in stacked layers. Moreover, results have been validated by comparison with existing processed data, such as from the ADNI database, which provides an excellent source of raw and postprocessed data for validating the various functions of NWSI.

Methods

Data

Data included in NWSI involves both AD data and epilepsy data, which are recorded on a regular basis as more cases are routinely scheduled.

Alzheimer Disease Data

AD imaging data included in this Web interface were obtained from the 1Florida Alzheimer's Disease Research Center (1Florida ADRC) and ADNI databases.

1Florida Alzheimer's Disease Research Center

NWSI was piloted using a database of MRI and amyloid PET images obtained through 1Florida ADRC. For this pilot project, currently 273 structural MRI, 43 18F-Florbetapir PET scans, and 89 18F-Florbetaben PET scans are available. The MRI images were obtained using a Siemens Medical System Skyra 3 Tesla Scanner. The DTI scans were used to measure radial, axial, and mean diffusivity, as well as fractional anisotropy (FA). PET images were obtained from a Siemens Biograph 16 Hi-Rez PET-CT machine.

Alzheimer's Disease Neuroimaging Initiative

Data used for validation purposes were obtained from the AD Neuroimaging Initiative (ADNI) database .The primary goal of ADNI has been to test whether serial MRI, 18F-Florbetapir PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. MRI scans were acquired from 1.5 T or 3 T scanners at multiple sites across United States and Canada.

Epilepsy Data

Epilepsy PET scans were obtained from Baptist Health South Florida, and Nicklaus Children's Hospital. There are currently 10 cases in the database, with more cases to be included in the near future as the schedule allows. The intent here is to include EEG data for 3D source localization in context with hypometabolism as observed through PET. Subjects from the Baptist Hospital of Miami were scanned with a Philips PET or computed tomography scanner with an FDG imaging agent. Subjects from Nicklaus Children's Hospital were scanned with a GE Discovery ST PET or CT system with an FDG imaging agent.

Neuroimaging Web Services Interface and Hardware Architecture

Neuroimaging Web Services Interface

The Web interface driving NWSI is based on Drupal, a popular open source content management system, which is the platform for BBC, University of Oxford, the US Department of Energy, among other well-established organizations. Drupal provides a user-based platform, in which the core code for security and design tools are updated and patched frequently to address vulnerabilities, as well as to add new functionalities. New features can also be added to Drupal via modules that can be integrated with its core code, allowing new code to run on Drupal, while maintaining the core software secure and intact.

The Web interface of NWSI has a simplified design, utilizing forms and uploaded files for most of the data input. Users of the interface, who are not familiar with Linux or its command line arguments, will be able to upload, view, and visualize existing data. Figure 1 shows the MRI upload form as an example. All data are deidentified before being uploaded to the server, and the user determines whether or not the data on NWSI will be shared with other users. Access to the site is provided through password-protected accounts.

Although the user interaction occurs through the Web interface, a set of replica servers (RSs), which run on Linux, perform a variety of asynchronous tasks, such as running FreeSurfer [13] on anatomical MRIs, or registering structural MRI to PET or DTI images. To keep the interface responsive, new tasks are sent to a work loader on the Web interface, which in turn sends tasks to one or more available RS(s) on the basis of their current status. Once the data is copied back to the Web server, the PostProcessing Core incorporates it into the system's database and file system. The architecture is scalable, such that new RSs, which are easy to maintain clones with identical software, can be added on demand to the system.

The asynchronous communication between the Web interface and the RSs is achieved by securely copying files. Some of these files are data to be processed, whereas others are status reports and workload balance data. MRI, PET, DTI, and fMRI images are processed on the RSs, but smaller tasks, such as brain extraction or registration, are done synchronously on the Web server by the Short Task Module. Tasks selected to run on the Small Task Modules must be brief, no longer than few minutes in duration to keep the Web server responsive.

The interface is illustrated in a PowerPoint presentation and video included in Multimedia Appendix 1 and Multimedia Appendix 2, respectively.

Neuroimaging Web Services Interface Hardware Architecture

Virtual technology and Modular Smart Array Systems are used to host the NWSI Web interfaces and its RSs. The cluster-aware infrastructure has two ProLiant DL 380 G7 servers, forming a centralized pool of resources that is used to create virtual machines (VMs) which run their own operating system (OS). Figure 2 illustrates the virtual architecture of NWSI. The Web interface runs on a VM running an Ubuntu open source software OS, Apache software, and a MySQL database management system. Since the processing time for a single MRI ranges from 8 to 12 hours, as many as 16 MRIs can be processed simultaneously, using RSs with 8 cores each. PETs and DTIs are processed in 15 and 10 min, respectively. The use of a virtual-server environment adds availability, security, and scalability to the Neuroimaging Web-interface application.

Figure 1. Data upload form, with magnetic resonance imaging (MRI) upload selected. Same function could be done with other modalities (positron emission tomography, diffusion tensor imaging, and functional magnetic resonance imaging).

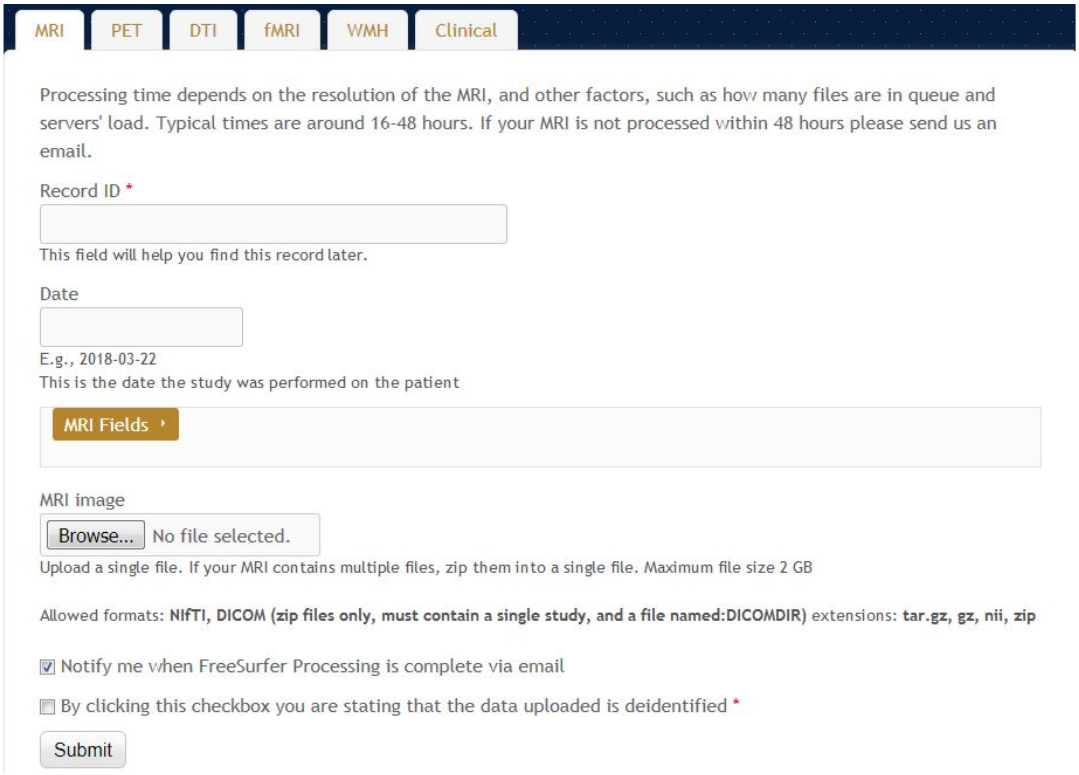
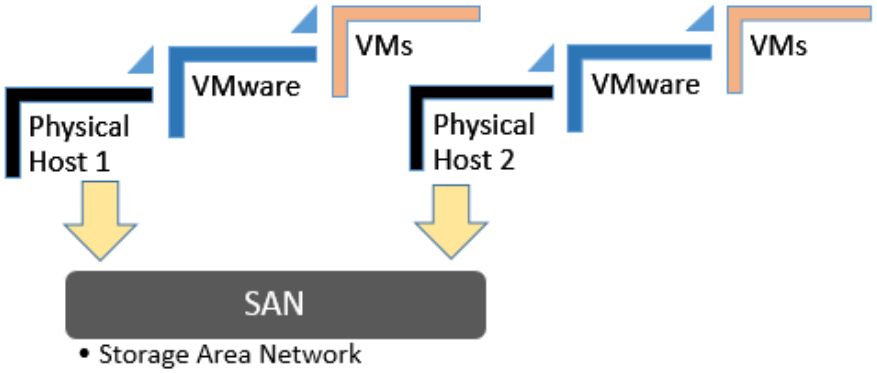


Figure 2. Virtual architecture.



Furthermore, new VMs can be added to the current design to increase the capability and performance of the system. RSs can reside in a private cloud or the Internet, as long as files can be securely copied between them and the Web interface. The Web interface can be accessed anywhere in the world with a fast Internet connection and a browser. It is both device and OS independent. The Drupal theme is responsive and tablet or phone friendly. Figure 3 shows the architecture of the neuroimaging Web-interface system, wherein the user interacts with the Web server through the Web browser.

Results from completed tasks are readily viewable. New tasks are sent by the Work loader, to an available RS, which sends the completed task to the PostProcessing Core, from which new values are entered into the database and raw and new images are stored on the file system. Registration, Brain Extraction, and other smaller tasks are processed on the Web server by the Small Tasks Module.

Image Viewer

Papaya, developed by the University of Texas Health Science Center at San Antonio, is a powerful open source, interactive, JavaScript-based image viewer that is incorporated within NWSI. The Web interface accepts the 3 most common formats as input—DICOM, ANALYZE, and NIfTI [14], but converts all files to NIfTI, which is versatile, more compact, and widely used. It is noted that supporting software, such as FSL [15], only accepts NIfTI as input.

The version of Papaya in NWSI has been modified to display FreeSurfer labels and custom color tables. The Web interface is also capable of displaying specific color-coded FreeSurfer regions; whole brain segmentation; interactive surfaces; and PET, fMRI, and DTI images. As part of our processing pipeline, all image files are registered to the structural MRI scan, making it possible to display several modalities as layers in the same viewer. Among the tools embedded on the Web interface and

available to the user by Papaya are color selection, a measuring tool, an axis viewer, and an image transparency modification tool, all of which are standard in many other viewers. The user can display these images online without having to save any files to the local hard disk (see Figures 4 and 5 below for illustrative examples). Furthermore, the ROI explorer page displays a color-coded segmentation of FreeSurfer regions. This is especially useful for researchers who are not familiar with FreeSurfer labels, but are familiar with the human brain anatomy. Pertinent information can be visualized by clicking on specific regions to scrutinize what the different regions reveal (see Figures 6 and 7).

Figure 3. Architecture of the Neuroimaging Web Interface System.

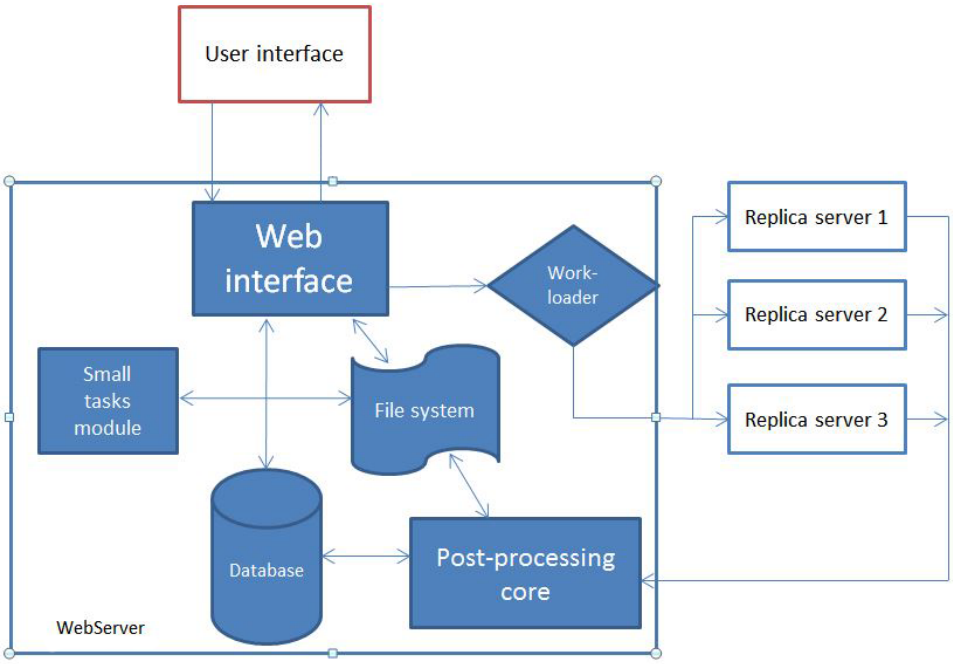


Figure 4. Interactive viewer, showing the surface reconstruction of and anatomical magnetic resonance imaging (MRI), Green represents the gray matter, Red represents the white matter.

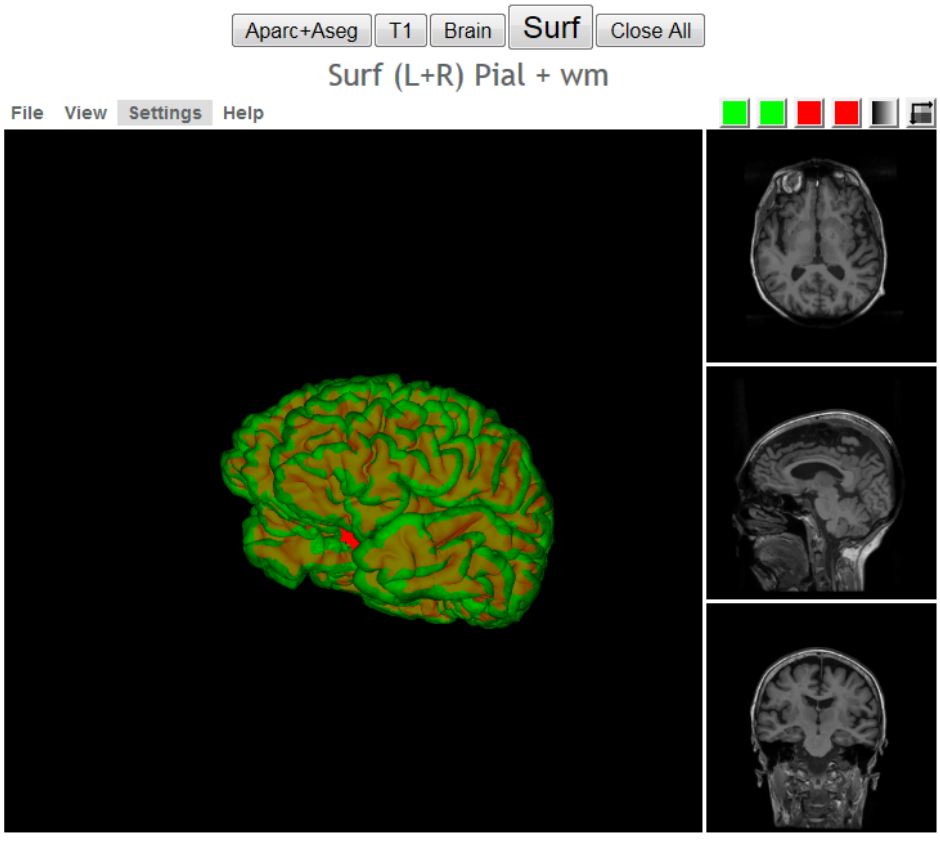


Figure 5. Image shows positron emission tomography (PET) viewer's full interface.

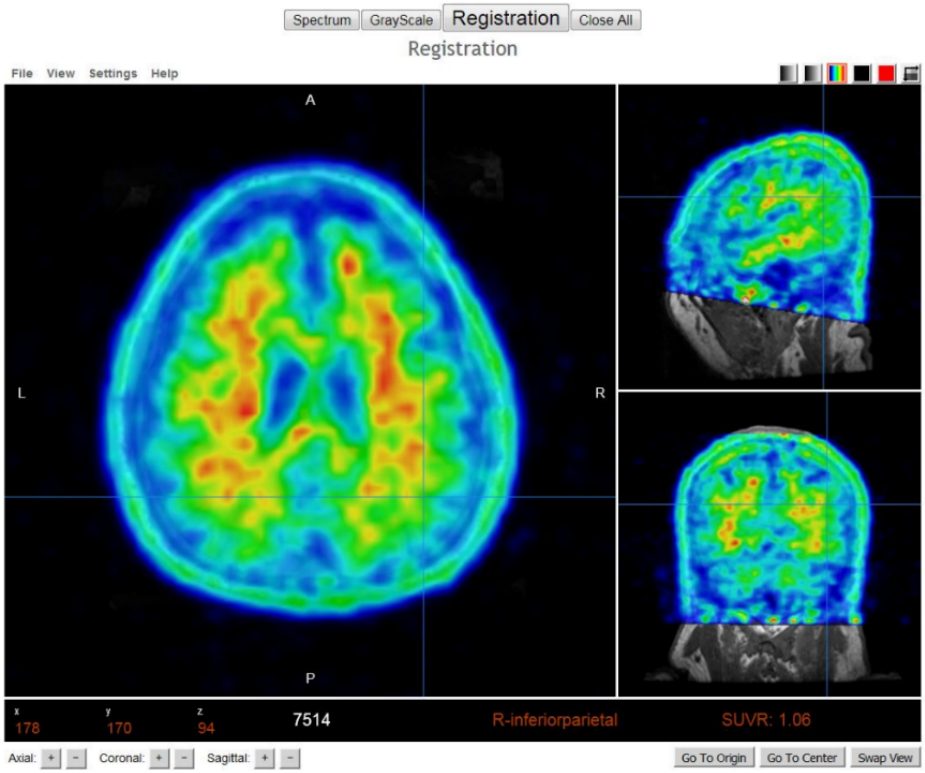


Figure 6. Regions of interest (ROI) Explorer: visualizing FreeSurfer segmentation. By default, all regions, cortical and subcortical, are shown.

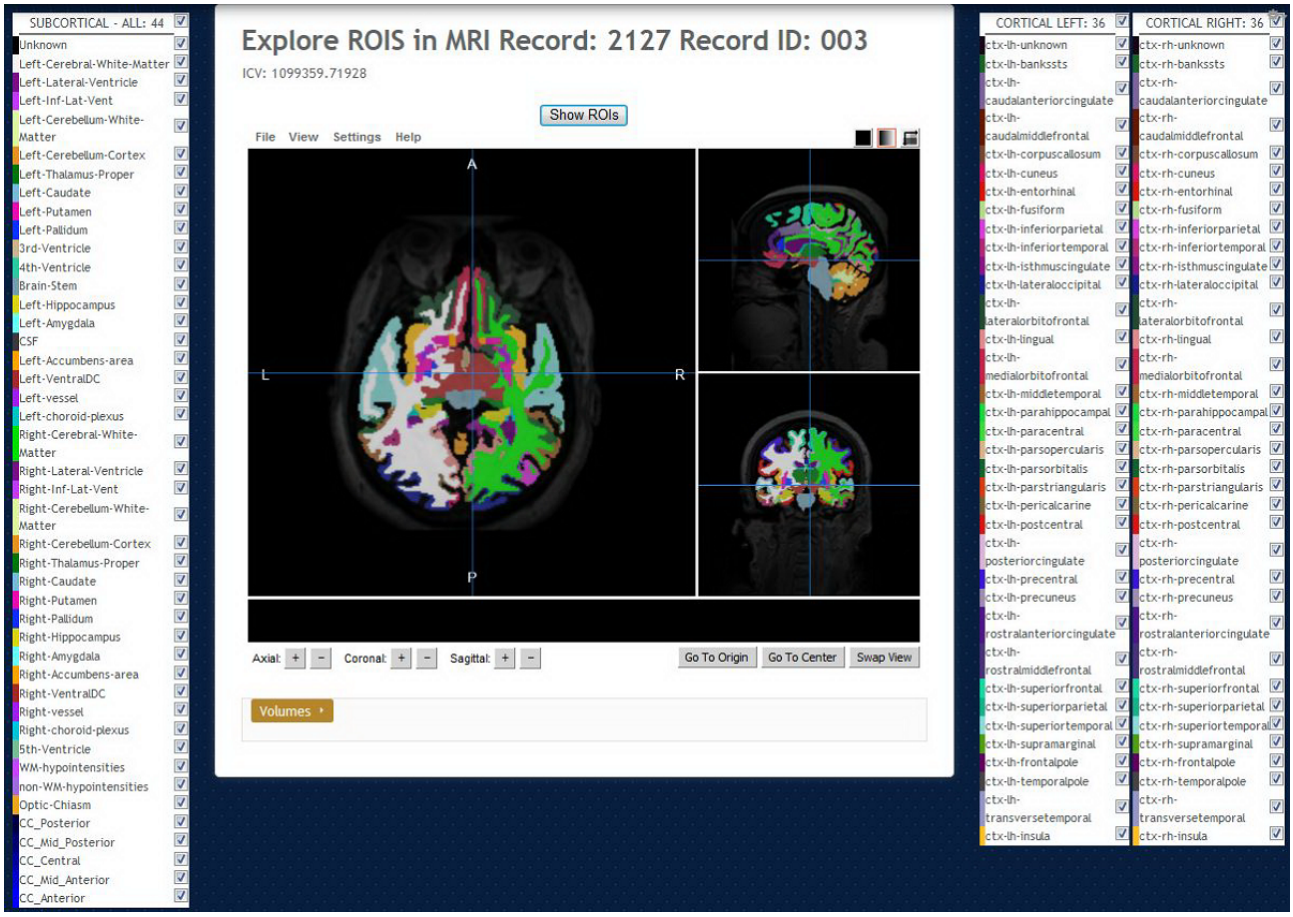
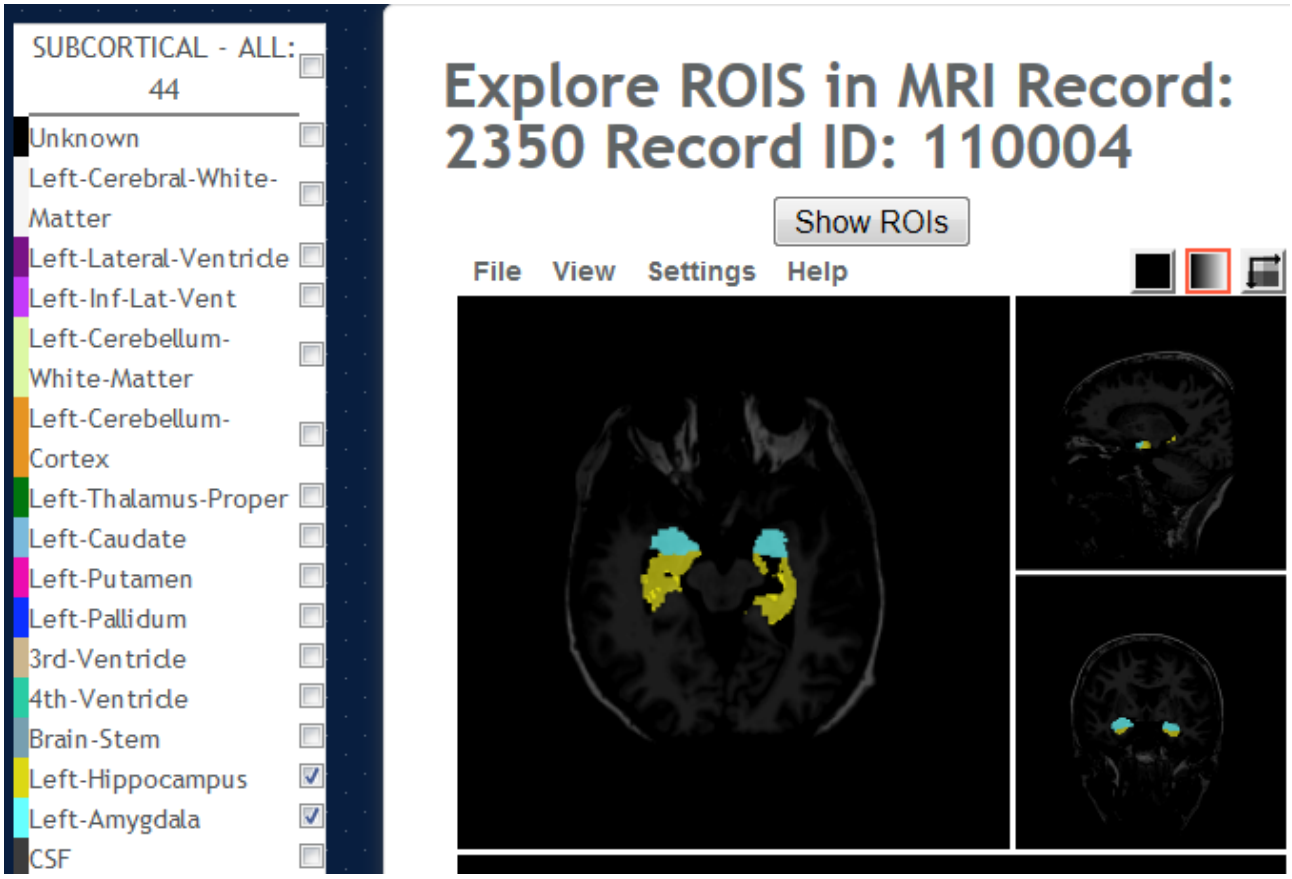


Figure 7. Regions of interest (ROI) Explorer: showing the selected subcortical left and right-hippocampus and amygdala regions.



Results

Volumetric and Cortical Thickness Calculations From Anatomical Magnetic Resonance Imaging

The basic functionality of NWSI depends on using FreeSurfer to reconstruct cortical surface models (gray-white boundary surface and pial surface) from structural MRIs and to output regional cortical and subcortical volumes, cortical thickness, and other values derived from input image segmentation (see FreeSurferWiki or a complete list of such measures). FreeSurfer also outputs image files that define the segmentation and replaces intensity on these files by numeric values representing the segmented regions. These files are used in PET standardized uptake values (SUV) calculations.

NWSI processes all structural MRIs on a local server, ensuring that the FreeSurfer results are validated and not affected by the OS version adopted [16]. FreeSurfer specifies in *fswiki* that certain OS-level libraries might affect the results. Thus, as new RSs are added to the system, it is imperative to test them before deployment to make sure the results are validated against established ones.

An important issue, which is resolved by NWSI, is in the ability of merging of data from different sources, which is a nontrivial task due to factors such as scanner bias, scanner field strength, etc [17]. This problem is best addressed by processing all values with the same hardware and software. Consequently, subjects from one institution can be merged with subjects from another institution, and the results can be downloaded and tabulated in a format as shown in Figure 8. In fact, NWSI has managed to achieve that by combining subjects from our own 1Florida ADRC to subjects from the ADNI database.

Quality Control

Quality control within NWSI is performed by visual inspection of the output image files, similar to the process used by ADNI. The image viewer can be used to inspect the image to determine whether segmentation of the structural image by FreeSurfer was subject to errors, such as truncated sections of the structural image, poor resolution, or contamination by noise in the system. In cases of such errors, it is still possible to isolate and study only those regions that segmented correctly. Depending on the outcome, each MRI is catalogued by Quality Control as Fail, Pass, Hippocampus-only, or Partial. Many studies implement normalization using global values, such as the intracranial volume, derived from FreeSurfer. If an image receives a Fail grade, it is not possible to include it in studies that depend on global values.

18F-Florbetapir Positron Emission Tomography or 18F-Florbetaben Positron Emission Tomography Analysis

Quantification of regions of interest (ROI) is still defined manually, but automatic standardized uptake value ratio (SUVR) calculations and segmentation of PET images have become the gold standard [18]. NWSI implements several PET analysis pipelines for FDG and 18F-Florbetapir images. Before uploading a PET scan, a related structural MRI must already exist in the system; the user is presented with a form in which an existing

MRI must be selected. After the PET scan is uploaded, it is copied to one of the RSs for processing and can be accessed through a form that lists all uploaded records, as shown in Figure 9. This form also contains graphs showing the distribution of all PET scans uploaded by the user as shown in Figure 10. Once a PET scan is processed, it can be displayed on the interactive image viewer, as shown in Figure 11 and with white and gray matter surfaces segmented as shown in Figure 12, and then quantitative data can be downloaded from the PET scan page, as shown in Figure 13.

Figure 8. Sample tabulated output for subcortical regions.

A	B	C	D	E	F	G	H	I
SITE_ID	Record_ID	Date	QC	Measure:v	Left-Lateral-Ventricle	Left-Inf-Lat-Vent	Left-Cerebellum-White-Matter	Left-Cerebellum-Cortex
2055	110025	4/6/2016	No Data	215_-1_-1	23778.1	270.1	15368.7	46407.1
2056	110026	4/6/2016	No Data	215_-1_-1	13807	438.2	13988.8	40207.9
2057	110033	4/12/2016	No Data	215_-1_-1	35533.9	3090	17376.9	47329
2058	110034	4/26/2016	No Data	215_-1_-1	37264.6	2386	4892	26241.9
2059	110038	6/17/2016	No Data	215_-1_-1	14027.4	574.6	12256.4	49011.8
2060	110041	4/8/2016	No Data	215_-1_-1	20431.8	540.6	14090.2	38588.9
2061	110042	5/12/2016	No Data	215_-1_-1	10929.7	531.8	18459.8	55887.7
2062	110044	5/19/2016	No Data	215_-1_-1	16508.6	455.8	16232.2	40618.9
2063	110045	5/23/2016	No Data	215_-1_-1	19612.1	510.2	10461.2	45045.4
2064	110046	4/28/2016	No Data	215_-1_-1	11629.1	299.8	12619.9	42953.7
2065	110047	5/12/2016	No Data	215_-1_-1	14649.8	765.2	14659.4	48621.5
2066	110048	5/10/2016	No Data	215_-1_-1	12113.8	509.7	14818.2	50881.8
2067	110049	4/26/2016	No Data	215_-1_-1	5548.2	184.8	15444	38776.4
2068	110051	3/30/2016	No Data	215_-1_-1	9556.8	310.2	11621.7	35095.4
2069	110052	5/2/2016	No Data	215_-1_-1	17536.9	1121	17532.4	49117.9
2070	110054	6/28/2016	No Data	215_-1_-1	26411.8	1401.7	16145.6	53678.7
2071	110055	4/20/2016	No Data	215_-1_-1	4972.2	145.7	14453	39841.8
2072	110057	5/18/2016	No Data	215_-1_-1	14633.8	252.3	13407.8	39170
2073	110059	4/15/2016	No Data	215_-1_-1	18519.3	896.8	13542.6	41373.6
2074	110060	5/4/2016	No Data	215_-1_-1	15615	870.9	13658.1	45631.9
2075	110073	5/11/2016	No Data	215_-1_-1	6686.7	274	17355	47082
2076	110074	5/10/2016	No Data	215_-1_-1	58060.7	4074.2	9181	46118.1
2077	110075	5/3/2016	No Data	215_-1_-1	22469.9	432.6	14291.5	45991.6
2078	110077	5/2/2016	No Data	215_-1_-1	18309.5	651	16066.1	52512.4

Figure 9. Page listing all positron emission tomographies (PETs) in the account, with links to the magnetic resonance imaging (MRI) used as reference.

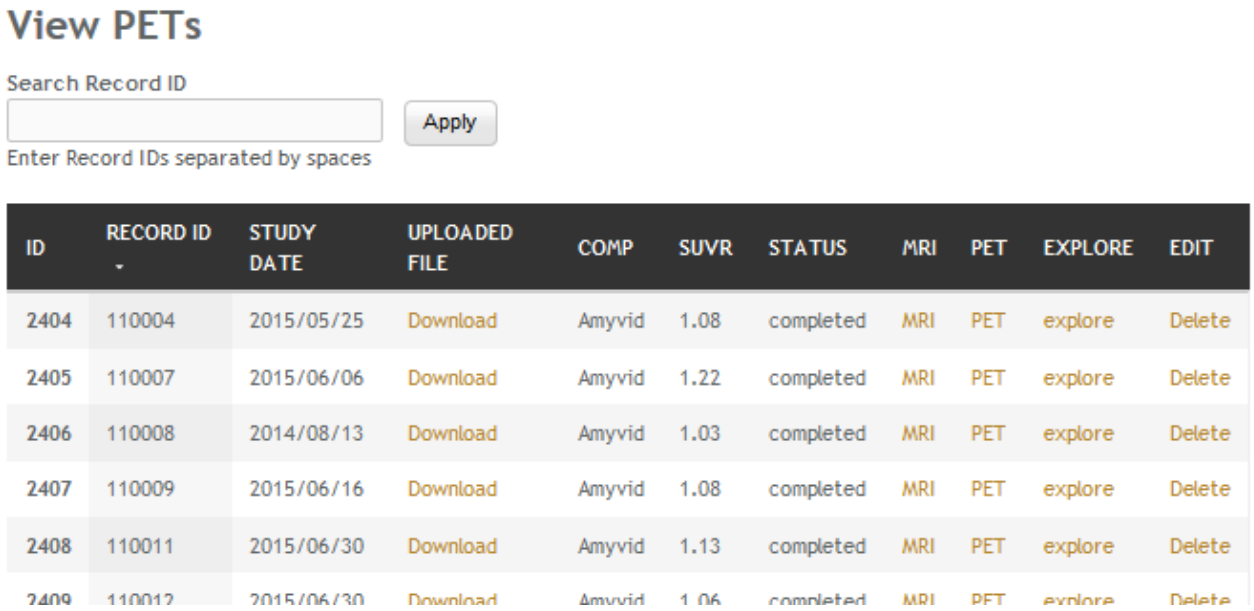


Figure 10. Graph showing the Standardized uptake value ratio (SUVR) distribution of all 18F-Florbetapir positron emission tomographies (PETs) uploaded by a user.

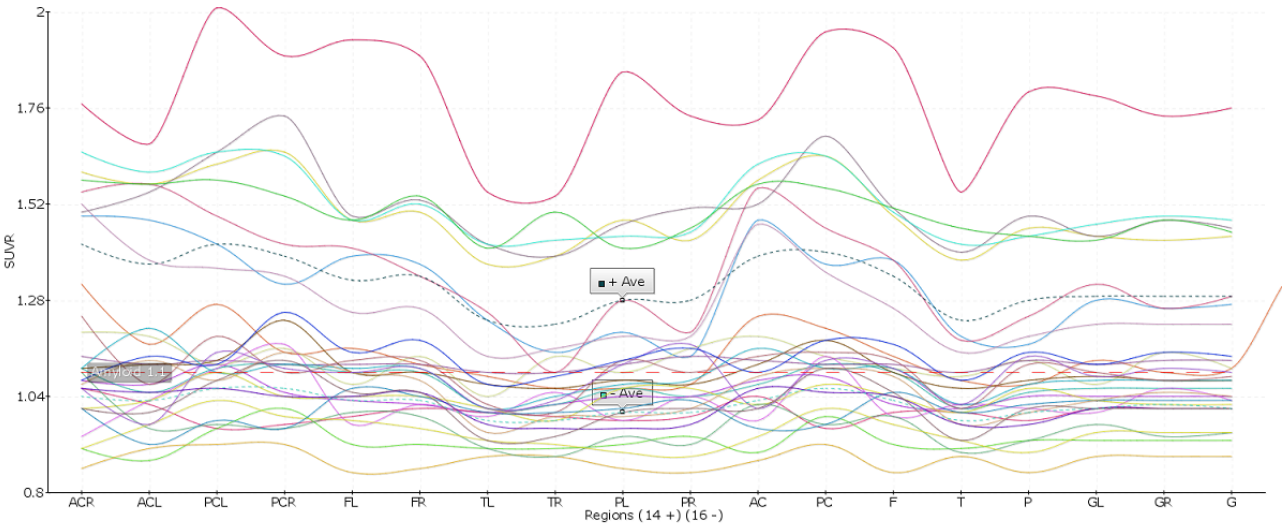


Figure 11. Higher concentration of 18F-Florbetapir shown in warmer colors of the spectrum look up Table (LUT).

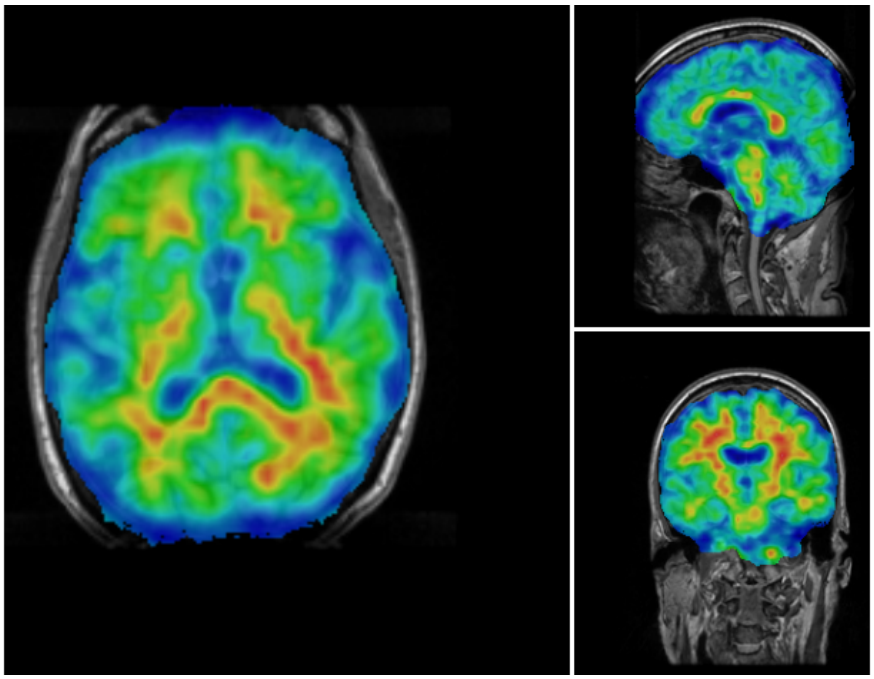


Figure 12. Positron emission tomography (PET) image is overlaid with the white matter surface (shown in White) and gray matter surface (shown in Black). Selected region's standardized uptake value ratio (SUVR) and name of region are displayed in Red at the bottom of the screen.

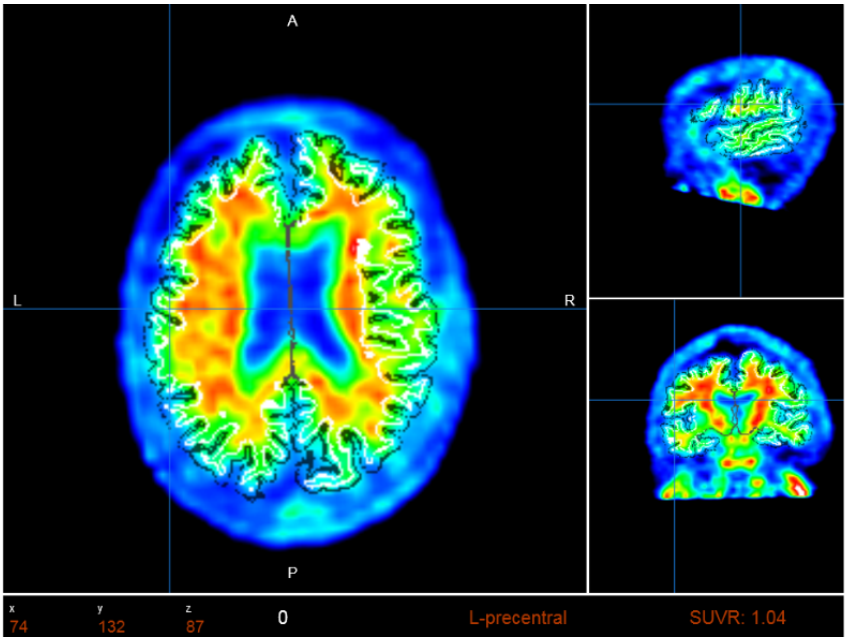
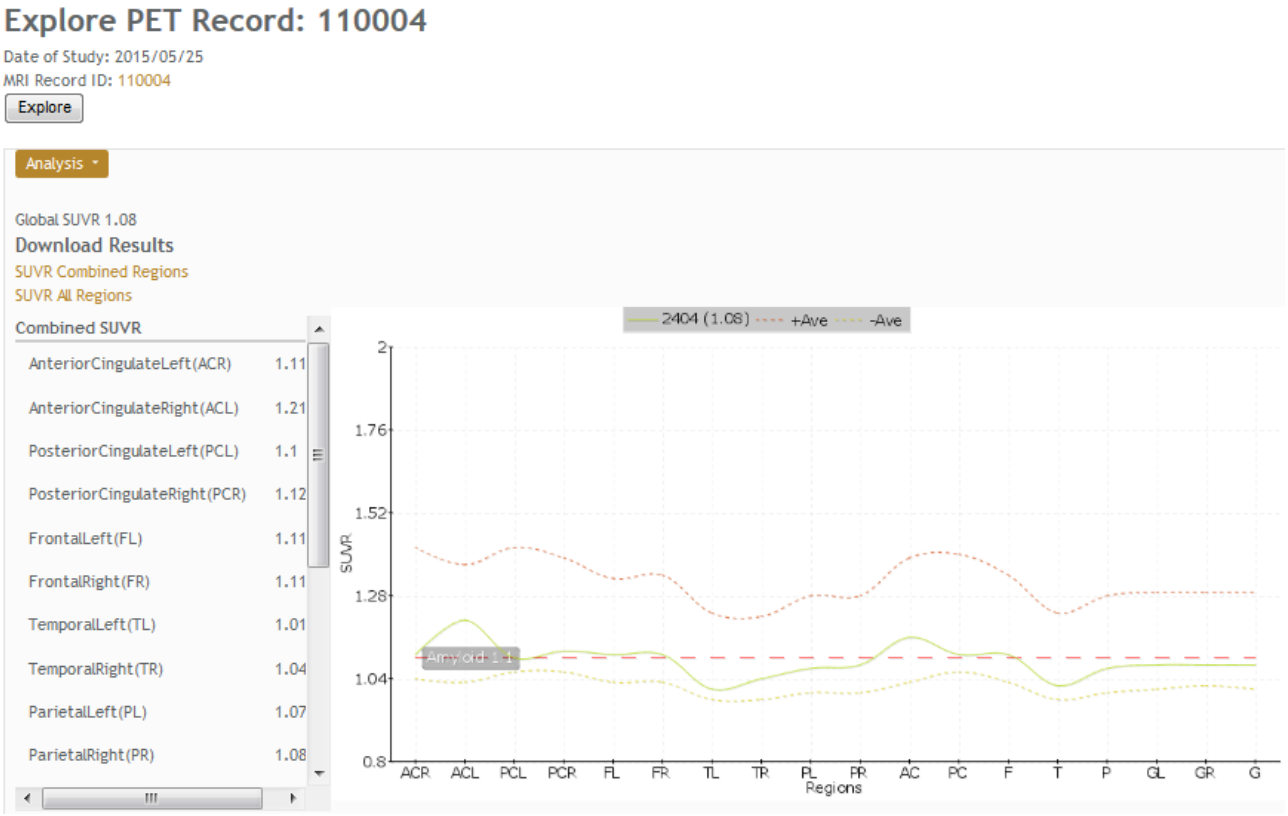


Figure 13. Positron emission tomography (PET) results page, which shows all standardized uptake value ratio (SUVR) values, links to download tabulated results, and a graph comparing the result to the averages determined in the system.



Magnetic Resonance Imaging-Positron Emission Tomography Image Registration

It is essential that the PET image is coregistered to the anatomical MRI since all calculations depend on how closely the anatomical regions of the 2 images overlap. A particular challenge for aging and AD and other neurodegenerative diseases is the issue of atrophy correction. It should be indicated that coregistration to MRI images largely reduces artifacts related to atrophy. The current implementation aligns the images using FSL. The alignment uses 12 degrees of freedom (3 translations, 3 rotations, 3 scalings, and 3 Skews or Shears). Then, a custom R script (R is an open source statistics software developed by Bell Laboratories) opens the coregistered PET and the FreeSurfer segmentation volume, performing a voxel by voxel analysis of the intensities of the PET file and

accumulating the values per anatomical ROI. In relation to these SUVR graphs, it is noted that the average over a region provides the *SUV* value over that region of interest (*ROI*) as given in Figure 14. The SUVR values are then determined by dividing the SUV_{ROI} by the SUV of a region of reference (SUV_{RG}) as expressed in Figure 15. Larger regions, aggregated from FreeSurfer subregions can also be calculated. SUVR, for combined FreeSurfer regions (CB), are calculated by a volume-weighted average of previously calculated SUVR as described in Figure 16.

Current literature reveals the merits of using an assortment of reference regions, such as the total or eroded subcortical white matter, the brain stem, the whole cerebellum, or the cerebellar white matter. SUVR results, normalized by the whole cerebellum, have been validated using the values reported by ADNI. NWSI also calculates SUVR using unilateral or bilateral cerebellar white matter. These values can be exported on tabulated files. PET imaging with 18F-Florbetaben also measures global cortical amyloid load and uses a similar processing pipeline to 18F-Florbetapir images. However, statistics and results on NWSI are reported separately to avoid bias. Studies have shown that there are no marked differences in the diagnostic accuracy of the amyloid binding ligand.

Fluorodeoxyglucose-Positron Emission Tomography in Epilepsy

PET imaging using FDG, labeled with a positron emitting tracer (Fluorine-18), or FDG-PET, is used to measure regional glucose metabolism, which is strongly correlated to the regional neuronal activity in the brain [19]. To study epileptic conditions using FDG-PET, in combination with structural MRI, regional SUVR is calculated using the cerebellar white matter, or the whole cerebellum, as a reference region [20]. In studying epilepsy, special consideration needs to be given to identifying focal conditions in one hemisphere and to account for surgical resection of regions in the brain. The FDG-PET pipeline in NWSI allows a choice of several reference regions, including the whole cerebellum, the cerebellar white matter, the average of all bilateral cortical regions (global cortical SUV), or all regions for a single intact hemisphere, especially in subjects who have had a prior resection in one hemisphere. The PET images are superimposed on MRI brain scans for defining the underlying structure and brain regions which have been resected. Regional SUVR is derived similar to the procedure for amyloid PET scans, and asymmetry in corresponding bilateral regions is calculated by dividing the difference in SUVR among corresponding bilateral regions by their sum and multiplying by 100%, as shown in Figure 17. Reference regions are not required in this calculation of asymmetry; a difference of 10% or greater between bilateral regions is typically considered to be consequential.

Figure 14. Formula to calculate the SUV value. N(ROI) is the total number of voxels in ROI(i) , and Intensity K is the value of voxel k in ROI (i) in PET. SUV: standardized uptake value; ROI: region of interest; PET: positron emission tomography.

$$SUV_{ROI_i} = \frac{\sum_{k=1}^{N_{ROI_i}} Intensity_k}{N_{ROI_i}}$$

Figure 15. Formula to calculate the regional SUVR values, by normalizing with SUV of reference region. SUVR: standardized uptake value ratio.

$$SUVR_{ROI_i} = \frac{SUV_{ROI_i}}{SUV_{RG}}$$

Figure 16. Formula to calculate the SUVR value of combined regions. CB represents combined regions, and V (ROI(i)) is the volume of the region. SUVR: standardized uptake value ratio; ROI: region of interest.

$$SUVR_{CB} = \frac{\sum_{i=1}^{m} SUVR_{ROI_i} \times V_{ROI_i}}{\sum_{i=1}^{m} V_{ROI_i}}$$

Figure 17. Formula to calculate the difference between the standardized uptake value ratio (SUVR) of bilateral regions. L and R represent Right and Left hemisphere, respectively.

$$Dif_{ROI_i} = \frac{SUVR_{ROI_i}{}^{L} - SUVR_{ROI_i}{}^{R}}{SUVR_{ROI_i}{}^{L} + SUVR_{ROI_i}{}^{R}} \times 100\%$$

Diffusion Tensor Imaging

DTI analysis on NWSI is still under development. Currently, DTI is obtained by executing the DTIFit FSL tool on an anatomically coregistered diffusion-weighted image (DWI) that has been corrected for Eddy currents. The DTIFit program fits a diffusion tensor model at each voxel. The resulting DTI eigenvalues and related eigenvectors, which reflect the direction of water movement and diffusion properties of a tissue, can be shown on the Web interface viewer modulated by the FA image. DTI is obtained from DWIs. A diffusion tensor is calculated for each voxel (3×3 matrix). The direction of the fibers is indicated by the tensor's eigenvector. The images are color coded, with Red, Blue, and Green; indicating right to left, foot to head, and anterior to posterior directions, respectively. As these images are coregistered to the anatomical MRI, further analysis can be done to obtain numeric values for anatomical ROIs (see Figures 18 and 19).

Data Convert, Registration, and Brain Extraction Tools

NWSI offers an assortment of other tools. The Data Convert (DC) tool, as shown in Figure 20, provides an interface to convert from DICOM and ANALYZE to NIfTI format. The DC tool also handles image compression, image orientation, and other issues which arise from transfer syntax in DICOM images. The interface is simple to use; it asks the user to upload a file and provides an identification field (record ID) and then converts the file to NIfTI. The output result can be downloaded and used as input to other forms within the site.

The CoRegistration Tool

This tool coregisters one brain image to another, from similar or different modalities. This form uses FSL tools to align the images and exports many of the options to the user, such as (1) extract the brain from source images before registration and (2) define the degrees of freedom, cost function, or angle to rotate the images. For most of the AD data processed in NWSI, default FSL registration parameters work well. However, some images have noise or artifacts and cannot be used for registration. CRT allows the user to find the registration parameters for individual images before they are uploaded to other processing pipelines. Current and previous results can be inspected in an embedded viewer. Coregistration is the main step to many processing pipelines, especially for multimodal imaging.

The Brain Extraction Tool

This tool interfaces with FSL and is extremely useful for extracting the brain structure from an image in any modality.

Evaluation and Validation

Magnetic Resonance Imaging Values

As cautioned earlier, FreeSurfer values obtained from the same FreeSurfer version can vary among different hardware and software platforms. Hence, for this proposed Web interface, the RSs are calibrated, making sure they always provide the same results before their deployment. ADNI renders FreeSurfer results calculated on FreeSurfer 5.1. In order to validate NWSI, a paired *t* test was performed to compare the values reported in the ADNIMerge file and the NWSI results. A total of 20 ADNI cases were selected at random from the 4 main diagnoses (AD; early mild cognitive impairment, EMCI; late MCI; and cognitively normal). One of the selected subjects failed ADNI's quality control for the Mid Temporal region, but it was successfully processed by NWSI. Table 1 shows the validation of the NWSI results in comparison with ADNI data, with ADNI being the gold standard for Alzheimer's MRI analysis.

Figure 18. Sample processed Alzheimer's disease (AD) diffusion tensor imaging (DTI). V1 modulated by fractional anisotropy (FA). Colors represent direction of water movement: Green is front to back. Blue is head to foot. Red is left to right.

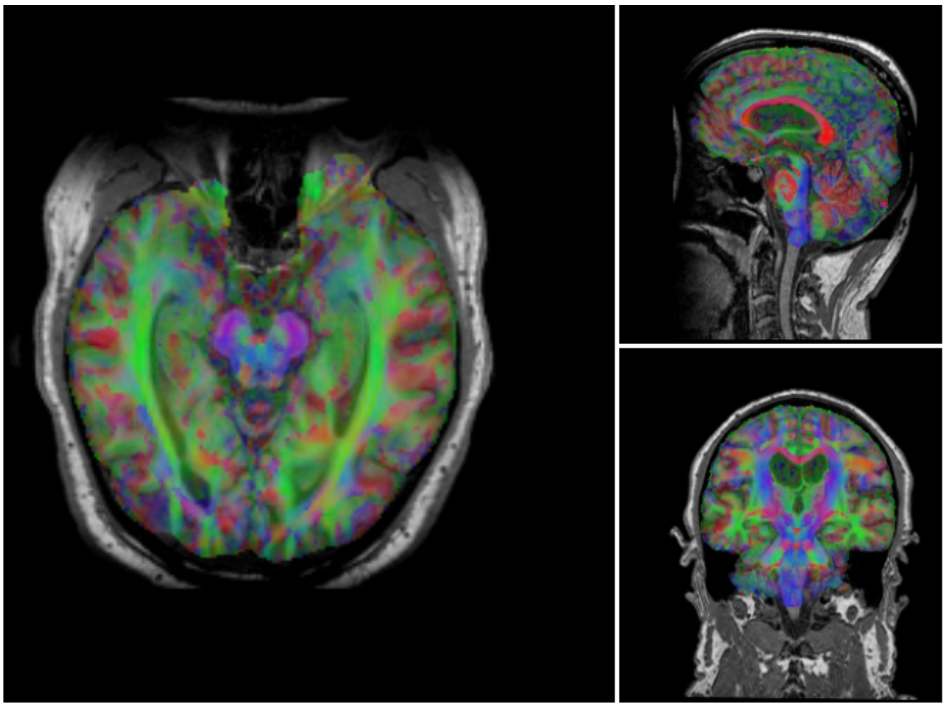


Figure 19. Close up of Diffusion Tensor Imaging (DTI) image showing the V1 eigenvectors pointing at the direction of water diffusion around the ventricle.

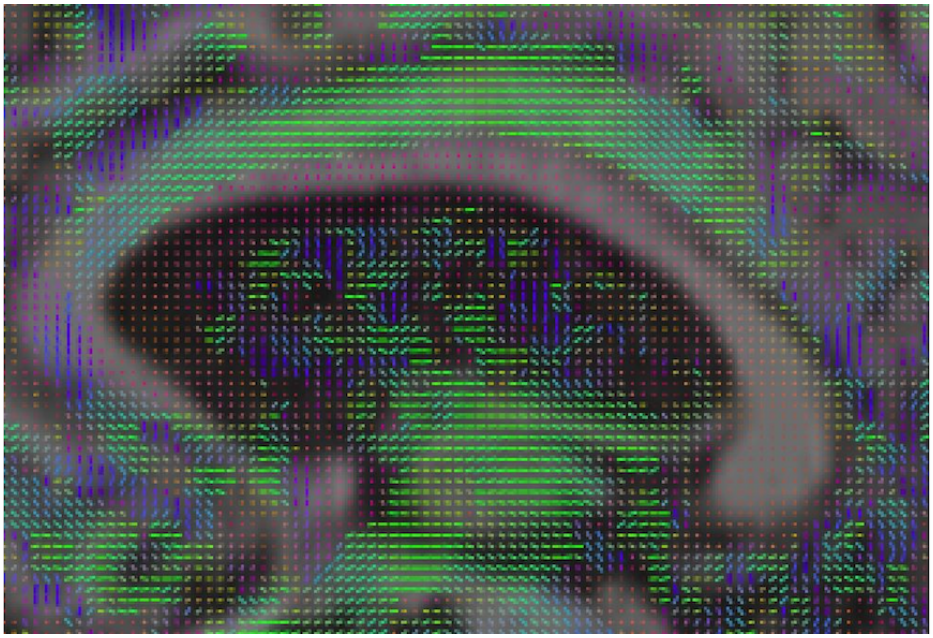


Figure 20. The data convert tool.

Table 1. Magnetic Resonance Imaging (MRI) processing comparison by Neuroimaging Web services interface (NWSI) quality control.

Region	NWSI[a] FreeSurfer 5.3	ADNI[b] Merge FreeSurfer 5.1	Difference (%)	Pearson correlation coefficient	Paired *t* test[c]
Hippocampus	7041 (SD 529)	7043 (SD 1404)	−3.00	.99 (*P*<. 001)	−0.03 (.9788)
Entorhinal cortex	3252 (SD 717)	3504 (SD 827)	−6.50	.88 (*P*<.001)	−2.8 (.01)
Middle temporal	19040	20020	−3.40	.97 (*P*<.001)	−5.3 (.0001)
Intracranial volume	1486538 (SD 154661)	1482872 (SD 152534)	2.40	.99 (*P*<.001)	−1.0 (0.33)

[a]NWFI: Neuroimaging Web Services Interface.

[b]ADNI: Alzheimer's Disease Neuroimaging Initiative.

[c]H0: Difference=0.

Table 2. Paired *t* test comparing Neuroimaging Web Services Interface (NWSI) and Alzheimer's Disease Neuroimaging Initiative (ADNI) Merge 18F-Florbetapir Positron Emission Tomography (PET) Global standardized uptake value ratio (SUVR) values.

Results	NWSI[a]	ADNIMERGE[b]
Mean	1.22	1.217
Variance (SUVR[c])	.047	.044
Observations	18	18
t Stat(degrees of freedom 17)	.26	—
P (T≤*t*) two-tail	.79	—
t critical two-tail (SUVR)	2.11	—

[a]NWSI: Neuroimaging Web Services Interface.

[b]ADNI: Alzheimer's Disease Neuroimaging Initiative.

[c]SUVR: standardized uptake value ratio.

ADNIMerge and NWSI FreeSurfer results are highly correlated. There is a small statistical difference for Mid Temporal and Entorhinal cortex. ADNIMerge was processed with FS5.1, and NWSI uses FS5.3. FreeSurfer 5.3 was a major upgrade to 5.1. Different FreeSurfer versions produce different results, but it does not imply lack of validity. Classification results can still be reliably used [21].

18F-Florbetapir Positron Emission Tomography

NWSI PET SUVR values have been validated with the values on ADNI data as reported by Jagust [22]. All ADNI PET results use 18F-Florbetapir. A random selection of 20 subjects from ADNI was used. Native PETs and MRIs were processed in NWSI, and the calculated global SUVR values were compared with the values reported by ADNI. Table 2 shows the paired *t* test results for comparing FLORBETAPIR (a PET radiopharmaceutical compound used for AD diagnostic) PET data obtained in NWSI, generated from ADNI cases, to the results provided by ADNI. This 2-tailed *t* test shows no statistical significance (*t* Stat<*t* Critical 2 tails) as can be seen from Table 2.

Cost

NWSI scalability allows a large number of additional RSs. The basic requirement is 2 servers—one hosting the Web interface and the other for processing images. This setup was the initial prototype which worked effectively for small batches of less than 20 or 30 sets uploaded simultaneously. The cost of maintaining 2 dedicated servers is low; services such as GoDaddy Operating Company provide each server from US $69.99 per month for a dedicated Linux server. The current prototype is installed on a distributed system located at Florida International University. This type of setup is more secure and easier to manage, and has a larger price tag. FIU (Florida International University) paid US $62,000, including hardware and software licenses. If all the resources are utilized, the FIU setup can service many requests, with the capability to process hundreds of MRIs and PETs per day. The prototype only uses 3 servers from up to 15 servers which can be created on this distributed system with 8 cores and 32 GB each.

Discussion

Principal Findings

The feedback provided by medical doctors (MDs) and other researchers has been invaluable. The system, as currently developed, is the result of many hours dedicated to understanding the domain of neuroimaging and the needs and requirements of MDs. The medical images displayed in the Papaya viewer contain layers defining the edge of the gray and white matter, custom LUTs created based on how MDs visualize the specific results, and many other enhancements that emerged from interacting with the different medical teams who have used NWSI.

Data entry into NWSI has been evaluated based on user feedback. The forms, as explained before, are intuitive and similar to many other forms on the Internet. Uploading an MRI into NWSI is as simple as updating a Facebook or Instagram status.

New studies are added every month to NWSI. As more data is uploaded, it is possible to create methods for merging similar data from different sources. This allows, for example, using the control scans from one account to enhance another study lacking controls. Multimodal pipelines can also be created based on merging PET, structural and functional MRI, and DTI. The current implementation did not include fMRI, but our colleagues at the University of Florida will be processing all resting state fMRIs to be included in this Web interface, and the DTI processing is currently limited to few cases, but the intent is to enhance this work with more cases to be processed over the next year. Expanding those pipelines will allow multimodal pipelines to be created for enhanced multimodal studies. New processing pipelines can be exposed to the user, allowing inclusion of previously processed cases and broadening the scope of new studies.

FreeSurfer and FSL were the natural choices for segmentation and registration, especially since ADNI data were readily available and already processed using the well-established FreeSurfer and FSL software. There are other software packages that also provide excellent results, such as 3D Slicer (an open source platform for medical image informatics developed by an international community), AFNI (Analysis of Functional

NeuroImages: a set of programs for processing, analyzing, and displaying functional MRI data, developed by the National Institute of Health), and SPM (Statistical Parametric Mapping is a software designed for the analysis of brain imaging data sequences, developed by members and collaborators of the Wellcome Trust Centre for Neuroimaging). For future development, it will be possible to add pipelines employing these software packages and for the user to select which one to use at the upload-form phase.

Conclusions

NWSI provides a platform for storing and processing neuroimaging data. All data are deidentified before being uploaded to the server. NWSI is accessible worldwide. The user interacts with the processing pipelines through a simple interactive Web interface, which allows the users to upload and process images of the brain and view the results directly on the browser. The multiuser interface allows privacy among researchers, as well as data sharing. Data are protected on the secured server, whereas communication with the user is encrypted. Pipelines that process structural MRI and amyloid PET scans have been validated with existing and well-established databases such as ADNI. NWSI stores all results in SQL tables and files, facilitating the selection and processing of existing data into new pipelines. As such, NWSI offers a complete solution for neuroimaging studies with multiuser tools for data processing and visualization, as well as for downloading to other platforms for further processing.

Acknowledgments

We are grateful for the continued support from the National Science Foundation (NSF) under NSF grants CNS-1532061, CNS-0959985, CNS-1551221, HRD- 0833093, and IIP 1338922. We also greatly appreciate the support of the Ware Foundation. This research is also supported through the Florida Department of Health, Ed and Ethel Moore Alzheimer's Disease Research Program, and the 1Florida ADRC (Alzheimer’s Disease Research Center, 1P50AG047266-01A1) and R01 AG047649-01A1.

Data collection and sharing for this project was funded by ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following—AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc; Cogstate; Eisai Inc; Elan Pharmaceuticals, Inc; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc; Fujirebio; GE Healthcare; IXICO Ltd; Janssen Alzheimer Immunotherapy Research and Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co, Inc; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institute of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Slideshow presentation describing the system.

[PPTX File, 3MB - medinform_v6i2e26_app1.pptx]

Multimedia Appendix 2

Video of system description.

[WMV File (Windows Media Video), 46MB - medinform_v6i2e26_app2.wmv]

References

1. Morita T, Asada M, Naito E. Contribution of neuroimaging studies to understanding development of human cognitive brain functions. Front Hum Neurosci 2016;10:464 [FREE Full text] [doi: 10.3389/fnhum.2016.00464] [Medline: 27695409]
2. Governing Council of the Organization for Human Brain. Neuroimaging databases. Science 2001 Jun 01;292(5522):1673-1676. [doi: 10.1126/science.1061041] [Medline: 11387466]
3. Van Horn JD, Grethe JS, Kostelec P, Woodward JB, Aslam JA, Rus D, et al. The functional Magnetic Resonance Imaging Data Center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. Philos Trans R Soc Lond B Biol Sci 2001 Aug 29;356(1412):1323-1339 [FREE Full text] [doi: 10.1098/rstb.2001.0916] [Medline: 11545705]

4. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Cedarbaum J, Alzheimer's Disease Neuroimaging Initiative. 2014 update of the Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimers Dement 2015 Jun;11(6):e1-120 [FREE Full text] [doi: 10.1016/j.jalz.2014.11.001] [Medline: 26073027]
5. Marenco LN, Wang R, Bandrowski AE, Grethe JS, Shepherd GM, Miller PL. Extending the NIF DISCO framework to automate complex workflow: coordinating the harvest and integration of data from diverse neuroscience information resources. Front Neuroinform 2014 May 28;8:58 [FREE Full text] [doi: 10.3389/fninf.2014.00058] [Medline: 25018728]
6. Malone IB, Cash D, Ridgway GR, MacManus DG, Ourselin S, Fox NC, et al. MIRIAD--Public release of a multiple time point Alzheimer's MR imaging dataset. Neuroimage 2013 Apr 15;70:33-36 [FREE Full text] [doi: 10.1016/j.neuroimage.2012.12.044] [Medline: 23274184]
7. Poldrack RA, Gorgolewski KJ. OpenfMRI: Open sharing of task fMRI data. Neuroimage 2017 Jan;144(Pt B):259-261 [FREE Full text] [doi: 10.1016/j.neuroimage.2015.05.073] [Medline: 26048618]
8. Gorgolewski KJ, Varoquaux G, Rivera G, Schwarz Y, Ghosh SS, Maumet C, et al. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front Neuroinform 2015 Apr 10;9:8 [FREE Full text] [doi: 10.3389/fninf.2015.00008] [Medline: 25914639]
9. Schirner M, Rothmeier S, Jirsa VK, McIntosh AR, Ritter P. An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data. Neuroimage 2015 Aug 15;117:343-357 [FREE Full text] [doi: 10.1016/j.neuroimage.2015.03.055] [Medline: 25837600]
10. Keator DB, Helmer K, Steffener J, Turner JA, Van Erp TG, Gadde S, et al. Towards structured sharing of raw and derived neuroimaging data across existing resources. Neuroimage 2013 Nov 15;82:647-661 [FREE Full text] [doi: 10.1016/j.neuroimage.2013.05.094] [Medline: 23727024]
11. Harrigan RL, Yvernault BC, Boyd BD, Damon SM, Gibney KD, Conrad BN, et al. Vanderbilt University Institute of Imaging Science Center for Computational Imaging XNAT: A multimodal data archive and processing environment. Neuroimage 2016 Jan 01;124(Pt B):1097-1101 [FREE Full text] [doi: 10.1016/j.neuroimage.2015.05.021] [Medline: 25988229]
12. Kargupta H. Data mining : next generation challenges and future directions. Menlo Park, California, United States: AAAI; 2004.
13. Fischl B. FreeSurfer. Neuroimage 2012 Aug 15;62(2):774-781 [FREE Full text] [doi: 10.1016/j.neuroimage.2012.01.021] [Medline: 22248573]
14. Larobina M, Murino L. Medical image file formats. J Digit Imaging 2014 Apr;27(2):200-206 [FREE Full text] [doi: 10.1007/s10278-013-9657-9] [Medline: 24338090]
15. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004;23(Suppl 1):S208-S219. [doi: 10.1016/j.neuroimage.2004.07.051] [Medline: 15501092]
16. Gronenschild EH, Habets P, Jacobs HI, Mengelers R, Rozendaal N, van Os J, et al. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS One 2012;7(6):e38234 [FREE Full text] [doi: 10.1371/journal.pone.0038234] [Medline: 22675527]
17. Pfefferbaum A, Rohlfing T, Rosenbloom MJ, Sullivan EV. Combining atlas-based parcellation of regional brain data acquired across scanners at 1.5 T and 3.0 T field strengths. Neuroimage 2012 Apr 02;60(2):940-951 [FREE Full text] [doi: 10.1016/j.neuroimage.2012.01.092] [Medline: 22297204]
18. Tuszynski T, Rullmann M, Luthardt J, Butzke D, Tiepolt S, Gertz HJ, et al. Evaluation of software tools for automated identification of neuroanatomical structures in quantitative β-amyloid PET imaging to diagnose Alzheimer's disease. Eur J Nucl Med Mol Imaging 2016 Jun;43(6):1077-1087. [doi: 10.1007/s00259-015-3300-6] [Medline: 26739328]
19. Sarikaya I. PET studies in epilepsy. Am J Nucl Med Mol Imaging 2015 Oct 12;5(5):416-430 [FREE Full text] [Medline: 26550535]
20. Miller-Thomas MM, Benzinger TL. Neurologic applications of PET/MR imaging. Magn Reson Imaging Clin N Am 2017 May;25(2):297-313. [doi: 10.1016/j.mric.2016.12.003] [Medline: 28390530]
21. Chepkoech JL, Walhovd KB, Grydeland H, Fjell AM, Alzheimer's Disease Neuroimaging Initiative. Effects of change in FreeSurfer version on classification accuracy of patients with Alzheimer's disease and mild cognitive impairment. Hum Brain Mapp 2016 May;37(5):1831-1841. [doi: 10.1002/hbm.23139] [Medline: 27018380]
22. Jagust WJ, Landau SM, Koeppe RA, Reiman EM, Chen K, Mathis CA, et al. The Alzheimer's disease neuroimaging initiative 2 PET core: 2015. Alzheimers Dement 2015 Jul;11(7):757-771 [FREE Full text] [doi: 10.1016/j.jalz.2015.05.001] [Medline: 26194311]

Abbreviations

AD: Alzheimer disease
ADNI: Alzheimer’s Disease Neuroimaging Initiative
BET: Brain Extraction Tool
CB: Combined FreeSurfer Regions

CT: computed tomography
DC: Data Convert
DTI: Diffusion Tensor Imaging
DWI: Diffusion-Weighted Image
EMCI: early mild cognitive impairment
FA: fractional anisotropy
FDG: fluorodeoxyglucose
fMRI: functional magnetic resonance imaging
fswiki: FreeSurferWiki
MCI: mild cognitive impairment
MD: medical doctor
MRI: magnetic resonance imaging
NWSI: Neuroimaging Web Services Interface
OS: operating system
PET: positron emission tomography
ROI: regions of interest
RS: replica server
SUV: standardized uptake values
SUVR: standardized uptake value ratio
VM: virtual machine
1Florida ADRC: 1Florida Alzheimer's Disease Research Center

Edited by G Eysenbach; submitted 28.09.17; peer-reviewed by F Wang, S Kim, A Yadollahpour, W You; comments to author 09.12.17; revised version received 12.01.18; accepted 29.01.18; published 26.04.18.

Please cite as:
Lizarraga G, Li C, Cabrerizo M, Barker W, Loewenstein DA, Duara R, Adjouadi M
A Neuroimaging Web Services Interface as a Cyber Physical System for Medical Imaging and Data Management in Brain Research: Design Study
JMIR Med Inform 2018;6(2):e26
URL: http://medinform.jmir.org/2018/2/e26/
doi:10.2196/medinform.9063
PMID:29699962

Original Paper

Secure Logistic Regression Based on Homomorphic Encryption: Design and Evaluation

Miran Kim[1], MSc, PhD; Yongsoo Song[2,3], PhD; Shuang Wang[1], MSc, PhD; Yuhou Xia[4], MSc; Xiaoqian Jiang[1], MSc, PhD

[1]Division of Biomedical Informatics, University of California, San Diego, San Diego, CA, United States

[2]Department of Mathematical Sciences, Seoul National University, Seoul, Republic Of Korea

[3]Department of Computer Science and Engineering, University of California, San Diego, San Diego, CA, United States

[4]Department of Mathematics, Princeton University, Princeton, NJ, United States

Corresponding Author:
Miran Kim, MSc, PhD
Division of Biomedical Informatics
University of California, San Diego
9500 Gilman Drive
San Diego, CA, 92093
United States
Phone: 1 8582499978
Email: mrkim@ucsd.edu

Abstract

Background: Learning a model without accessing raw data has been an intriguing idea to security and machine learning researchers for years. In an ideal setting, we want to encrypt sensitive data to store them on a commercial cloud and run certain analyses without ever decrypting the data to preserve privacy. Homomorphic encryption technique is a promising candidate for secure data outsourcing, but it is a very challenging task to support real-world machine learning tasks. Existing frameworks can only handle simplified cases with low-degree polynomials such as linear means classifier and linear discriminative analysis.

Objective: The goal of this study is to provide a practical support to the mainstream learning models (eg, logistic regression).

Methods: We adapted a novel homomorphic encryption scheme optimized for real numbers computation. We devised (1) the least squares approximation of the logistic function for accuracy and efficiency (ie, reduce computation cost) and (2) new packing and parallelization techniques.

Results: Using real-world datasets, we evaluated the performance of our model and demonstrated its feasibility in speed and memory consumption. For example, it took approximately 116 minutes to obtain the training model from the homomorphically encrypted Edinburgh dataset. In addition, it gives fairly accurate predictions on the testing dataset.

Conclusions: We present the first homomorphically encrypted logistic regression outsourcing model based on the critical observation that the precision loss of classification models is sufficiently small so that the decision plan stays still.

(JMIR Med Inform 2018;6(2):e19) doi:10.2196/medinform.8805

KEYWORDS

homomorphic encryption; machine learning; logistic regression; gradient descent

Introduction

Biomedical data are highly sensitive and often contain important personal information about individuals. In the United States, health care data sharing is protected by the Health Insurance Portability and Accountability Act [1], whereas biomedical research practitioners are covered under federal regulation governing the “Common Rule,” a federal policy that protects people who volunteer for federally funded research studies [2]. These policies set high standards on the protection of biomedical data and violations will lead to financial penalties and lost reputation. On the other hand, cloud computing, which significantly simplifies information technology environments, is the trend for data management and analysis. According to a recent study by Microsoft, nearly a third of organizations work with four or more cloud vendors [3]. The privacy concern, therefore, becomes a major hurdle for medical institutions to outsource data and computation to the commercial cloud. It is

imperative to develop advanced mechanisms to assure the confidentiality of data to support secure analysis in the cloud environment.

An intuitive solution is to train a model without accessing the data and only obtain the estimated model parameters in a global manner. Assuming summary statistics can be shared, this can be done in a joint manner and we have developed the "grid logistic regression" [4-6] to show the feasibility of estimating the global parameters from distributed sources (eg, by only exchanging gradients and Hessian matrices). However, there are still vulnerabilities in sharing even the summary statistics; for example, the difference in mean age between a cohort of n patients and another cohort of n –1 overlapped patients can reveal the actual age of a single patient.

Many medical decision-making systems rely on the logistic regression model [7-9]. However, to use them appropriately, we need to provide a sufficient sample, which requires a sample size calculation. Peduzzi et al [10] suggested a simple guideline for a minimum number of cases to include in the study: let p be the smallest of the proportions of negative or positive cases in the population and k the number of covariates (the number of independent variables), then the minimum number of cases to include is $N = 10 \cdot k / p$. For example, one has three covariates to be included in the model and the proportion of positive cases in the population is 0.2 (20%). The minimum number of cases required is $10 \cdot 3 / 0.20 = 150$. For rare disease studies with many variables, it is even harder to collect enough samples from a single institution to meet this goal. We need to circumvent the privacy barriers to feed the model with more samples from different sources. As shown in Figure 1, homomorphic encryption techniques can support typical secure computations (eg, secure outsourcing and secure multiparty computation) and mitigate the privacy risks by allowing all computation to be done in the encrypted format.

Graepel et al [11] shed light on machine learning with homomorphically encrypted data. The article discussed scenarios that are appropriate and inappropriate to exercise machine learning with homomorphic encryption techniques. The authors provided two examples: linear means classifier and linear discriminative analysis, which can be achieved by using low-degree polynomials in homomorphic encryption. However, these simple parametric models do not handle complex datasets well and they do not represent the mainstream machine learning technologies used in biomedical research [12,13]. Additional work was carried out by Bos et al [14] to demonstrate the feasibility of making a prediction on encrypted medical data in Microsoft's Azure cloud. However, instead of learning from the data, this model only makes predictions using learned logistic regression models in a privacy-preserving manner. Similarly, a more recent work called CryptoNets applied trained neural networks to encrypted data only for the evaluation purpose [15]. Related works are summarized in Table 1.

In the current literature, most similar to our work are Aono et al [16] and Mohassel et al [17], but they are also very different from ours in assumptions and methods. Aono et al introduced an approximation to convert the likelihood function into a low-degree polynomial and used an additive homomorphic encryption to aggregate some intermediary statistics [16]. However, their scenario relies on the client to decrypt these intermediary statistics so that it can minimize the parameters locally. This is not a completely secure outsourcing scenario as ours, which works on encrypted data to obtain encrypted parameters without any client involvement. Mohassel et al developed secure two-party computation protocols to conduct the stochastic gradient descent for solving logistic regression and neural network problems [17].

Figure 1. Two secure models: (a) secure storage and computation outsourcing and (b) secure model outsourcing.

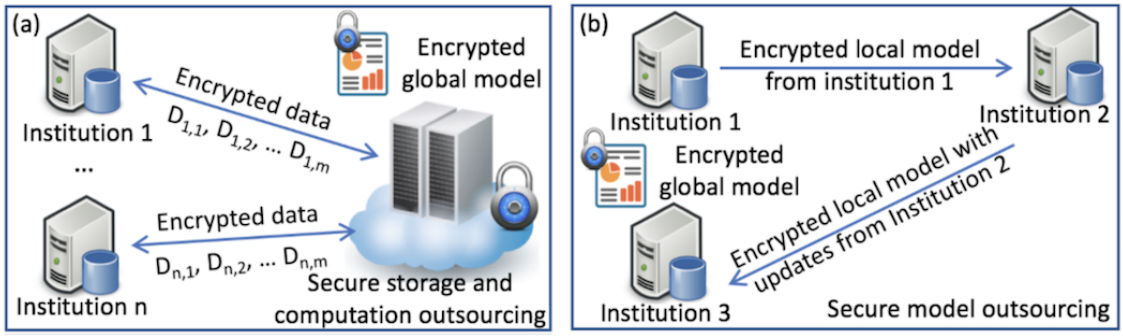


Table 1. Research works in secure analysis.

Reference	Problem	Techniques
Graepel et al [11]	Linear means classifier/discriminative analysis	Homomorphic encryption
Bos et al [14]	Prediction using learned logistic regression model	Homomorphic encryption
Dowlin et al [15]	Prediction using learned neural networks	Homomorphic encryption
Aono et al [16]	Logistic regression	Additive homomorphic encryption
Mohassel et al [17]	Logistic regression	Multiparty computation
This work	Logistic regression	Homomorphic encryption

This method takes a completely different approach (garbled circuit and secret sharing vs homomorphic encryption) and the assumptions are widely different from ours (secure multiparty computation vs secure outsourcing). There are several prominent challenges related to scalability and efficiency. Traditional methods cannot handle many iterations of multiplications, which leads to a deep circuit and exponential growth in computational cost and storage size of the ciphertext. On the other hand, it is a nontrivial task to approximate certain critical functions in machine learning models using only low-degree polynomials. Naive approximation may lead to big errors and makes the solutions intractable. Our framework proposes novel methods to handle these challenges and makes it possible to learn a logistic regression model on encrypted data based completely on homomorphic encryption.

Methods

Logistic Regression

Logistic regression is a widely used learning model in biomedicine [13]. Data for supervised learning consist of pairs (x_i, y_i) of a vector of covariates $x_i = (x_{i\,1},..., x_{i\,d})$ and a class label y_i for $i = 1,..., n$. We assume that $y_i = 1 / -1$ for binary classification. The model looks like:

for the sigmoid function $\sigma(x) = 1 / [1 + \exp(-x)]$ where $\beta = (\beta_0, \beta_1,..., \beta_d)$ are the model parameters to be estimated. Training methods of logistic regression aim to find the optimal parameters β, which minimizes the cost (negative log-likelihood)

Homomorphic Encryption for Approximate Arithmetic

Homomorphic encryption is an encryption technique that allows computations on ciphertexts and generates encrypted results that match those of plaintext computation. We adopted a special cryptosystem developed by Cheon et al [18], which supports an approximate arithmetic of encrypted messages. Different from existing methods, this cryptosystem trades precision for efficiency so that the size of parameters does not grow too large (thus computationally feasible). Interested readers can refer to Multimedia Appendix 1 for more details. The cryptosystem supports key generation, encryption, decryption, addition, and multiplication operations. It also supports message packing and rotation, which are important to parallelize similar tasks.

A unique property of this cryptosystem is the following rescaling procedure, which plays an important role in controlling the magnitude of messages and, therefore, achieving the efficiency of approximate computation. The rescaling procedure coverts an encryption ct of a message m with a ciphertext modulus q into an encryption ct' of $r^{-1} \cdot$ m under the same secret key but a smaller modulus $q' = r^{-1} \cdot q$, in which r is a scaling factor. We denote the output ciphertext by RS(ct; r). It enables us to round the message and reduce the size of significand by removing some inaccurate least significant bits as in the floating-point arithmetic. Informally, we will say that the input ciphertext modulus is reduced by log r bits after this procedure where the binary logarithm will be simply denoted by log(·).

Least Squares Approximation of the Sigmoid Function

Unlike linear regression, logistic regression does not have a closed-form solution in most cases. As a result, we need to use nonlinear optimization methods to find the maximum likelihood estimators of the regression parameters. The Newton-Raphson [19] and the gradient descent [20] are the most commonly used methods for training. Because the Newton-Raphson method involves matrix inversion and most homomorphic encryption schemes do not naturally support division or matrix inversion, it is difficult to evaluate the method with homomorphic encryption schemes. On the other hand, gradient descent does not require the division operation and, therefore, it is a better candidate for homomorphically encrypted logistic regression. Thus, we choose the gradient descent algorithm as the training method for logistic regression.

Let (x_i, y_i) be the supervised learning samples for $i = 1,..., n$. If we write $z_i = y_i \cdot (1, x_i)$, the cost function for logistic regression is defined by:

Its gradient with respect to β is computed by $-1 / n\Sigma_{1\le i\le n}\sigma(-z_i^T\beta) \cdot z_i$. To find a local minimum point, the gradient descent method updates the regression parameters using the following formula until β converges:

where α is the learning rate.

Although the gradient descent method seems better suited than other training methods for homomorphic evaluation, some technical problems remain for implementation. In the preceding update formula, the sigmoid function is the biggest obstacle for evaluation, since the existing homomorphic encryption schemes only allow evaluation of polynomial functions. Hence, the Taylor polynomials $T_d(x) = \Sigma_{0\le k\le d}(f^{(k)}(0) / k!) \cdot x^k$ have been commonly used for approximation of the sigmoid function [14,17]:

However, we observed the input values $z_i^T\beta$ of the sigmoid function during iterations on real-world datasets and concluded that the Taylor polynomial $T_9(x)$ of degree 9 is still not enough to obtain the desired accuracy (see Figure 2a). The size of error grows rapidly as $|x|$ increases. For instance, we have $T_9(4) \approx 4.44$, $T_9(6) \approx 31.23$, and $T_9(8) \approx 138.12$. In addition, we have to use a higher degree Taylor polynomial to guarantee the accuracy of regression, but it requires too many homomorphic multiplications to be practically implemented. In summary, the Taylor polynomial is not a good candidate for approximation because it is a local approximation near a certain point. Therefore, we adopted a global approximation method that minimizes the mean squared error (MSE). For an integrable function f, its mean square over an interval I is defined by

$(1 / |I|) \int_I f(x)^2 dx$, where $|I|$ denotes the length of I. The least squares method aims to find a polynomial $g(x)$ of degree d which minimizes the MSE $(1 / |I|) \int_I (f(x) - g(x))^2 dx$. The least squares approximation has a closed formula that can be efficiently calculated using linear algebra.

In our implementation, we used the degree 3 and 7 least squares approximations of the sigmoid function over the interval [–8,8], which contains all of the input values $(-z_i^T\beta)$ during iterations. The least squares polynomials are computed as:

where the coefficients vectors are $(a_1, a_3) \approx (1.20096, -0.81562)$ and $(b_1, b_3, b_5, b_7) \approx (1.73496, -4.19407, 5.43402, -2.50739)$. The degree 3 least squares approximation requires a smaller depth for evaluation, whereas the degree 7 polynomial has a better precision (see Figure 2b).

Homomorphic Evaluation of Gradient Descent Algorithm

We will describe how to encode data and explain how to analyze logistic regression on encrypted data. To speed up the computation, we will use the packing mechanism to batch n slots and perform n evaluations in parallel, where n is the number of training data samples.

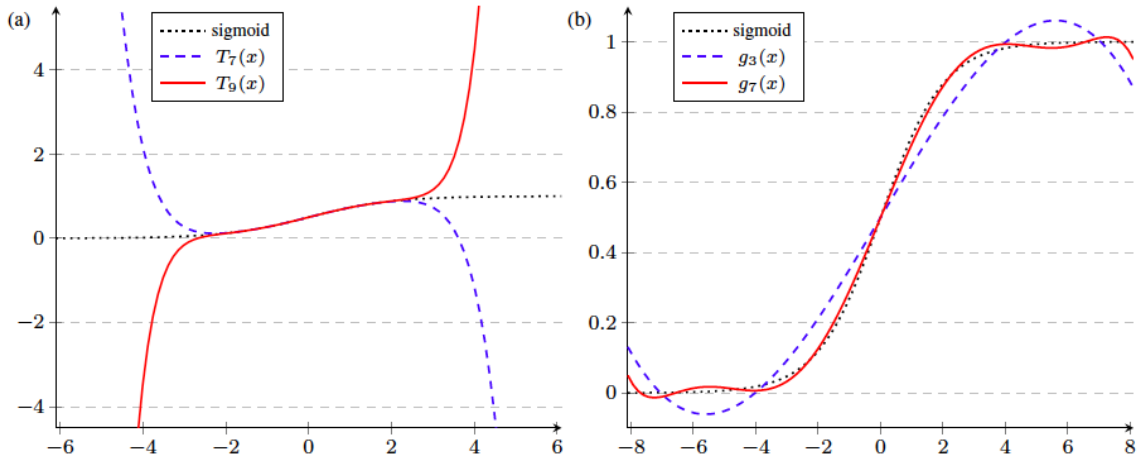
We start with a useful aggregation operation across plaintext slots from the literature [21-23]. Specifically, given a ciphertext representing a plaintext vector $(m_1, m_2, ..., m_k)$, we introduce an algorithm (denoted by AllSum) that generates a ciphertext representing a value of $\Sigma_{1 \le i \le k} m_i$ in each plaintext slot. Assume that k is chosen as a power-of-two integer. The cyclic rotation by one unit produces a ciphertext encrypting the plaintext vector $(m_2, ..., m_k, m_1)$. Then an encryption of the vector $(m_1 + m_2, m_2 + m_3, ..., m_k + m_1)$ is obtained by adding the original ciphertext. We repeatedly apply this method $(\log k - 1)$ times with a rotation by a power of two, which generates the desired ciphertext; that is, every plaintext slot contains the same value of $\Sigma_{1 \le i \le k} m_i$. The AllSum algorithm is explicitly described in Textbox 1.

Let us assume that we are given n training data samples z_i with $(d + 1)$ features. As mentioned before, our goal is to securely evaluate the following arithmetic circuit:

where $g(x)$ denotes the approximating polynomial of the sigmoid function chosen in the previous subsection. We set the initial β parameters as the zero vector for simplicity.

Because our cryptosystem only supports integer computation, all the elements are scaled by a factor of an integer p and then converted into the nearest integers for quantization. The client first receives the ciphertexts encrypting the vector $(p \cdot z_i)$ from n users, and then compromises them to obtain $(d + 1)$ ciphertexts $ct.z_i$ for all $j = 0, 1, ..., d$, each of which encrypts the vector $p \cdot (z_{1j}, ..., z_{nj})$ of the j-th attributes using batching technique. If n is not a power of two, the plaintext slots are zero padded so that the number of slots divides $N / 2$. Finally, these resulting ciphertexts $(ct.z_0, ..., ct.z_d)$ are sent to the server for the computation of gradient descent.

Figure 2. Graphs of (a) sigmoid function and Taylor polynomials and (b) sigmoid function and least squares approximations.



Textbox 1. The AllSum algorithm.

0: **Inputs:** ciphertext ct encrypting plaintext vector $(m_1, m_2, ..., m_k)$.

1: **For** $i = 0, 1, ..., \log k - 1$ **do**

2: Compute $ct \leftarrow \text{Add}(ct, \text{Rot}(ct; 2^i))$

3: **end for**

4: **Outputs:** ciphertext ct encrypting $\Sigma_{1 \le i \le k} m_i$ in each plaintext slot

Textbox 2. Secure logistic regression algorithm.

0: **Inputs:** Ciphertexts $\{ ct.z_j \}_{0\le j\le d}$, a polynomial $g(x)$, a number of iterations *IterNum*

1: **For** $j = 0,1,\ldots, d$ **do**

2: $ct.beta_j \leftarrow \mathbf{0}$

3: **end for**

4: **For** $k = 1,2,\ldots,$ *IterNum* **do**

5: $ct.ip \leftarrow \mathrm{RS}(\sum_{0\le j\le d}\mathrm{Mult}(ct.beta_j, ct.z_j); p)$

6: $ct.g \leftarrow \mathrm{PolyEval}(-ct.ip,\ p \cdot g(x))$

7: **For** $j = 0,1,\ldots, d$ **do**

8: $ct.grad_j \leftarrow \mathrm{RS}(\mathrm{Mult}(ct.g, ct.z_j); p)$

9: $ct.grad_j \leftarrow \mathrm{RS}(\mathrm{AllSum}(ct.grad_j);\ n/\alpha)$

10: $ct.beta_j \leftarrow \mathrm{Add}(ct.beta_j, ct.grad_j)$

11: **end for**

12: **end for**

13: **Outputs:** Ciphertexts $\{ ct.beta_{>}\, j \}_{0\le j\le d}$

The public server generates the initial ciphertexts ($ct.beta_0,\ldots, ct.beta_d$) as zero polynomials in R_q (the residue ring of $R = Z[X] / (X^N + 1)$ modulo an integer q). At each iteration, it performs a homomorphic multiplication of ciphertexts $ct.beta_j$ and $ct.z_j$, and outputs a ciphertext encrypting the plaintext vector $p^2 \cdot (z_{1j}{}^T\beta_j,\ldots, z_{nj}{}^T\beta_j)$ for all $j = 0,\ldots, d$. Then it aggregates the ciphertexts and performs the rescaling operation with a scaling factor of p to manipulate the size of plaintext, returning a ciphertext $ct.ip$ that represents a plaintext vector approximating to $p \cdot (z_1{}^T\beta,\ldots, z_n{}^T\beta)$.

For the evaluation of the least squares polynomial $g(x)$ at $(-z_i{}^T\beta)$, we adapt the polynomial evaluation algorithm, denoted by PolyEval(·), suggested in [18]. Each coefficient of the polynomial should be scaled by a factor of p to be transformed into an integral polynomial. The output ciphertext $ct.g$ contains $p \cdot g(-z_i{}^T\beta)$ in the i-th slot. Finally, the server performs a homomorphic multiplication of the ciphertexts $ct.g$ and $ct.z_j$, AllSum procedure, and rescaling by a factor of n/α (nearest integer to n/α). These procedures generate ciphertexts $ct.grad_0,\ldots, ct.grad_d$ corresponding to the entries of the gradient vector weighted by the learning rate and the sample size. Then it only needs to perform an addition with the model parameters β and the gradient vector over encryption, which yields a new ciphertext $ct.beta_j$ that encrypts an approximation of the j-th scaled value of the gradient update in Equation 7. Our secure logistic regression algorithm is described in Textbox 2.

Our solution can compute the gradient descent algorithm securely; however, its direct implementation is not efficient and requires a total ciphertext modulus of $\log p \cdot (\log \deg(g) + 3) + \log(n/\alpha)$ bits at each iteration, where x denotes the smallest integer that is not less than x. We further optimized this algorithm by manipulating the arithmetic circuit for the update term $(\alpha / n)\Sigma_{1\le i\le n} g(-z_i{}^T\beta) \cdot z_i$ and could reduce the ciphertext modulus to $3 \cdot \log p + \log(n/4\alpha)$ bits or $4 \cdot \log p + \log(n/4\alpha)$ bits when $g(x) = g_3(x)$ or $g(x) = g_7(x)$, respectively. Interested readers can refer to Multimedia Appendix 2 for more details.

Results

Implementation Details

All experiments were performed on an Intel Xeon running at 2.3 GHz processor with 16 cores and 64 GB of RAM, which is an m4.4xlarge AWS EC2 instance. In our implementation, we used a variant of a fixed-point homomorphic encryption scheme of Cheon et al [18,24] with C++-based Shoup's Number Theory Library [25]. Our implementation is publicly available at GitHub [26].

Datasets

We developed our approximation algorithm using the Myocardial Infarction dataset from Edinburgh [27]. The others were obtained from Low Birth Weight Study, Nhanes III, Prostate Cancer Study, and Umaru Impact Study datasets [28-31]. All these datasets have a single binary outcome variable, which can be readily used to train binary classifiers such as logistic regression. Table 2 illustrates the datasets with the number of observations (rows) and the number of features (columns), respectively. We utilized five-fold cross-validation that randomly partitions the original datasets into five folds with the approximately equal size; we used four subsets for learning (with the learning rate $\alpha \approx 1$) and one subset for testing the trained model..

Parameters and Timings for the Homomorphic Encryption Scheme

We assumed that all inputs had $\log p = 28$ bits of precision and set the bit length of the output ciphertext modulus as $\log q_0 = \log p + 10$. As discussed previously, when evaluating

the gradient descent algorithm with $g(x) = g_7(x)$, a ciphertext modulus is reduced more than $g(x) = g_3(x)$ at each iteration. Thus, we set the number of iterations as *IterNum* = 25 (resp *IterNum* = 20) when $g(x) = g_3(x)$ (resp. $g(x) = g_7(x)$) to take an initial ciphertext modulus of similar size. We could actually obtain the approximate bit length of fresh ciphertext modulus log q around 2204 to 2406. The parameter set provides 80 bits of security (see Multimedia Appendix 3 for more details). Because all the computations were performed on encrypted data, the security against a semi-honest adversary follows from the semantic security of the underlying homomorphic encryption scheme. For this setting, the size of the public key and a freshly encrypted ciphertext is 75 MB. The key generation takes approximately 56 to 58 seconds and the encryption takes approximately 1.1 to 1.3 seconds.

In Table 3, we evaluated our models performance based on average running time (encryption, evaluation, and decryption) and storage (encrypted dataset size) in each fold.

We used a popular metric, area under the receiver operating characteristic curve (AUC), to measure the model's classification performance when the true positive rate was plotted against the false positive rate at various thresholds. Figure 3 plots the average AUC values from five-fold cross-validation for datasets. The program was implemented by MATLAB 2017a.

We can converge to the optimum within a small number of iterations (20~25), which makes it very promising to train a homomorphically encrypted logistic regression model and mitigate the privacy concerns.

In Table 4, we compared the produced models using our encrypted approach and unencrypted logistic regression. In the unencrypted cases, we used the original sigmoid function on the same training dataset with the same iteration numbers as the encrypted cases. For discrimination, we calculated the accuracy (%), which is defined by the percentage of the correct predictions on the testing dataset. For a more accurate comparison, we used the MSE that measures the average of the squares of the errors. We could also normalize it by dividing by the average of the squares of the (unencrypted) model parameters, called a normalized mean squared error (NMSE).

Table 2. Description of datasets.

Dataset	Number of observations	Number of features
Edinburgh Myocardial Infarction	1253	10
Low Birth Weight Study	189	10
Nhanes III	15,649	16
Prostate Cancer Study	379	10
Umaru Impact Study	575	9

Table 3. Experiment results of our homomorphic encryption-based logistic regression algorithm

Dataset and degree of $g(x)$	Encryption (sec)	Evaluation (min)	Decryption (sec)	Storage (GB)
Edinburgh Myocardial Infarction				
3	12	131	6.3	0.69
7	12	116	6.0	0.71
Low Birth Weight Study				
3	11	101	4.9	0.67
7	11	100	4.5	0.70
Nhanes III				
3	21	265	12	1.15
7	21	240	13	1.17
Prostate Cancer Study				
3	11	119	4.4	0.68
7	11	100	4.5	0.70
Umaru Impact Study				
3	10	109	5.1	0.61
7	10	94	4.3	0.63

Figure 3. Average AUC of encrypted logistic regression. FPR: false positive rate; TPR: true positive rate.

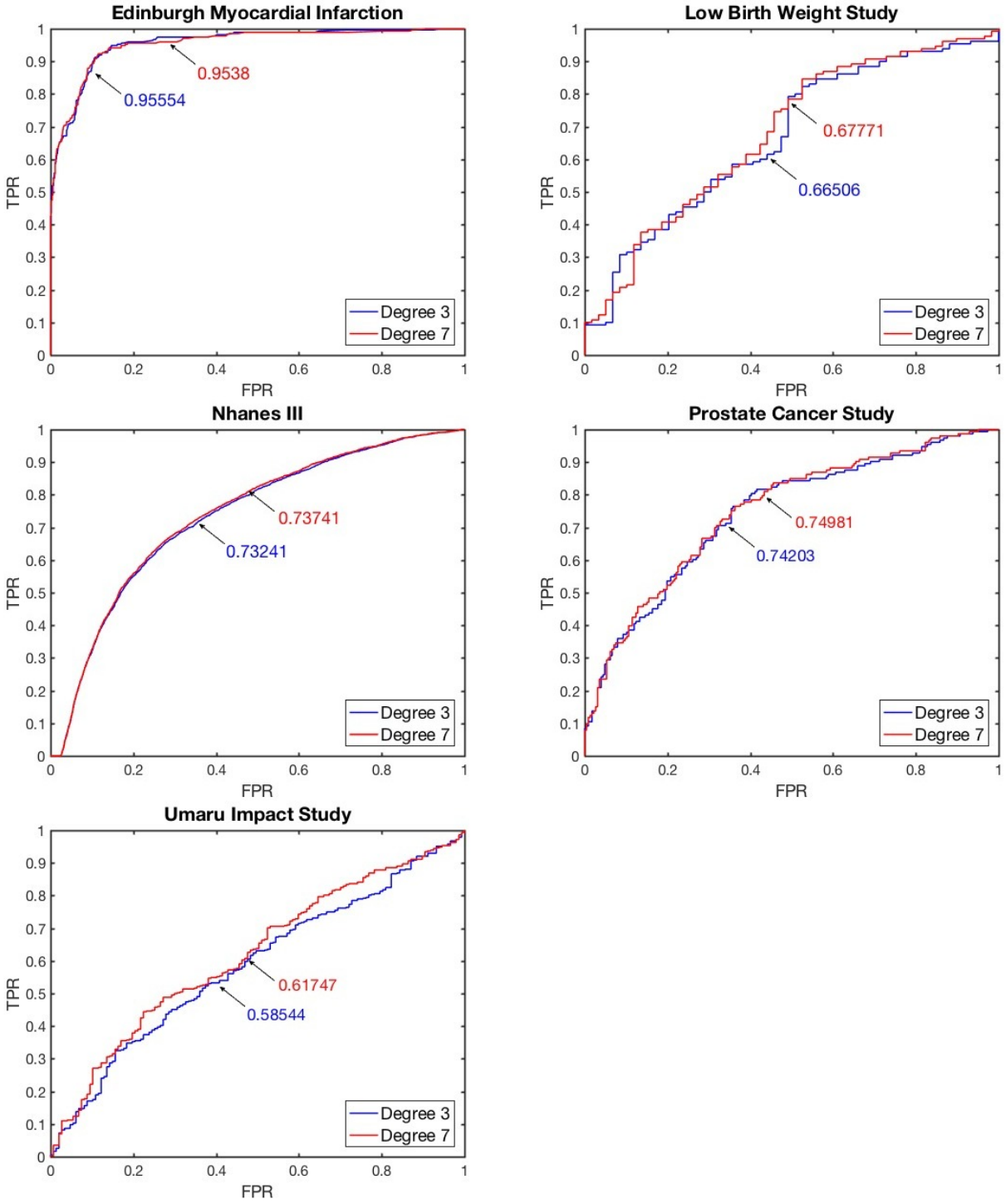


Table 4. Comparison of encrypted/unencrypted logistic regression. AUC: area under the receiver operating characteristic curve. MSE: mean squared error; NMSE: normalized mean squared error.

Dataset and iteration number	Degree of $g(x)$	Our homomorphic encryption-based logistic regression		Unencrypted logistic regression		MSE	NMSE
		Accuracy	AUC	Accuracy	AUC		
Edinburgh Myocardial Infarction							
25	3	86.03%	0.956	88.43%	0.956	0.0259	0.0261
20	7	86.19%	0.954	86.19%	0.954	0.0007	0.0012
Low Birth Weight Study							
25	3	69.30%	0.665	68.25%	0.668	0.0083	0.0698
20	7	69.29%	0.678	69.29%	0.678	0.0003	0.0049
Nhanes III							
25	3	79.23%	0.732	79.26%	0.751	0.0033	0.0269
20	7	79.23%	0.737	79.23%	0.737	0.0002	0.0034
Prostate Cancer Study							
25	3	68.85%	0.742	68.86%	0.750	0.0085	0.0449
20	7	69.12%	0.750	69.12%	0.752	0.0002	0.0018
Umaru Impact Study							
25	3	74.43%	0.585	74.43%	0.587	0.0074	0.0829
20	7	75.43%	0.617	74.43%	0.619	0.0004	0.0077

Discussion

Principal Findings

Our implementation shows that the evaluation of the gradient descent algorithm with the degree 7 least squares polynomial yields better accuracy and AUC than degree 3. It is quite close to the unencrypted result of logistic regression using the original sigmoid function with the same number of iterations; for example, on the training model of Edinburgh dataset, we could obtain the model parameters β as follows:

(–1.7086, 0.0768, 0.1119, 0.3209, 1.2033,
0.3684, 0.9756, 0.2020, 0.2259, –0.1641),

which can reach 86.19% accuracy and 0.954 AUC on the testing dataset. When using the sigmoid function on the same training dataset, the model parameters β are

(–1.6308, 0.0776, 0.1097, 0.3155, 1.1809,
0.3651, 0.9599, 0.2083, 0.2298, –0.1490),

which give the same accuracy and AUC. On the other hand, as shown in Table 4, the MSE and NMSE values of degree 7 are closer to zero which inspires us that the polynomial approximation of that degree is fairly accurate for logistic regression.

One of the inherent properties of our underlying homomorphic encryption scheme is that the inserted errors for security may increase after some homomorphic operations. Hence, the size of error and the precision loss should be discussed carefully to guarantee the correctness of the resulting value. On the other hand, the gradient descent method has a property of negative feedback on computational error. Because we use the gradient at the current weight vector β to move it closer to the optimal point of minimized cost, the effect of noise disappears after some iterations. Therefore, there is no need to manage the precision of messages to confirm the correctness of resulting value because the noises are not amplified during evaluation. In our experimentation on the Edinburgh dataset, for instance, the difference between the model parameters obtained from encrypted/unencrypted evaluations was less than 2^{-11}. This means that we can precisely compute at least most significant 11 bits after the radix point of the model parameters and this approximate vector is accurate enough to achieve a good performance in testing data samples.

Limitations

There are still a number of limitations in the application of our evaluation model to an arbitrary dataset. First, the use of homomorphic encryption yields the overheads in computation and storage. The size of the dataset should be limited for practical evaluation, but this is not a big problem because there have been significant improvements in the existing homomorphic encryption schemes. The development of homomorphic encryption technology will achieve much better practical performance in our protocol.

Another issue arises from the polynomial approximation. We suggested the least squares method on a certain interval [–8,8], but the precision of the result can increase by managing approximation error from wider range inputs. Finally, our model is based on fixed hyperparameters that should be decided before starting of the evaluation. It would be highly beneficial if we could detect convergence of the loss function in the training process and support early stop instead.

Conclusions

This paper presents the first effective methodology to evaluate the learning phase of logistic regression using the gradient descent method based on homomorphic encryption. We have demonstrated the capability of our model across the experiments with different biological datasets. In particular, our solution can be applied to a large-scale dataset, which shows the feasibility of our approach.

Acknowledgments

The authors would like to thank Kristin Lauter for helpful discussions and suggestions. The authors would also like to thank Andrey Kim for extensive assistance with the code for the homomorphic encryption scheme.

This research of MK, SW, and XJ was supported in part by the National Institute of Health under award numbers R00HG008175, R01GM118574, R01GM118609, and U01EB023685. YS was supported by the National Research Foundation of Korea grant funded by the Korean Government (No: 2017R1A5A1015626).

Authors' Contributions

MK led the algorithm development and the writing of the methodology. YS, YX, SW, and XJ contributed to the approximation algorithm and evaluation. YS also developed the parallelization for the proposed protocol. XJ and SW motivated the study and blended novel algorithms and new homomorphic schemes to enable secure learning. All authors carefully reviewed and edited the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Homomorphic encryption for approximate arithmetic.

[PDF File (Adobe PDF File), 91KB - medinform_v6i2e19_app1.pdf]

Multimedia Appendix 2

Further optimization of secure logistic regression algorithm.

[PDF File (Adobe PDF File), 128KB - medinform_v6i2e19_app2.pdf]

Multimedia Appendix 3

How to set parameters.

[PDF File (Adobe PDF File), 65KB - medinform_v6i2e19_app3.pdf]

References

1. Trinckes Jr J. The Definitive Guide to Complying with the HIPAA/HITECH Privacy and Security Rules. Boca Raton, FL: CRC Press; 2012.
2. Rousseau D. A&R Translations. 2017 Feb 24. Biomedical research: changing the common rule URL: https://www.ammon-rousseau.com/changing-the-rules-by-david-rousseau/ [accessed 2017-08-19] [WebCite Cache ID 6spHgiYRI]
3. Microsoft. 2016 May 12. Survey results from the Hosting and Cloud Study 2016 URL: https://www.microsoft.com/en-us/download/details.aspx?id=52045 [accessed 2018-03-04] [WebCite Cache ID 6xg4oxOQh]
4. Wu Y, Jiang X, Kim J, Ohno-Machado L. Grid Binary LOgistic REgression (GLORE): building shared models without sharing data. J Am Med Inform Assoc 2012;19(5):758-764 [FREE Full text] [doi: 10.1136/amiajnl-2012-000862] [Medline: 22511014]
5. Jiang W, Li P, Wang S, Wu Y, Xue M, Ohno-Machado L, et al. WebGLORE: a Web service for Grid LOgistic REgression. Bioinformatics 2013 Dec 15;29(24):3238-3240 [FREE Full text] [doi: 10.1093/bioinformatics/btt559] [Medline: 24072732]
6. Wang S, Jiang X, Wu Y, Cui L, Cheng S, Ohno-Machado L. EXpectation Propagation LOgistic REgRession (EXPLORER): distributed privacy-preserving online model learning. J Biomed Inform 2013 Jun;46(3):480-496 [FREE Full text] [doi: 10.1016/j.jbi.2013.03.008] [Medline: 23562651]
7. Skinner E, Octo Barnett G, Singer D, Mulley A, Lew R, Stickler S, et al. The use of logistic regression in diagnostic and prognostic prediction in a medical intensive care unit. In: Proc Annu Symp Comput Appl Med Care. 1980 Nov 05 Presented at: Proceedings of the Annual Symposium on Computer Application in Medical Care American Medical Informatics Association; ;1; Nov 2-5, 1980; Washington, DC p. 222-227 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2203702/

8. Dietz E. Application of Logistic Regression and Logistic Discrimination in Medical Decision Making. Biometrical J 1987;29(6):747-751 [FREE Full text] [doi: 10.1002/bimj.4710290614]
9. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med Decis Making 2001;21(1):45-56. [doi: 10.1177/0272989X0102100106] [Medline: 11206946]
10. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-1379. [Medline: 8970487]
11. Graepel T, Lauter K, Naehrig M. ML confidential: machine learning on encrypted data. In: Lecture Notes in Computer Science, vol 7839. Berlin, Heidelberg: Springer; 2012 Presented at: International Conference on Information Security and Cryptology; Nov 28-30, 2012; Seoul p. 1-21 URL: https://link.springer.com/chapter/10.1007%2F978-3-642-37682-5_1
12. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2017 May 06:1. [doi: 10.1093/bib/bbx044] [Medline: 28481991]
13. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35(5-6):352-359. [Medline: 12968784]
14. Bos JW, Lauter K, Naehrig M. Private predictive analysis on encrypted medical data. J Biomed Inform 2014 Aug;50:234-243 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.003] [Medline: 24835616]
15. Dowlin N, Gilad-Bachrach R, Laine K, Lauter K, Naehrig M, Wernsing J. CryptoNets: applying neural networks to encrypted data with high throughput and accuracy. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning. 2016 Presented at: 33rd International Conference on Machine Learning; Jun 19-24, 2016; New York p. 201-210.
16. Aono Y, Hayashi T, Trieu PL, Wang L. Scalable and secure logistic regression via homomorphic encryption. In: Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy. 2016 Presented at: Sixth ACM Conference on Data and Application Security and Privacy; Mar 9-11, 2016; New Orleans, LA p. 142-144 URL: https://dl.acm.org/citation.cfm?doid=2857705.2857731
17. Mohassel P, Zhang Y. SecureML: a system for scalable privacy-preserving machine learning. 2017 Presented at: 2017 IEEE Symposium on Security and Privacy; May 22-26, 2017; San Jose, CA p. 19-38 URL: http://ieeexplore.ieee.org/document/7958569/
18. Cheon J, Kim A, Kim M, Song Y. Homomorphic encryption for arithmetic of approximate numbers. 2017 Presented at: 23rd International Conference on the Theory and Applications of Cryptology and Information Security; Dec 3-7, 2017; Hong Kong p. 409-437 URL: https://link.springer.com/chapter/10.1007/978-3-319-70694-8_15
19. Ypma T. Historical development of the Newton–Raphson method. SIAM Rev 1995 Dec;37(4):531-551. [doi: 10.1137/1037125]
20. Bottou L. Large-scale machine learning with stochastic gradient descent. In: Proceedings of the 19th International Conference on Computational Statistics. 2010 Presented at: 19th International Conference on Computational Statistics; Aug 22-27, 2010; Paris p. 177-186 URL: https://link.springer.com/chapter/10.1007/978-3-7908-2604-3_16
21. Halevi S, Shoup V. Algorithms in HElib. In: Advances in Cryptology – CRYPTO 2014. 2014 Presented at: 34th Annual Cryptology Conference; Aug 17-21, 2014; Santa Barbara, CA p. 554-571.
22. Cheon J, Kim M, Kim M. Search-and-compute on encrypted data. In: International Conference on Financial Cryptography and Data Security. 2015 Presented at: FC 2015 International Workshops; Jan 30, 2015; San Juan, Puerto Rico p. 142-159.
23. Cheon J, Kim M, Kim M. Optimized Search-and-Compute Circuits and Their Application to Query Evaluation on Encrypted Data. IEEE T Inf Foren Sec 2016 Jan;11(1):188-199 [FREE Full text] [doi: 10.1109/TIFS.2015.2483486]
24. Cheon J, Kim A, Kim M, Song Y. GitHub. 2016. The HEAAN library URL: https://github.com/kimandrik/HEAAN [accessed 2017-08-19] [WebCite Cache ID 6spMzVJ6U]
25. Shoup V. NTL: A Library for doing Number Theory. 2001. URL: http://www.shoup.net/ntl/index.html [accessed 2018-03-04] [WebCite Cache ID 6xgAtJaXs]
26. Kim M, Song Y. GitHub. 2017. Implementation of secure logistic regression based on homomorphic encryption URL: https://github.com/K-miran/HELR [accessed 2017-08-19] [WebCite Cache ID 6spN8xvke]
27. Kennedy RL, Burton AM, Fraser HS, McStay LN, Harrison RF. Early diagnosis of acute myocardial infarction using clinical and electrocardiographic data at presentation: derivation and evaluation of logistic regression models. Eur Heart J 1996 Aug;17(8):1181-1191. [Medline: 8869859]
28. LogisticDx: Diagnostic Tests for Models with a Binomial Response. lbw: Low Birth Weight study data URL: https://rdrr.io/rforge/LogisticDx/man/lbw.html [accessed 2017-08-19] [WebCite Cache ID 6spNFX2b5]
29. LogisticDx: Diagnostic Tests for Models with a Binomial Response. nhanes3: NHANES III data URL: https://rdrr.io/rforge/LogisticDx/man/nhanes3.html [accessed 2017-08-19] [WebCite Cache ID 6spNJJFDx]
30. LogisticDx: Diagnostic Tests for Models with a Binomial Response. pcs: Prostate Cancer Study data URL: https://rdrr.io/rforge/LogisticDx/man/pcs.html [accessed 2017-08-19] [WebCite Cache ID 6spNLXr5a]
31. LogisticDx: Diagnostic Tests for Models with a Binomial Response. uis: UMARU IMPACT Study data URL: https://rdrr.io/rforge/LogisticDx/man/uis.html [accessed 2017-08-18] [WebCite Cache ID 6spNOLB9n]

Abbreviations

AUC: area under the curve
MSE: mean squared error
NMSE: normalized mean squared error

Edited by G Eysenbach; submitted 22.08.17; peer-reviewed by J Bian, C Bauer; comments to author 26.10.17; revised version received 21.12.17; accepted 10.01.18; published 17.04.18.

Please cite as:
Kim M, Song Y, Wang S, Xia Y, Jiang X
Secure Logistic Regression Based on Homomorphic Encryption: Design and Evaluation
JMIR Med Inform 2018;6(2):e19
URL: http://medinform.jmir.org/2018/2/e19/
doi:10.2196/medinform.8805
PMID:29666041

Original Paper

Developing a Third-Party Analytics Application Using Australia's National Personal Health Records System: Case Study

Niranjan Bidargaddi[1,2], BE, PhD; Yasmin van Kasteren[3], PhD; Peter Musiat[1], PhD; Michael Kidd[4], MD, PhD

[1]Digital Psychiatry & Personal Health Informatics Lab, College of Medicine and Public Health, Flinders University, Adelaide, Australia

[2]South Australian Health and Medical Research Institute, Adelaid, Australia

[3]Flinders Digital Health Centre, College of Nursing and Health Sciences, Flinders University, Adelaide, Australia

[4]Department of Family & Community Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Niranjan Bidargaddi, BE, PhD
Digital Psychiatry & Personal Health Informatics Lab
College of Medicine and Public Health
Flinders University
GPO Box 2100
Adelaide, 5001, SA
Australia
Phone: 61 423416543
Email: niranjan.bidargaddi@flinders.edu.au

Abstract

Background: My Health Record (MyHR) is Australia's national electronic health record (EHR) system. Poor usability and functionality have resulted in low utility, affecting enrollment and participation rates by both patients and clinicians alike. Similar to apps on mobile phone app stores, innovative third-party applications of MyHR platform data can enhance the usefulness of the platform, but there is a paucity of research into the processes involved in developing third-party applications that integrate and use data from EHR systems.

Objective: The research describes the challenges involved in pioneering the development of a patient and clinician Web-based software application for MyHR and insights resulting from this experience.

Methods: This research uses a case study approach, investigating the development and implementation of *Actionable Intime Insights (AI*^{2}*)*, a third-party application for MyHR, which translates Medicare claims records stored in MyHR into a clinically meaningful timeline visualization of health data for both patients and clinicians. This case study identifies the challenges encountered by the Personal Health Informatics team from Flinders University in the MyHR third-party application development environment.

Results: The study presents a nuanced understanding of different data types and quality of data in MyHR and the complexities associated with developing secondary-use applications. Regulatory requirements associated with utilization of MyHR data, restrictions on visualizations of data, and processes of testing third-party applications were encountered during the development of the application.

Conclusions: This study identified several processes, technical and regulatory barriers which, if addressed, can make MyHR a thriving ecosystem of health applications. It clearly identifies opportunities and considerations for the Australian Digital Health Agency and other national bodies wishing to encourage the development of new and innovative use cases for national EHRs.

(JMIR Med Inform 2018;6(2):e28) doi:10.2196/medinform.7710

KEYWORDS

computer software applications; electronic health record; software design; medication compliance

Introduction

Background

Across the member countries of the Organisation for Economic Co-operation and Development, national electronic health records (EHRs) are at varying stages of implementation [1]. The overarching objective of national EHRs is to improve the delivery and administration of health care [1]. Australia is one of the countries that have implemented a national rollout of EHRs. The vision for My Health Record (MyHR), as with other national EHR systems, is to improve health care outcomes through coordinated care and the sharing of essential patient information at the point of care [2]. In the United Kingdom, which, like Australia, has a universal health care system, the National Health Service took a top-down approach in developing EHRs. In the United Kingdom, 97% (55 million) of health care consumers (hereafter referred to as patients) have online access to their Summary Care Record, which contains information on prescriptions, allergies, and adverse reactions and allows patients to access their record, book online appointments, and order repeat prescriptions. Only 0.4% of patients in the United Kingdom have ever used this service [3].

As in the United Kingdom, Australia's MyHR is suffering from lack of buy-in from patients. Since its launch in 2012, uptake of MyHR has been insufficient to achieve the critical mass of patient participation necessary to galvanize health care provider engagement [4]. Unlike the United Kingdom, Australia adopted an opt-in system, and as of November 2017, only 22% of Australians were registered for MyHR. To mobilize health provider engagement, from 2018, MyHR will shift from an opt-in model to an opt-out model. Equally, health care providers have also been slow on the uptake as they have little incentive to use MyHR as a communication platform, because a vast majority already use practice management software with advanced functionalities including most of the information intended to be made available through MyHR [5]. For a national EHR system to succeed, both health care consumers and providers must buy into the system. However, it is not the uptake, but rather the usefulness of MyHR, which will ultimately drive the necessary widespread adoption and everyday use of MyHR for providers and patients alike [6]. Only then, the goals of improving health care outcomes with a national EHR system can be achieved.

Unlike the United Kingdom and Australia, the United States adopted a bottom-up approach to the development of a national EHR system. The strategy in the United States was a staged approach focusing on "meaningful use" [7]. The first stage saw the development of the necessary infrastructure for capturing and storing EHRs, and the second stage, currently in progress, incentivizes hospitals and health care providers to find ways to engage with electronic records in meaningful ways. The final stage will be the realization of the benefits of a national EHR system [7,8]. Research suggests that maximizing added value for patients is most likely to drive the widespread adoption of EHRs and realize the benefits and the outcomes of national EHR systems [6] because patient portals are important tools for engaging patients in their own health care [9-11].

In Australia, the initial focus of third-party software development for MyHR was on software integrating MyHR into existing clinical information systems. By clinical information systems, we mean computer-based systems for managing and storing in-patient or practice-based medical records and test results for the delivery of patient care in local settings such as a general practice (GP) or a hospital [12]. Since the inception of Australia's national EHR, most conformant third-party applications have been clinical information systems or contracted service provider hosts [13]. In 2012-2013, 26 software vendors linked to MyHR through their clinical information systems, which included applications for GPs, hospitals, aged care providers, and pharmacies [14]. By 2016, a further 19 clinical information systems were registered plus 3 systems for contracted service providers to host or provide infrastructure for use by health care providers. On the other hand, development of software applications that are patient-centered has been limited.

In Australia, there are 2 factors that influence the usefulness of MyHR—the type of data sources available and the functionality of the system. First, MyHR is currently missing a coherent approach to automatically source data from health care providers and other repositories of health data, resulting in incomplete data. The most complete and up-to-date data currently available on MyHR are Medicare data. Medicare is Australia's universal care provider. The Medicare claims database records medical intervention claims reimbursed through the Medical Benefits Scheme (MBS) and pharmaceutical prescriptions and dispensing claims reimbursed through the Prescription Benefit Scheme (PBS). Second, although access to fragmented medical history in a single place is one of the main perceived benefits and utility of EHRs [15], MyHR cannot currently create a collated health summary [5], which limits the usability of MyHR for the coordination of care.

In this research, we review the development of a third-party application called *Actionable Intime Insights (AI^2)*, the first analytic software developed for MyHR by Digital Psychiatry & Personal Health Informatics Lab, Flinders University, Adelaide, South Australia. Designed as an application for both patients and clinicians, AI^2 capitalizes on the demonstrated utility of Medicare claims data in understanding treatment [16,17] by collating and organizing medical visits and prescriptions and dispensing information from the Medicare database to allow patients and clinicians to view patient history in an intuitive timeline format. The research describes the challenges involved in pioneering the development of a patient and clinician Web-based software application for MyHR and insights from this experience. Recommendations for improvements to MyHR development environment are discussed.

Data Stored in Health Record

EHRs involve the exchange of information with other health care systems and repositories of health care data [18]. Developing a national EHR system is a complex challenge both in terms of design and implementation because of the involvement of multiple stakeholders and concerns over interoperability, privacy, and security [19]. MyHR collates data

from many sources; however, different sources of data have different guidelines, procedures, and permissions that regulate their inclusion in MyHR. Figure 1 summarizes the sources of data that feed into MyHR and how they are generated as records in MyHR through the MyHR portal.

Health data stored in MyHR are programmatically accessible in a format known as views. There are 11 different views for different types of data as shown in Figure 1. Except for the Medicare view, all views are populated by data sourced from various clinical information systems used by different health care providers such as GPs, hospitals, or radiology and pathology clinics. The health data in each of these views are populated under different conditions with varying level of data completeness. First, data integration requires patients having a MyHR. Second, certain records, eg, GP records, require action by the GP and patients' prior permission for them to be uploaded onto MyHR [20]. Third, health care providers must have the clinical information systems in place that can access and upload data to MyHR [21]. Finally, despite the availability of interfaces, not all available data are automatically extracted from these sources and populated into corresponding views. Data uploads depend on if health care systems and repositories are linked to MyHR and if they are set up to push data into MyHR automatically or on request. As a result of these factors, it has been claimed that health data in MyHR are often incomplete, thus making it unsuitable for patient care [22]. Incomplete data are also problematic for analysis.

A notable exception to the above is data stored under the Medicare view, which contains details directly populated from the Medicare claims database. It consists of a list of reimbursed MBS and PBS interventions. The Medicare records are complete and, on creation of a MyHR, the system automatically uploads retrospective data for 2 years from the date of activation of a MyHR account and then subsequently automatically updates records after each and every intervention.

Medicare Data in My Health Record

Medicare data record reimbursement claims for dispensed prescriptions and administered interventions such as diagnostic and therapeutic procedures, oral and maxillofacial, diagnostic imaging, and pathology services, as well as data from allied health and optometry services.

Figure 1. Overview of data input into My Health Record (myHR).

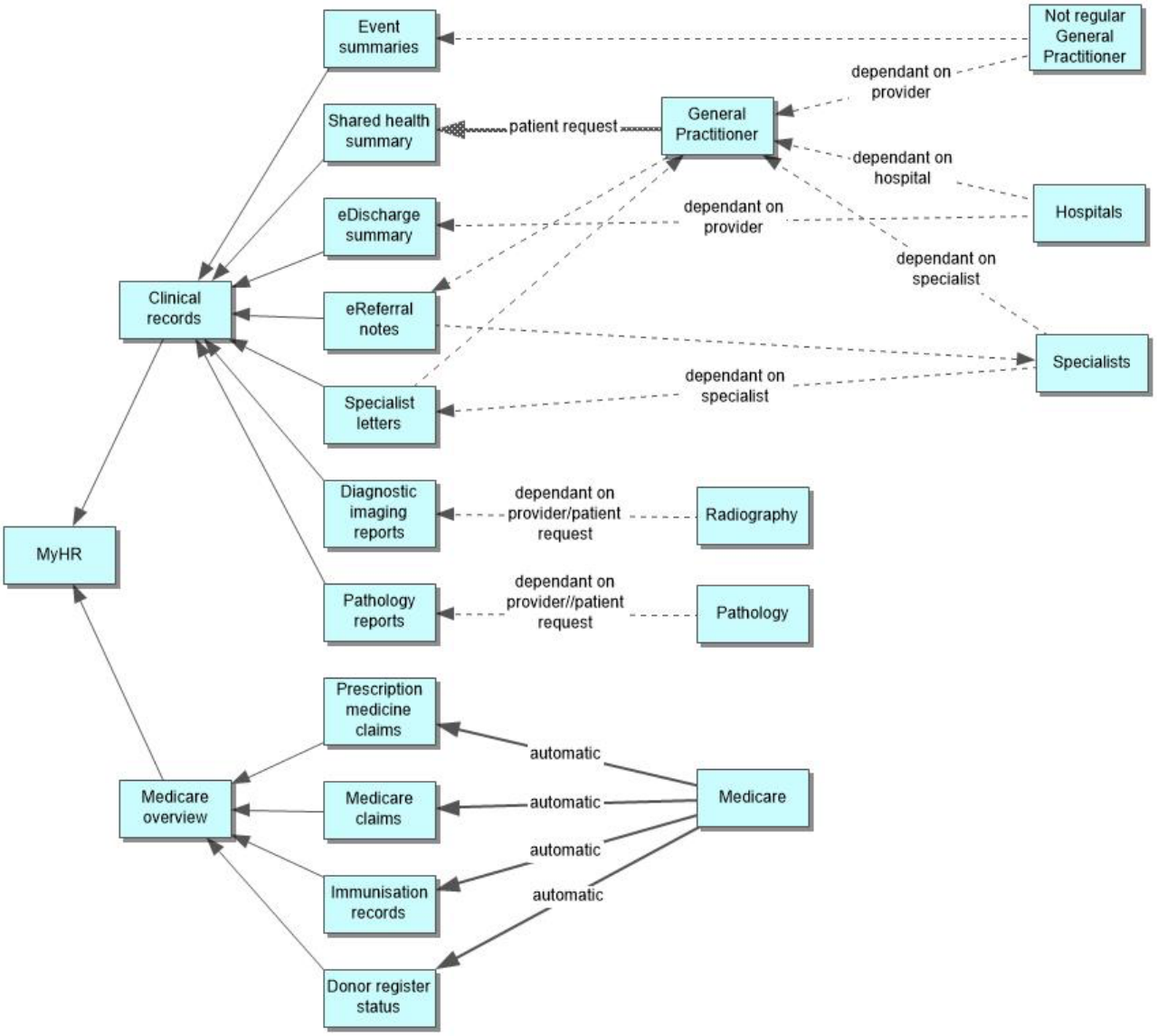


Table 1. Data fields for Medicare data.

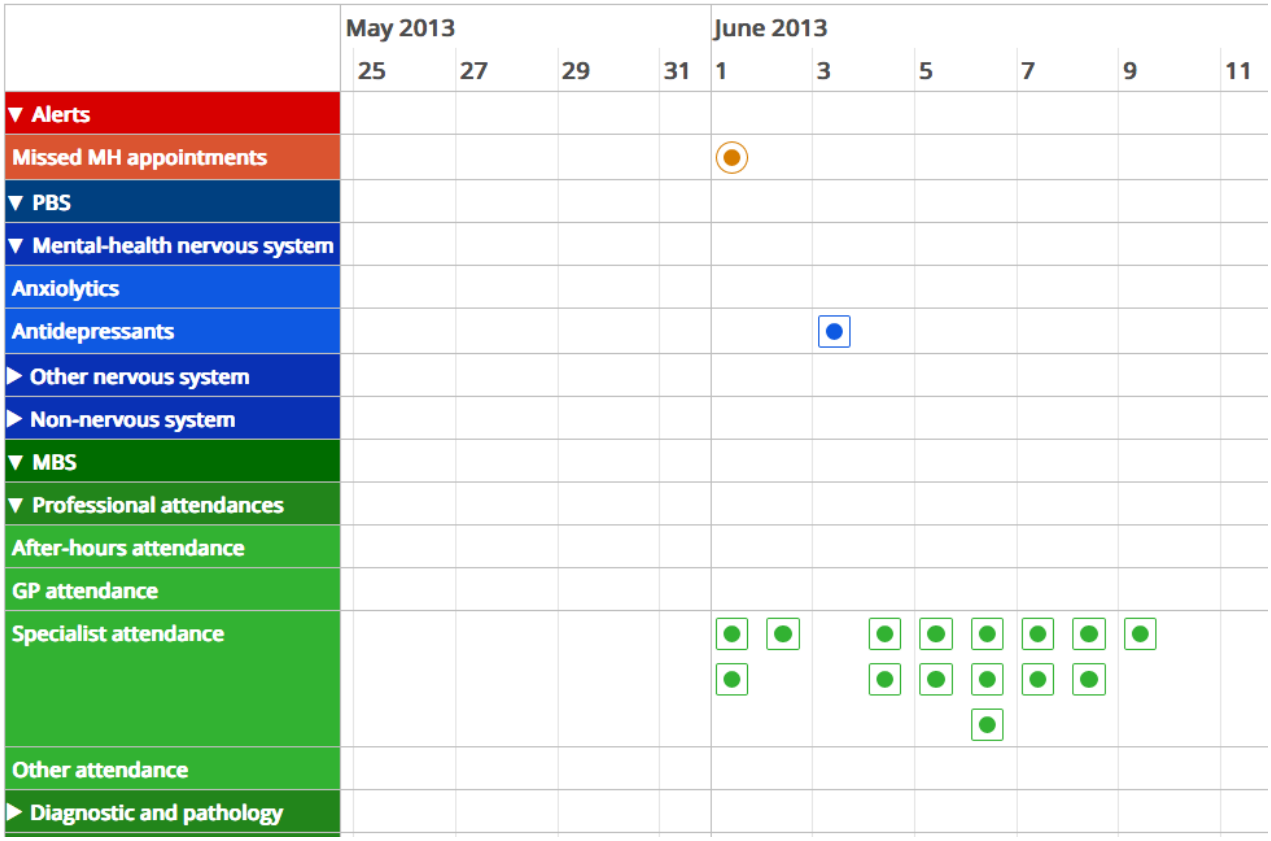
Medical Benefits Scheme data	Prescription Benefit Scheme data
Date of service	Generic name
Item number[a]	Brand
Item description[b]	Date prescribed
Name of practitioner	Date supplied
In hospital (yes/no)	Form and strength
	Quantity
	Number of repeats[c]
	Code

[a]Medicare Benefits Schedule outlining item numbers [23].

[b]For example, "consultation and consulting rooms, Level B" or "Initial Specialist Attendance, MRI scan for derangement of shoulder or its supporting structures."

[c]Number of repeat prescriptions left.

Figure 2. AI Squared Medical Benefits Scheme (MBS) and Prescription Benefit Scheme (PBS) data timeline visualization. GP: general practice.



The data detail the date of visit to a supplier of Medicare-funded services and the specific services or tests performed as described in the medical benefits schedule [23]. It also contains information about the type of medication, amount of medication supplied, and the date of supply. Table 1 shows the data fields as displayed in MyHR from Medicare's PBS and the MBS. This information can currently be viewed by both patients and health care providers in the form of 2 separate tables.

Benefits of AI Squared for Patients and Clinicians

AI^2 is designed as a responsive Web application that visualizes MyHR Medicare view data in the form of a timeline view. It has patient and clinician interfaces but is designed to be a standalone application accessible by registered patients even when there is no interaction with health professionals involved. Thus, it is the first patient analytics-oriented third-party application interfacing with MyHR.

Medicare view data are a by-product of administrative claims processing. So, to provide an initial level of useful analytics,

AI^2 uses a simple taxonomy to map administratively used item types and prescription codes into clinically relevant categories. The clinical categories are based on the intervention provider type or medication names for each disorder. The mapping involved creating a parser that groups item numbers associated with similar service providers (eg, GPs, psychiatrists) and grouping of prescriptions associated with same disorders such as antidepressants. The result of grouping data and using the timeline format is that it is easier to quickly visualize a patients' trajectory and gaps over time; see Figure 2.

The benefit for patients is that AI^2 provides an intuitive timeline interface that can facilitate the management of their health, by collating pharmaceutical and health provider encounters. AI^2 is most likely to benefit patients with severe mental illnesses or chronic disease, with ongoing medications and treatment through multiple care providers, by easing the burden of managing or recalling extensive history of polypharmacy and medical information across multiple health care providers [9,18,24].

For clinicians, understanding the patient history is an essential part of the patient interview [25]. For patients with complex conditions and in a time-constrained environment, a visual overview of the patient's history assembled from the MBS and PBS data can improve accuracy, save time, and serve as a point of discussion between clinician and patient. The benefit of the timeline visualization of medication and medical services is that it can quickly reveal gaps in medications, and typical service uses patterns and changes, which provide clinical contexts relevant to assessment and treatments [26,27]. An overview of compliance with medication and treatment can also potentially reduce problems with polypharmacy and conflicts over care plans across multiple care providers [18,28,29].

Methods

This research is a case study of a pioneering development of a third-party application for MyHR. Although the AI^2 application has not been officially released as it is currently undergoing trials, it was nonetheless the first third-party application developed for MyHR. As such, it was a test case for the Australian Digital Health Agency and other developers of MyHR. This case study identifies the challenges encountered by the Personal Health Informatics team from Flinders University, in the development of a third-party application for MyHR. The research identifies opportunities and considerations for the Australian Digital Health Agency and other national bodies wishing to encourage the development of new use cases for national EHRs. Having a functional and efficient third-party application development environment is important because application can "stimulate open innovation and competition for products that deliver on consumer and health care provider expectations in the digital health space, and ultimately contribute to improved health care outcomes for people" [30].

Results

AI Squared Application

The development of the AI^2 application involved implementing interface programs to access Medicare data, applying timeline visualizations on obtained Medicare data, and testing conformance requirements. The application was developed using open-source development and hosting tools. It is developed in Java Enterprise Edition using JBoss Seam and Hibernate frameworks. The database is created in PostgreSQL. JBoss Application servers and APACHE Web servers are required to run the compiled application. Implementation of a third-party application that interfaces with MyHR involved integration with 2 Web services known as Health care Identifiers verification (HI) Service, maintained by Medicare Australia, and MyHR Service from the Australian Digital Health Agency. HL7 SOAP protocol is used for integration with HI and MyHR Web services. The resulting application is developed to be registered as a stand-alone "health service" that can be accessed directly by the patients with or without a treating clinician and satisfy the requirements of 2012 Australian MyHR legislation. The source code of the application is available for the development of future patient-oriented applications and the corresponding author can be contacted to request access.

Creating User Records in AI Squared Registration and Login Process

AI^2 was designed with 2 types of users in mind: patients and clinicians. The application can be used independently either by patients, for self-monitoring, or by clinicians for checking on care plans or compliance or as a tool for both patient and clinician to verify patient history and discuss events, care plans, or treatments. Either the patient or the clinician can create an AI^2 record.

Using test data, we created a patient record on MyHR. Creating a patient record on AI^2 involves setting up a username, which is a mobile phone number and password as well as providing details required for the purposes of verifying Individual Health Identifier (IHI) with the HI service. To do this, patients can either provide their 16-digit IHI number, or alternatively provide other details (first name, last name, date of birth, gender, and Medicare number). The application sends this information to HI service for verification, and on receiving a valid user confirmation, a user record is created in the AI^2 database.

The registration page for clinicians is similar, as it also involves entering a mobile number and password similar to the patient users. The registration page also has the capacity to collect the 16-digit Healthcare Provider Identifier (HPI-I), another unique identifier used by MyHR for verifying health professionals. All clinicians automatically have a Healthcare Provider Number, which is their Australian Health Practitioner Regulation Agency ID number with the addition of "800361" digits before [31]. The application sends this information to HI service's "Healthcare provider directory search" application programming interfaces (API) for verification purposes.

Process of Accessing Data From My Health Record

The process for fetching data from MyHR within AI^2 involves, first, verifying that the individual has an activated MyHR, and on successful confirmation, gaining access to MyHR database. After receiving successful confirmation, the application then fetches health data using "Getview" APIs. The application is programmed to extract the latest records from MyHR whenever a patient profile is opened by either a patient or a clinician. Patients can also view a list of AI^2-registered clinicians and can control which of these clinicians has access to their data using a connect or disconnect button.

Registered clinicians have to be connected to a patient user before they can view the patient's record. The application can be configured to allow clinicians either to be automatically connected to all or selectively connected to registered patient users. If a clinician is not designated to have automatic connection with all individuals, they can search for patients by name, date of birth, gender, and Medicare number to find matching patients and connect, as is current practice in hospitals.

Applying Timeline Visualizations on My Health Record Data

The AI^2 application applies a visually rich interactive timeline visualization on Medicare view data, implemented using vis.js and timeline.js, 2 open-source, browser-based visualization libraries. The timeline visualization displays each MBS and PBS claim stored in Medicare view as an event against a timescale on horizontal axis (see Figures 2 and 3). The events are shown at the corresponding start date position displayed on the horizontal axis, and the width of the event box can be adjusted to the corresponding end date position if it exists. Each event is associated with a group, and all events with the same group are shown in the same row or adjacent rows as a box, and the common value of their group property is used as a label at the left edge of the timeline. A simple taxonomy was developed to create a hierarchical grouping of PBS prescriptions and MBS interventions such that they are clinically meaningful and that variables required for assigning them into the groups were available in the Medicare view data. PBS prescriptions were grouped into generic medication classes, and MBS interventions were grouped into 4 categories—namely, GPs, specialists, imaging, and pathology. Within each MBS group, there were 2 subgroups based on if the service was provided in hospital or out of hospital, and within each subgroup, further levels were based on the nature of service. The group labels are displayed at the left side of the timeline visualization. Each MBS intervention and PBS prescription contained in the Medicare view (also see Figure 2) is then displayed as an event on timeline in the row of the group it belongs to. The timescale on the horizontal axis can be adjusted with zooming options, which dynamically scale the visualized data. For each event box, additional contextual information is displayed either in the box or on the top of a banner. In the case of MBS interventions, contextual information includes the description of the claim such as "Optometrist visit" and the details of the service provider as well as if the service was provided "in hospital" or "out of hospital." For PBS prescriptions, the contextual information includes prescription name, dispensed date, and number of medications supplied. In addition to the above visualization, a simple notes entry page was also included. This was designed to encourage patients and health professionals to record insights or alternatively enable patients to make notes or observations they may like to follow up on.

Process of Developing Software for My Health Record

MyHR consists of a centralized infrastructure to verify, obtain, and transfer health information of individuals from repositories holding health and clinical documents. It is designed to interact with several distributed repositories, some of which are established and managed centrally by the Australian Digital Health Agency, and others are maintained by registered external organizations, including Medicare. The central infrastructure is designed to query and identify the repositories that contain information about an individual patient using their unique identifier, known as Individual Healthcare Identifier (IHI) and to collate available data on request from the repositories.

Figure 3. Timeline visualization of Medical Benefits Scheme (MBS) data over a number of years. PBS: Prescription Benefit Scheme.

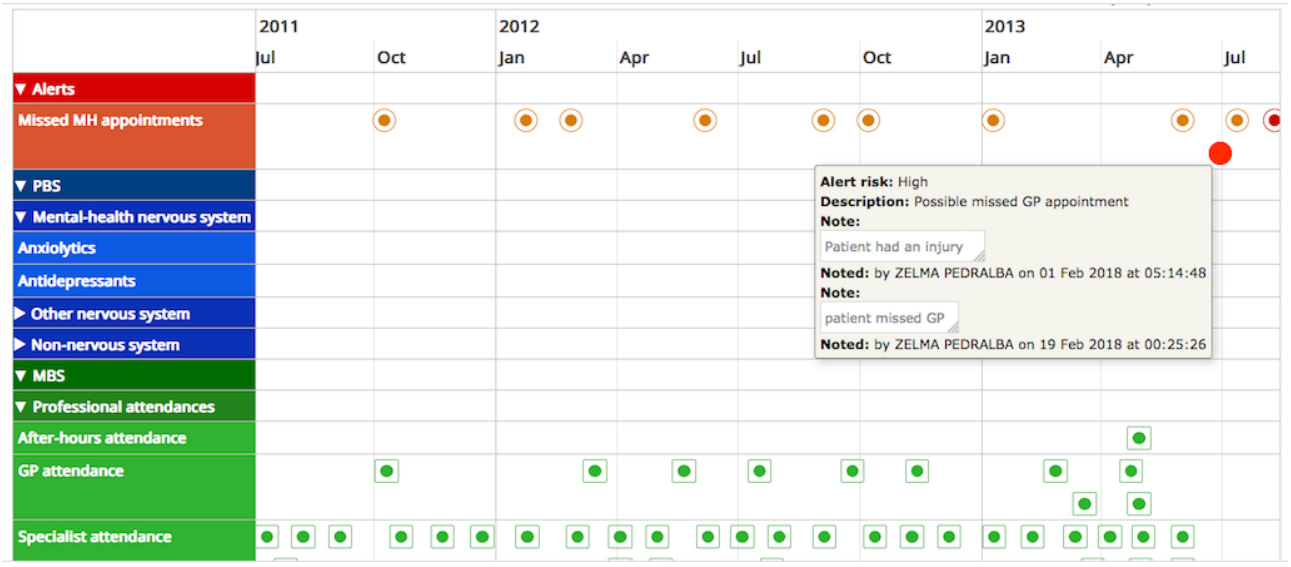


Figure 4. Links between systems in the third-party application development environment. Continuous lines represent automatic push of information or data. Dashed lines represent push data on request of information or data. GP: general practice; MyHR: My Health Record.

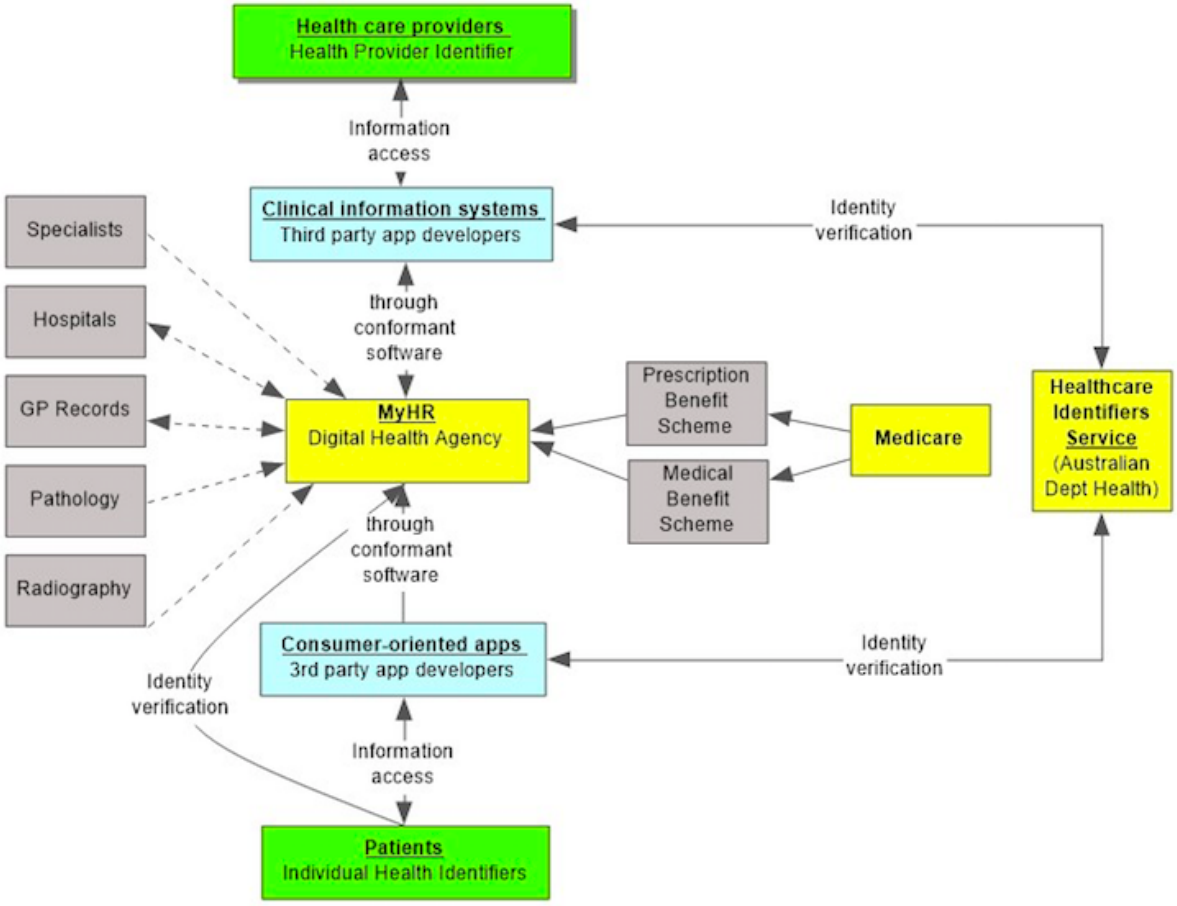


Figure 4 illustrates the way in which repositories and providers connect via MyHR.

Regulatory Requirements for Using My Health Record

Registering as an Application Developer

Developing and administering the AI[2] application involves hosting the infrastructure on a server and meeting eligibility criteria to be a host organization by the Australian Digital Health Agency. The host organization, in this case, Flinders University, must adopt a MyHR use policy, register to be a MyHR "participating organization," and apply to obtain a Health Provider Identifier. The purpose of MyHR's use policy was to ensure that organizations are accessing data through conformant software for providing health care only.

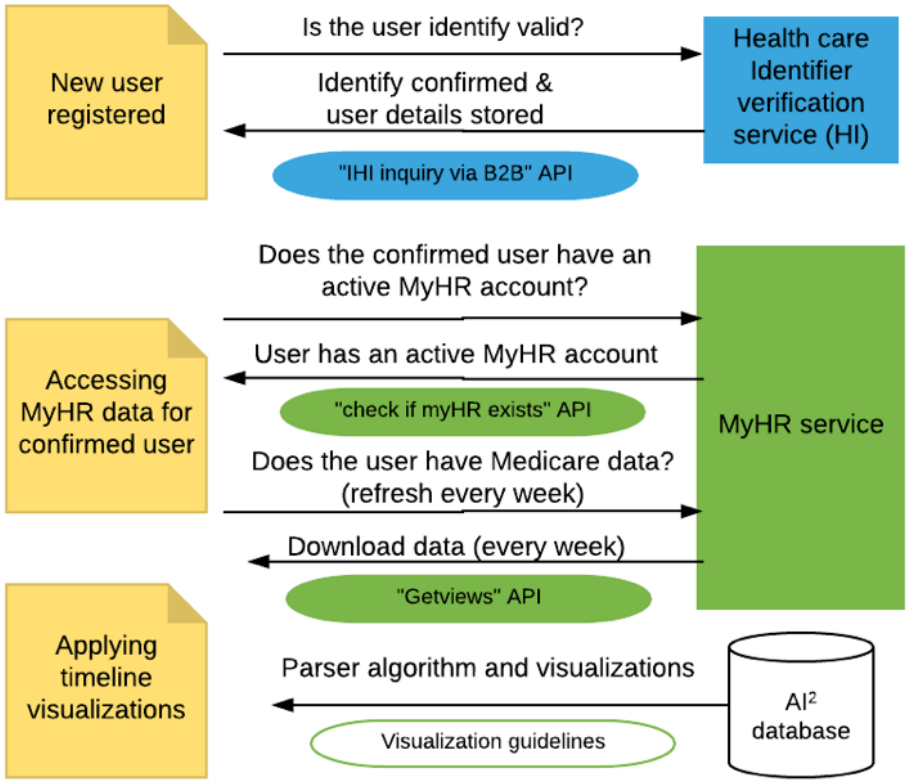
To develop a MyHR application for patients and or clinicians, the first step was to register as an application developer with Medicare and the Australian Digital Health Agency , after which access is granted to a development environment and a test kit containing sample data for testing and descriptions of supported integration use cases.

The Personal Health Informatics team from Flinders University was registered as a software developer with Medicare Australia and the Australian Digital Health Agency to gain access to developer and testing environments. At the time of the development of AI[2], access was restricted to applications developed for the purposes of providing "health service" as defined in the Privacy Act 1988 (Cwlth) as follows:

> *(a) an activity performed in relation to an individual that is intended or claimed (expressly or otherwise) by the individual or the person performing it:*
>
> *(i) to assess, record, maintain or improve the individual's health; or*
>
> *(ii) to diagnose the individual's illness or disability; or*
>
> *(iii) to treat the individual's illness or disability or suspected illness or disability; or*
>
> *(b) the dispensing on prescription of a drug or medicinal preparation by a pharmacist.*

The interpretation of this act presupposes the involvement of a clinician. There were no special provisions for the development of patient portals, which therefore were also required to meet the definition of providing a "health service." Thus, a case was argued with the Department of Health that the analysis provided by the AI[2] met the criteria of providing a "health service" by providing patients with health data insights. Flinders University was given as a unique 16-digit identifier as a "health services provider" under the auspices of Healthcare Identifiers Act 2010 (Cwlth) (21).

Figure 5. Software integration with Healthcare Identifiers verification service (HI) and My Health Record (MyHR) service. AI^2: AI Squared; B2B API: Business 2 Business application programming interfaces.



Software Implementation

Figure 5 shows the different software components a third part party application must implement. AI^2 had to integrate the Business 2 Business API's functionality from the HI service into the application to verify the identity of their application users via the HI service to exchange data with MyHR by integrating. Third-party applications that can successfully verify users through the HI service are then eligible to read and write MyHR data using API functionality offered by MyHR.

Technical Conformance Testing and Gaining Access to the My Health Record Production Server

As would be expected on a national EHR, there are strict technical and intended use requirements. AI^2 and other third-party software that integrates with MyHR had to meet conformance requirements of both the HI service and the My Health Record to demonstrate that the application could first exchange information with the HI service, and second, pass conformance tests to ensure the information exchange was consistent with approved use cases and rendering guidelines for displaying information from MyHR. Each test, 4 tests in all, took between 3 and 6 months.

The testing process involved first creating test cases, and subsequently, remotely assessing these test cases using the sample data provided in the test kit. Sample datasets in the test kit were mainly aimed at testing the authentication and data access process and not utilization needs. They neither contained longitudinal records nor did they have a sufficient number of test cases required for refining the visualization categories. Thus, it was necessary for us to use data sourced from a different study to develop the timeline analytics [32].

The first test, with the HI Service, the Notice of Connection (NOC) test, involved taking a screenshot and log file based on test cases, which were then approved by testers at the Department of Health.

After completing the NOC test, the second test, the HI conformance test, was conducted by IV&V Australia, an accredited tester at a cost of Aus $10,000. On passing the HI service conformance tests, the Department of Health issued an approval letter with details to gain access to the production HI Service.

The third test, the MyHR NOC test, was carried out by Ventura Inc. at no cost. Testing involved verifying that appropriate warnings and alerts were displayed in the application for incorrect individual details included in sample data. A tester from Accenture PLC remotely monitored the application, while the developer tested different patient and clinician scenarios, also at no cost. The fourth and final test, the MyHR conformance test of rendering guidelines, was done via a self-assessment form.

Displaying timeline visualizations of Medicare data obtained from MyHR required overcoming several standards-related challenges. To meet the software conformance standards, health data had to be displayed in a predefined format and style guidelines set by the Australian Digital Health Agency [33]. The guideline specifies 67 different requirements covering font size, format, and structure, and not all of which were applicable to the formats and style of timeline visualizations. As a

workaround, we created 2 different data visualizations on the same Medicare data, the first using rendering in the approved style as prescribed by the Australian Digital Health Agency and the second visualization using the desired timeline visualizations described above. Subject to implementation of a conformant approach, no restrictions applied to alternative secondary visualizations.

On completion of all 4 conformance tests, the Department of Health issued a letter of approval, granting access to the MyHR production server through the AI^2 application.

Requirements for Organizations to Use the AI Squared Application

To use the AI^2 application, an organization must register with the Australian Digital Health Agency and obtain an HPI-O number, which is a unique 16-digit organization number verifiable by the HI service as well as a National Authentication Service for Health Public Key Infrastructure Certificate, a digital certificate for activating the conformant software AI^2 application. Both these details need to be keyed into AI^2 before the platform can make connections with live MyHR. Both clinicians and patients have to read and consent to the terms of use for the AI^2 service and the privacy policy of the AI^2 application, which were developed in consultation with the Flinders University legal team. The privacy policy contains information on user obligations and outlines how information sourced from MyHR are used and managed. After this process, the South Australian Health and Medical Research Institute is registered as the "participating organization," and an instance of the AI^2 application for supporting psychiatric patients is hosted on Nectar, a cloud server infrastructure available for Australian university researchers under the AI^2 website [34].

Discussion

Overview

Developing a third-party patient application for EHRs has been previously attempted and deemed difficult [35]. This paper documents the process and challenges encountered in the development of an innovative third-party patient and clinician application, the AI^2 application, which displays Medicare data from MyHR in an intuitive timeline format. Although there were numerous third-party clinical information applications being developed for MyHR, AI^2 was the first application also designed for patients. Results identified several challenges encountered in the process of developing the application that can be grouped in 4 categories: (1) regulations related to use of MyHR in applications, (2) type of applications of MyHR data, (3) issues related to data processing, and (4) regulations related to testing.

Regulations Related to Use of My Health Record Data in Applications

First, at the time of the development of AI^2, the focus of development of third-party applications was predominantly on development of clinical health information systems. A major challenge in the development of AI^2 was to understand how a patient-centered third-party application might meet these requirements.

All third-party conformant software has to interface with MyHR via the provider portal. To do so, at the time, providers had to be recognized as a "health service provider," defined by the MyHR legislation [36]. The legislation stipulates that eligible entities can only utilize MyHR data for the purpose of providing "health services" as defined in the Privacy Act 1988 (Cwlth), regulation designed with development of clinical information systems in mind. Several revisions were made to act in 2006 after consultation and review, but the definition of use has not changed. This definition is problematic for the development of patient-focused third-party applications, in particular, for the new category of applications widely known as online or Internet personal health interventions, standalone applications of personal health interventions which have been shown to be effective in health, eg, in treatment of depression [37] or cardiovascular rehabilitation [38].

Circumventing the definition of "health service provider" has 2 possible interpretations, one is to consider the application itself as a "health service provider", and the other is to consider the organizations creating and hosting standalone patient-oriented applications of MyHR as "health service" providers. Either way, defining the software or the software development companies as "health service" may raise issue of regulation.

Treating the software itself as a "health service provider" feeds into the growing debate on the need for standards and regulation-related health service provision for standalone applications of personal health interventions [39,40]. In Australia, software offering personal health interventions might then arguably fall under the auspices of the Therapeutic Goods Administration, which regulates goods used in "preventing, diagnosing, curing or alleviating a disease, ailment, defect or injury in persons" or for "influencing, inhibiting or modifying a physiological process in persons," [41] and similar legislation would apply in other countries.

Treating software development companies as health service providers may result in more opportunities for creating new applications using MyHR data; however, health services providers, in particular, professional health services providers, are strictly regulated [42]. Defining software development companies as "health service providers" may potentially require registration and or accreditation, and it might also change the way in which complaints are handled [42].

Another challenge raised by framing MyHR third-party software development as "health service" provision is a restriction on use and, in particular, restrictions on use for research and analysis which do not meet the criteria of the My Health Record 2012 (Cwlth) legislation. Paradoxically, the restriction on use only applies to retrieving data from MyHR. Once the information is lawfully obtained from MyHR, local terms of use and privacy policy within the application can be applied for subsequent utilization of downloaded data to support new use cases such as in the case of AI^2, future plans for providing

treatment decision support, data linkage endeavors, or recruitment for clinical trials [43]. Although recently, the Australian Digital Health Agency has begun investigating frameworks for the secondary use of data in the MyHR system for research, policy planning, system use, quality improvement, and evaluation activities.

My Health Record Data Application Use Case Considerations—Completeness of Records

The quality of and completeness of records/data available via MyHR varies substantially and is a major limitation for the development of new use cases with MyHR. The completeness and therefore the value of information under each of the views depends on, for some records such as GP-shared summaries, whether the patient permission to upload has been granted, whether the service provider has the technology to upload to MyHR, and whether the data upload happens automatically. Currently, only Medicare data are complete and up-to-date and, therefore, potentially useful for innovative analytics and clinical trial use cases.

Data Processing Challenges for Analytics

Developing and using MyHR data for analytics applications for MyHR also presents 2 data processing challenges. The first relates to standards and interoperability. The data held by MyHR from the clinical repositories are analogous to shared file repositories, such as Dropbox. With the exception of Medicare data, data in MyHR from sources such as hospital and GPs are a collection of clinical documents with free text information. These data are characterized by clinical terminologies and jargons specific to the repositories in which they were created. To use this kind of data for analytics requires understanding and developing approaches to translate these terminologies using natural language processing analysis. Equally devising taxonomies, particularly suitable for analytics, can help reduce data complexity and potentially accelerate development of decision support applications using the data from MyHR. Efforts to reduce complexity in data using taxonomies have been attempted in apps and wearable devices [44] and online marketing [45].

The second challenge relates to visualization and rendering of data sourced from MyHR. Results showed that a time-consuming work around was needed to meet the very prescriptive rendering guidelines needed to obtain conformance, even though the purpose of the application was to develop an alternative visualization of data. AI^2 overcame this challenge by implementing both the default rendering and the new timeline visualization; however, developing the work around was time-consuming and ultimately did not contribute to the usefulness of the application. Relaxing the rendering guidelines would allow software developers to create new visualizations of data.

Regulations Related to Development and Testing

The ease of third-party application integration with MyHR system is another important factor in the development of secondary applications. Development with MyHR involves integration with 2 separate but interrelated services, which are coordinated by 2 different organizations. Medicare is responsible for the integration of user identity verification, and the Australian Digital Health Agency is responsible for the integration of provider identity verification. This dual verification process creates a complex workflow for application developers, particularly for startups and small businesses with limited resources. In addition, new applications have to be tested separately for an active connection and conformance to standards for each of these services. This complex process not only involves testing several elaborate scenarios, which may or may not be within the scope of the developed application, but also involves conducting these tests with 4 different teams. In this project, the time and resource allocated for meeting conformance requirements were substantially higher than was justified by the actual use cases. Although significant support is offered by the team at the Australian Digital Health Agency to help developers navigate the process, streamlining the testing process could actually reduce the time and costs for both application developers and the agency.

Streamlining integration and testing processes within MyHR will reduce the costs of third-party application development and reduce the time to market. It is more likely to create the impetus for an ecosystem of applications with innovative use cases of health data to emerge as it has for in the consumer wearables and health application spaces. Improvements to the way in which software developers can integrate and work with MyHR will ultimately both improve engagement with and generate value for MyHR. Finally, improving the size and richness of test data provided to application developers such that it reflects complete records of several patients can also assist in helping development of new applications.

Conclusions

Given the substantial investment by the Australian Government, and other governments worldwide, in developing National EHR system, MyHR system operators and other national EHR authorities are increasingly recognizing the value of innovative third-party applications in adding value for consumers by better engaging consumers in their own health care and for clinicians, through the addition of analytics to clinical portals. Innovative applications have a key role in realizing the initial vision for MyHR of improving health care outcomes and gaining efficiencies in the delivery of health care services. Consequently, it is important that the development environment facilitates rather than hinders third-party application development to attract and capitalize on research and entrepreneurial activities in this space.

Acknowledgments

NB is supported by funding from Country Health SA. The study was supported by grants from Country Health SA and Flinders ED IAPT service. The authors would like to acknowledge technical support and advise provided by staff at the Australian Digital

Health Agency (formerly NEHTA) and the contributions of Mr Seoyong Lim and Yang Yang on this project for programming and testing the software application. The authors wish to acknowledge the Medical Research Futures Fund Rapid Applied Research Translation Program and Commonwealth Department of Health in supporting the aim to trial AI^2 application with mental health patients.

Authors' Contributions

NB designed and led the study, supervised software development, and wrote the full draft. YvK, PM, and MK contributed to the writing and revision of the paper.

Conflicts of Interest

None declared.

References

1. Zelmer J, Ronchi E, Hyppönen H, Lupiáñez-Villanueva F, Codagnone C, Nøhr C, et al. International health IT benchmarking: learning from cross-country comparisons. J Am Med Inform Assoc 2017 Mar 01;24(2):371-379. [doi: 10.1093/jamia/ocw111] [Medline: 27554825]
2. Ley S. 2015. Patients to get new myHealth Record: $485m 'rescue' package to reboot Labor's e-health failures URL: http://www.health.gov.au/internet/ministers/publishing.nsf/Content/health-mediarel-yr2015-ley050.htm [accessed 2018-04-15] [WebCite Cache ID 6yi4MmArv]
3. The Parliamentary Office of Science and Technology. Electronic Health Records URL: http://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-0519 [accessed 2018-04-15] [WebCite Cache ID 6yi4GGINB]
4. Delimiter. 2013. Review of the Personally Controlled Electronic Health Record URL: http://delimiter.com.au/wp-content/uploads/2014/05/FINAL-Review-of-PCEHR-December-2013.pdf [accessed 2018-04-15] [WebCite Cache ID 6yi4Ap4aw]
5. Coiera E. Do we need a national electronic summary care record? Med J Aust 2011 Jan 17;194(2):90-92. [Medline: 21241224]
6. Otte-Trojel T, de Bont A, van de Klundert J, Rundall TG. Characteristics of patient portals developed in the context of health information exchanges: early policy effects of incentives in the meaningful use program in the United States. J Med Internet Res 2014;16(11):e258 [FREE Full text] [doi: 10.2196/jmir.3698] [Medline: 25447837]
7. Sheikh A, Jha A, Cresswell K, Greaves F, Bates DW. Adoption of electronic health records in UK hospitals: lessons from the USA. Lancet 2014 Jul 05;384(9937):8-9. [doi: 10.1016/S0140-6736(14)61099-0] [Medline: 24998803]
8. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
9. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. Br Med J 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]
10. Robert Wood Johnson Foundation. Forces4quality. 2012. The Value of Personal Health Records and Web Portals to Engage Consumers and Improve Quality URL: http://forces4quality.org/af4q/download-document/5596/Resource-value_of_personal_health_records_and_web_portals_12-000--for_rwjf_review.pdf [accessed 2018-04-15] [WebCite Cache ID 6yi45I2MW]
11. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008;15(1):1-7 [FREE Full text] [doi: 10.1197/jamia.M2562] [Medline: 17947615]
12. Blum B. Clinical Information Systems. Springer-Verlag Berlin Heidelberg: Springer-Verlag; 1986.
13. Australian Digital Health Agency. DigitalHealth. Software products using digital health URL: https://www.digitalhealth.gov.au/get-started-with-digital-health/what-is-digital-health/who-is-using-digital-health/software-products-using-digital-health [accessed 2018-02-06] [WebCite Cache ID 6x2f74xpX]
14. Commonwealth of Australia. 2013. Personally controlled electronic health record system operator. Annual report 1 July 2012 to 30 June 2013 URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/DA94DAE992F8CDFDCA257C35001DE1E6/$File/PCEHR-System-Operater-Annual-Report-12-13.pdf [accessed 2018-04-18] [WebCite Cache ID 6yleBhltR]
15. Armani R, Mitchell LE, Allen-Graham J, Heriot NR, Kotsimbos T, Wilson JW. Current patient and healthcare worker attitudes to eHealth and the personally controlled electronic health record in major hospitals. Intern Med J 2016 Jun;46(6):717-722. [doi: 10.1111/imj.13086] [Medline: 27040359]
16. Page E, Kemp-Casey A, Korda R, Banks E. Using Australian Pharmaceutical Benefits Scheme data for pharmacoepidemiological research: challenges and approaches. Public Health Res Pract 2015 Sep 30;25(4):e2541546 [FREE Full text] [doi: 10.17061/phrp2541546] [Medline: 26536508]
17. Mellish L, Karanges EA, Litchfield MJ, Schaffer AL, Blanch B, Daniels BJ, et al. The Australian Pharmaceutical Benefits Scheme data collection: a practical guide for researchers. BMC Res Notes 2015 Nov 02;8:634 [FREE Full text] [doi: 10.1186/s13104-015-1616-8] [Medline: 26526064]

18. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak 2008;8:45 [FREE Full text] [doi: 10.1186/1472-6947-8-45] [Medline: 18837999]
19. Williams PA. Does the PCEHR mean a new paradigm for information security? Implications for health information management. Health Inf Manag 2013;42(2):31-36. [doi: 10.12826/18333575.2013.0007.Williams] [Medline: 23695935]
20. Pearce C, Bartlett J, Mcleod A, Eustace P, Amos R, Shearer M. Effectiveness of local support for the adoption of a national programme--a descriptive study. Inform Prim Care 2014;21(4):171-178 [FREE Full text] [doi: 10.14236/jhi.v21i4.70] [Medline: 25479347]
21. Australian Digital Health Agency. 2017. Pathology and Diagnostic Imaging Updates URL: https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/news-033 [accessed 2018-02-06] [WebCite Cache ID 6x2nOdOLn]
22. Australian Medical Association. 2016. Shared electronic medical records URL: https://ama.com.au/position-statement/shared-electronic-medical-records-revised-2016 [accessed 2018-04-16] [WebCite Cache ID 6yi28u22T]
23. Medicare Benefits Schedule Book. Canberra: Australian Government, Department of Health; 2016.
24. van Dooren K, Lennox N, Stewart M. Improving access to electronic health records for people with intellectual disability: a qualitative study. Aust J Prim Health 2013;19(4):336-342. [doi: 10.1071/PY13042] [Medline: 23962451]
25. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med J 1975 May 31;2(5969):486-489 [FREE Full text] [Medline: 1148666]
26. Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B. LifeLines: using visualization to enhance navigation and analysis of patient records. Proc AMIA Symp 1998:76-80 [FREE Full text] [Medline: 9929185]
27. dosReis S, Camelo Castillo W, Gardner J, Tai M. Application of EventFlow to Model the Complexity and Continuity of Antipsychotic Use Among Youth. In: WORKSHOP: Temporal Event Sequence Analysis EventFlow and CoCo User Group Meeting. 2016 Presented at: 33rd Human-Computer Interaction Lab Symposium; May 28, 2015; University of Maryland.
28. Roxon N. Government of Australia. 2010. Personally controlled electronic health records for all Australians URL: http://www.health.gov.au/internet/budget/publishing.nsf/content/FCB15F5C6E04DB36CA257CA0003FF4F9/$File/hmedia09.pdf [accessed 2018-04-16] [WebCite Cache ID 6yi3GLCjM]
29. Bodenheimer T. Coordinating care--a perilous journey through the health care system. N Engl J Med 2008 Mar 6;358(10):1064-1071. [doi: 10.1056/NEJMhpr0706165] [Medline: 18322289]
30. Australian Government, Department of Health. Canberra: Department of Health; 2016. My health record system operator. Annual Report 1 July 2015 to 30 June 2016 URL: https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/0F2F30876A95E7D4CA257F8A0008E337/$File/MyHealthRecordSystemOperatorAnnualReport-2015-16.PDF [accessed 2018-04-18] [WebCite Cache ID 6yleh77oW]
31. Australian Digital Health Agency. 2018. The Healthcare Identifiers Service? How to get your HPI-I URL: https://www.digitalhealth.gov.au/get-started-with-digital-health/registration/workbook/general-practice-registration-workbook/DH-HI%20Service-How%20to%20get%20your%20HPI_v2_20180130.pdf [accessed 2018-04-15] [WebCite Cache ID 6yi3lndHE]
32. Bidargaddi N, Schrader G, Piantadosi C, Adams R, Ryan B, Williams J, et al. Embedding research into clinical practice: the Services, Pathways Access - Research and Knowledge (SPARK) study. Aust J Rural Health 2013 Jun;21(3):192-193. [doi: 10.1111/ajr.12029] [Medline: 23782289]
33. Australian Government, Australian Digital Health Agency. DigitalHealth. Clinical Documents - PCEHR Usability Recommendations v1 URL: https://www.digitalhealth.gov.au/implementation-resources/clinical-documents/EP-1962-2014/NEHTA-1923-2014 [accessed 2018-02-06] [WebCite Cache ID 6x2o0kn3b]
34. Actionable Intime Insights Project. Welcome to the Actionable Intime Insights (AI2) Project. Aisquared 2018 [FREE Full text]
35. Gay V, Leijdekkers P. Bringing health and fitness data together for connected health care: mobile apps as enablers of interoperability. J Med Internet Res 2015;17(11):e260 [FREE Full text] [doi: 10.2196/jmir.5094] [Medline: 26581920]
36. Government of Australia. Canberra; 2012. My health records act URL: https://www.legislation.gov.au/Details/C2017C00313 [accessed 2018-04-15] [WebCite Cache ID 6yi3eVw3b]
37. Twomey C, O'Reilly G. Effectiveness of a freely available computerised cognitive behavioural therapy programme (MoodGYM) for depression: Meta-analysis. Aust N Z J Psychiatry 2017 Mar;51(3):260-269. [doi: 10.1177/0004867416656258] [Medline: 27384752]
38. Blackford K, Jancey J, Lee AH, James AP, Waddell T, Howat P. Home-based lifestyle intervention for rural adults improves metabolic syndrome parameters and cardiovascular risk factors: A randomised controlled trial. Prev Med 2016 Aug;89:15-22 [FREE Full text] [doi: 10.1016/j.ypmed.2016.05.012] [Medline: 27196148]
39. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
40. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
41. Office of Parlimentary Counsel. Therapeutic Goods Act 1989. Canberra: Commonwealth of Australia; 1989.

42. Australian Health Practitioner Regulation Agency. AHPRA. 2018 Feb 06. About the National Scheme URL: http://www.ahpra.gov.au/~/link.aspx?_id=D4E5EF420D3C4EAB8B247FDB72CA6E0A&_z=z [accessed 2018-02-06] [WebCite Cache ID 6x2uhjqaN]
43. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nat Biotechnol 2011 May;29(5):411-414. [doi: 10.1038/nbt.1837] [Medline: 21516084]
44. Almalki M, Gray K, Sanchez FM. The use of self-quantification systems for personal health information: big data management activities and prospects. Health Inf Sci Syst 2015;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S1 [FREE Full text] [doi: 10.1186/2047-2501-3-S1-S1] [Medline: 26019809]
45. Anderl E, Schumann JH, Kunz W. Helping firms reduce complexity in multichannel online data: a new taxonomy-based approach for customer journeys. J Retailing 2016;92(2):185-203. [doi: 10.1016/j.jretai.2015.10.001]

Abbreviations

AI [2]: Actionable Intime Insights application
API: application programming interface
Cwlth: Commonwealth
EHR: electronic health record
HI: Healthcare Identifier verification service
HPI: Healthcare Provider Identifier
IHI: Individual Health Identifier
MBS: Medical Benefits Scheme
MyHR: My Health Record
NOC: Notice of Connection
PBS: Prescription Benefit Scheme

Edited by G Eysenbach; submitted 17.03.17; peer-reviewed by K Mandl, A Fogarty, J Parak, F Sasangohar; comments to author 10.08.17; revised version received 19.12.17; accepted 15.02.18; published 24.04.18.

Please cite as:
Bidargaddi N, van Kasteren Y, Musiat P, Kidd M
Developing a Third-Party Analytics Application Using Australia's National Personal Health Records System: Case Study
JMIR Med Inform 2018;6(2):e28
URL: http://medinform.jmir.org/2018/2/e28/
doi:10.2196/medinform.7710
PMID:29691211

Original Paper

Data Access and Usage Practices Across a Cohort of Researchers at a Large Tertiary Pediatric Hospital: Qualitative Survey Study

Hoi Ki Kiki Ho[1], BSc (Bio); Matthias Görges[1,2], PhD; Elodie Portales-Casamar[1,3], PhD

[1]BC Children's Hospital Research Institute, Vancouver, BC, Canada

[2]Department of Anesthesiology, Pharmacology & Therapeutics, University of British Columbia, Vancouver, BC, Canada

[3]Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Elodie Portales-Casamar, PhD
Department of Pediatrics
University of British Columbia
Room V3-321
950 West 28th Avenue
Vancouver, BC,
Canada
Phone: 1 604 875 2345 ext 4687
Email: eportales-casamar@bcchr.ca

Abstract

Background: Health and health-related data collected as part of clinical care is a foundational component of quality improvement and research. While the importance of these data is widely recognized, there are many challenges faced by researchers attempting to use such data. It is crucial to acknowledge and identify barriers to improve data sharing and access practices and ultimately optimize research capacity.

Objective: To better understand the current state, explore opportunities, and identify barriers, an environmental scan of investigators at BC Children's Hospital Research Institute (BCCHR) was conducted to elucidate current local practices around data access and usage.

Methods: The Clinical and Community Data, Analytics and Informatics group at BCCHR comprises over 40 investigators with diverse expertise and interest in data who share a common goal of facilitating data collection, usage, and access across the community. Semistructured interviews with 35 of these researchers were conducted, and data were summarized qualitatively. A total impact score, considering both frequency with which a problem occurs and the impact of the problem, was calculated for each item to prioritize and rank barriers.

Results: Three main themes for barriers emerged: the lengthy turnaround time before data access (18/35, 51%), inconsistent and opaque data access processes (16/35, 46%), and the inability to link data (15/35, 43%) effectively. Less frequent themes included quality and usability of data, ethics and privacy review barriers, lack of awareness of data sources, and efforts required duplicating data extraction and linkage. The two main opportunities for improvement were data access facilitation (14/32, 44%) and migration toward a single data platform (10/32, 31%).

Conclusions: By identifying the current state and needs of the data community onsite, this study enables us to focus our resources on combating the challenges having the greatest impact on researchers. The current state parallels that of the national landscape. By ensuring protection of privacy while achieving efficient data access, research institutions will be able to maximize their research capacity, a crucial step towards achieving the ultimate and shared goal between all stakeholders—to better health outcomes.

(JMIR Med Inform 2018;6(2):e32) doi:10.2196/medinform.8724

KEYWORDS

clinical data sharing, research barriers, data linkage, data sources, data management, environmental scan, research facilitation

Introduction

The use of data is a foundational component of both research and health care. Health and health-related data are generated at high volumes and are not limited to front-end clinical data [1]. Secondary sources of data include medical imaging, laboratory, insurance, and demographic data, and particularly patient-collected data like activity, nutrition, and other qualitative data; these data add substantial information to the mass of overall health-related data [1]. These complex and interconnected datasets are commonly referred to as "big data", which is often formally defined as large and complex datasets that require specialized software for manipulation and analysis [1,2]. "Big data" is projected to grow at an accelerated pace; for example, the size of health and health-related data in the United States is expected to reach the scale of yottabytes (10^{24} gigabytes) soon [1]. This rapid expansion of health care data is recognized globally, and the ability to access and analyze this wealth of information might allow us to better support a wide range of medical and health care functions, like public health surveillance, population health management, and real-time clinical decision support [1-6].

As research takes on an increasingly data intensive and global focus, there is an increased need for appropriate data sharing, storage and maintenance infrastructure at research institutes engaging in big data analytics [6-12]. There are many benefits of data sharing: it allows for replication and validation of scientific outcomes and results, projects can be extended and viewed from different perspectives, and data re-collection can be minimized [6,8,9]. Infrastructure that supports data sharing, along with the appropriate storage and maintenance of data, maximizes its value and contribution to research [6,7,9,10].

In a study conducted by the Publishing Research Consortium in 2010, approximately two-thirds of the 3823 respondents identified access to datasets, data models and algorithms, and programs as being important to very important, but only about a third of them perceived these resources to be easily accessible [8]. A subsequent survey administered by Tenopir et al [7] in 2014 around perceptions and practices pertaining to data sharing revealed that 85% of 1329 participating scientists would be interested in datasets generated by other researchers or institutions if they were easily accessible. Additionally, 67% viewed the lack of access to these datasets as an impediment to scientific progress, while less than half reported being satisfied with the integration of data from other sources or the availability of different types of data to answer research questions [7].

While there is consensus that data sharing is an integral part of scientific research, there are barriers that contribute to the disparity between the desire to share data and the perceived accessibility of data [2,3,6-10]. Logistical barriers to developing standardized data sharing systems or processes, or a centralized repository for data sharing is a shared challenge among research institutions [1-5,9-12]. For example, to consolidate disparate data sources, datasets must be generated in an "analysis-ready format." This poses several methodological challenges: data harmonization is complicated by the heterogeneity of data sources (the types of data collected and the mechanisms used to collect them) and the availability and usability of data hosted in current electronic health records systems [4,5,6,10,11]. Further, other concerns with data access and sharing, common across research institutions, are confidentiality of potentially re-identifiable data and ethical concerns around consent—has it been given and does it extend to data usage by other parties [1-6]?

In Canada, while health care systems and innovation are highly valued, researchers have faced challenges with striking a balance between enabling timely access to data for research purposes and protecting patient confidentiality [2,3,10]. A major barrier is the inconsistent interpretation of privacy legislation, which varies by province and has led to varying requirements for research ethics board approval, privacy impact assessments, and related data access processes, with turnaround times ranging from months to years [10].

Challenges and concerns around data access are especially pertinent to investigators at BC Children's Hospital Research Institute (BCCHRI), where discovery, translational, and clinical research is conducted to benefit the health of children and their families; at this center, many collaborations are national or global. Many frameworks identify big data through three dimensions: volume, variety, and velocity [13]. Much of the research work conducted at BCCHRI fits under one or more of these dimensions, as our hospital site sees over 200,000 patients annually, from which it collects a large volume of varied data from patients consenting to participate in local studies [14]. These data include clinical parameters and notes, questionnaire responses, medical imaging data, high-density vital sign recordings, multi-omics datasets, and many more. These data are collected in real-time, creating and contributing to various databases, databanks and registries. Specifically, the Clinical and Community Data, Analytics and Informatics Group (data group) engages in such work. Within the research institute's "Evidence-to-Innovation" theme, this group is composed of over 40 BCCHRI investigators with diverse expertise and interest in data who share a common goal of facilitating data collection, usage, and access across our community. Researchers on site have experienced increasing challenges with accessing data for research. Thus, a local environmental scan was performed to a) evaluate and review the state of the data access infrastructure at BCCHRI; b) identify barriers and opportunities; and c) provide feedback to the institute's leadership to help improve data access and usage on site.

Methods

This environmental scan was a quality improvement activity. The University of British Columbia and Children's and Women's Health Centre of British Columbia Research Ethics Board does not review quality assurance or quality improvement studies, in accordance with Article 2.5 of the Tri-Council Policy Statement 2. Following standard methodology for qualitative research [8], semistructured one-to-one interviews were conducted between May and August 2016, with members of the data group, focusing on both their data needs and their experiences accessing and using data on the BC Children's Hospital (BCCH) campus. With consent from the interviewee,

interviews were tape-recorded for transcription of notes, at which point the recordings were destroyed. Each respondent was assigned a participant code (P#). A full list of interview questions can be found in Multimedia Appendix 1, and the list of datasets provided to participants (referenced in Question 1) can be found in Multimedia Appendix 2.

Quantifiable metrics from multiple choice questions, like data needs and expertise, were gathered using paper questionnaires and summarized using Excel (Microsoft, Redmond, WA). The unstructured descriptions of individual experiences with data access and usage were analyzed and synthesized using a template analysis approach [15]. The initial template was defined a priori with three parent themes (barriers for data access, facilitators, and opportunities). The final template used to code and analyze all interview data can be found in Multimedia Appendix 3, which includes additional sub-themes to further describe the three parent themes in the initial template. Relevant quotes were extracted from the interview transcripts to further illustrate respondents' experience with data access and usage, which is a common means of textual data presentation in the template analysis approach. Each quote is attributed to the corresponding respondent using their participant code.

To prioritize and rank barrier items, a total impact score was calculated. This score is analogous to the severity ratings proposed in Jakob Nielsen's usability methodology [16], where a composite score is derived from both the frequency with which a problem occurs and the impact of the problem. For the purposes of this scan, we used the following terminology: *total impact score* for each barrier = *frequency* of mention x *mean effect score* across all items tagged under this barrier. The effect score for each item ranged from 1 (minimal) to 3 (severe) based on the participant's description of how much it affected their research.

Results

Thirty-five of the 43 data group members participated in the environmental scan, constituting an 81% response rate.

Participant Characteristics

Expertise in the data group represents a wide range of specialties, with respondents of the scan mainly identifying clinical data (22/35, 63%) as their "core" expertise, followed by data analysis (13/35, 37%), data standardization and harmonization (11/35, 31%), and administrative data (11/35, 31%; Table 1). Further, when asked if any further expertise was required to advance their work, respondents reported needing additional support in statistics (18/35, 51%) and navigating data access processes (17/35, 49%).

Data Needs

When asked to identify their current data needs, most respondents identified improved access and facilitated data linkage as important data needs (20/35, 57%), followed by the need to bridge clinical and research data (18/35, 51%) and improved usability of electronic health records data (14/35, 40%; Table 2).

Barriers

The three greatest challenges to accessing and using data for research were lengthy turnaround times (18/35, 51%), inconsistent and opaque data access processes (16/35, 46%), and the inability to link data (15/35, 43%; Figure 1, see part a). All barriers were ranked using their total impact score and analyzed in detail (Figure 1, see part b).

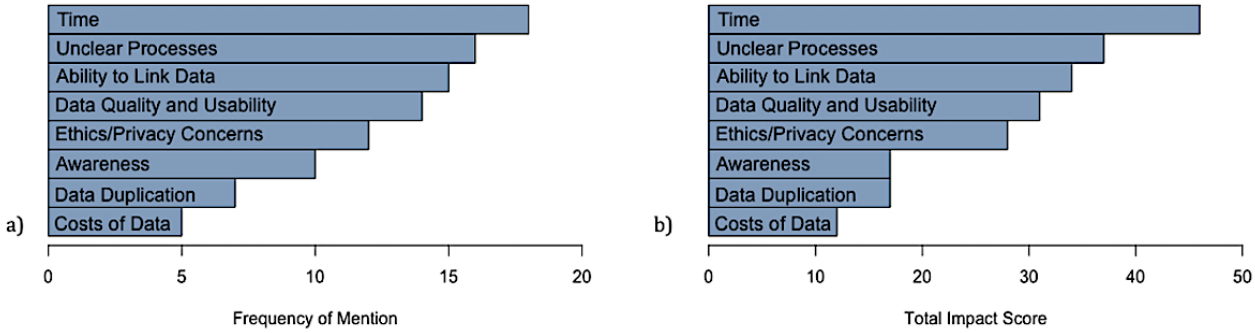
Table 1. Collective expertise of data group members. The data group hosts a wide breadth of expertise with clinical data being most prevalent, followed by data analysis and administrative data. This table lists all categories selected by 5 or more respondents. Other categories with <5 respondents included: clinical expertise (4, 11%), data linkage (4, 11%), database design and building (4, 11%), mobile apps (4, 11%), data integration across modalities, (3, 9%) experience with data stewards (3, 9%), population level data (2, 6%), child pyschology (1, 3%), health surveillance (1, 3%), intervention design (1, 3%), machine learning (1, 3%), and privacy and security (1, 3%).

Identified expertise	n (%)
Clinical data	22 (63)
Data analysis	13 (37)
Administrative data	11 (31)
Data standardization/harmoinzation	11 (31)
Registry/database	10 (29)
Biostatistics	8 (23)
National data networks	8 (23)
Data visualization	7 (20)
Genomics data	7 (20)
Data mining	6 (17)
Epidemiology	5 (14)
International data networks	5 (14)

Table 2. Identified data needs. Facilitated data linkage, improved data access and bridging clinical and research data were the three most frequently mentioned data needs.

Data needs	n (%)
Facilitated data linkage	20 (57)
Improved data access	20 (57)
Bridging clinical and research data	18 (51)
Improved quality of e-records	14 (40)
Access to expertise	10 (29)
Data governance	10 (29)
Permission to contact	5 (14)
Registry framework	3 (9)
Storage space	1 (3)

Figure 1. Frequency of mention (a) and total impact score (b) of the barriers on respondents' research. The barriers that have the most substantial impact on respondents' research are lengthy turnaround times, inconsistent and unclear data processes, and limited capacity for data linkage.



Time

The "Time" barrier was mentioned by 18/35 (51%) respondents, with a mean effect score of 2.56 (median 3; range=2-3), and a total impact score of 46. Most respondents identified the long turnaround time for processing and completing data requests as one of the greatest challenges when trying to access data outside of their primary collection, with instances of waiting up to 7 years to receive datasets and multiple rounds of back and forth communication with different data custodians. Some respondents report waiting for several years without the data request ever reaching completion or receiving approval:

> *I didn't want to be limited to local data and I wanted to be able to compare the patient population coming through BC Children's Hospital and provincial level data. I put members on my team on the task, but it never progressed anywhere. The process was too difficult and took too long: it was taking years. Eventually, we just gave up and moved on to other things that took priority.* [P23]

Unclear Processes

The "Unclear Processes" barrier was mentioned by 16/35 (46%) respondents, with a mean effect score of 2.31 (median 2; range=2-3), and a total impact score of 37. Issues were largely related to 1) lack of a central resource, and 2) a lack of consistency and standardization across different data custodians about access procedures. Lacking a central data access resource leaves researchers without guidance on how to approach accessing data outside their primary collection; that is, being unaware of who to contact, what the data access processes entail, and what data is available:

> *It's unclear as to who will make the decisions, who will provide the approval, who will review the paperwork and look at the privacy impact. This needs to be cleared up and formalized and communicated so that it's clear who to talk to in order to get access to this data, and we need someone to facilitate this.* [P10]

Furthermore, the data access processes are often inconsistent and unclear. Respondents noted that the data access processes are highly variable, especially between different data custodians. Researchers feel as though each time a new project is started, they are starting from scratch and responses emphasized the need to streamline these processes.

Ability to Link Data

The "Ability to Link Data" barrier was mentioned by 15/35 (43%) respondents, with a mean effect score of 2.27 (median 2; range=1-3), and a total impact score of 34. A common concern with many researchers is that the current data infrastructure encourages the creation of "silos," in which data exists isolated within certain divisions, or is restricted to certain projects. Respondents note a lack of official guidance or established infrastructure to facilitate data linkage between disciplines or between internal and external data sources:

> *If you look at other leading children's hospitals around the world, there are mechanisms by which patients and families donate their data and information for research purposes in very broad and powerful ways, and in the Canadian environment, that's more challenging in terms of how we handle data privacy and at the level of the Stanford's and Hopkins', one really needs to have a mechanism by which patients are able to donate or release their data for research and my experience in the BC Children's environment, most patients are actually shocked when they find out we're not using their data.* [P30]

Data Quality and Usability

The "Data Quality and Usability" barrier was mentioned by 14/35 (40%) respondents, with a mean effect score of 2.21 (median 2; range=1-3), and a total impact score of 31. Internally, the current state of electronic health records poses a challenge for researchers, as the data are not truly electronic, such that data is not stored in an electronically extractable format. Thus, manual transcription is still required to extract the data, with the possibility of transcription error. This greatly limits the campus' ability to contribute to and participate in larger national and international databases. Externally, many variables requested by researchers are unavailable, not defined clearly, or in an inappropriate format, thus requiring further back and forth communication between researchers and data custodians:

> *The data received from the steward is messy, as in it isn't organized in a common-sense way. I couldn't tell which participant answered which question, which then required a constant regeneration of the dataset to have it organized in a meaningful way. Unfortunately, I can't avoid this because investigators themselves do not have access to the raw data, so I couldn't even match the data up or re-organize the data on my own and had to engage in this constant back and forth with the steward.* [P2]

Ethics and Privacy Concerns

The "Ethics and Privacy Concerns" barrier was mentioned by 12/35 (34%) respondents, with a mean effect score of 2.33 (median 2; range=2-3), and a total impact score of 28. Some examples of these obstacles include not having permission to contact patients and their families, the lack of consistency as to when and if a Privacy Impact Assessment (PIA) is required for a project, and the varying requirements across ethics boards for multi-site projects.

> *The unknown variable and obstacle is that it is unclear when and for which projects the PIA is required, the process has been very inconsistent, with a lot of back and forth, often asking for information that has already been provided. I find the PIA processes inconsistent not only across health authorities but even within health authorities.* [P26]

Awareness

The "Awareness" barrier was mentioned by 10/35 (29%) respondents, with a mean effect score of 1.70 (median 1.5; range=1-3), and a total impact score of 17. Responses show that there are some researchers who are completely unaware of the data sources that are available to them outside their primary collection or collaborations with others:

> *I haven't heard of or used any of the sources listed here, so I haven't had any experience with these data holders as I didn't know that these sources existed. I've only used data through my own primary collection, but I would like to learn more about how to access these and what types of data is available.* [P18]

Data Duplication

The "Data Duplication" barrier was mentioned by 7/35 (20%) respondents, with a mean effect score of 2.43 (median 2; range=2-3), and a total impact score of 17. The manual transcription required to extract data from local systems, and the inability to link datasets across different projects and studies, leads to the continued duplication of data. Respondents noted that many studies collect the same basic package of information (eg, demographics), which further contributes to repeated and isolated datasets existing across the campus:

> *What I find happens a lot here is that there's duplication in data collection, and if we had a way to collect a base level of data on all the kids coming to the hospital, like a standardize form, especially to make it easier to be integrated into electronic health records and pulled, I think that would really save time as opposed to every time there's a new project, you pull the same data and some poor med student is manually extracting it. There could be errors there, if we could somehow connect it via a system with accurate and secure information that would be extractable, that would be great. I know there's lots of red tape around this, in the sense we can't even get such a system running, let alone use it for research, but I think ultimately that's what we need.* [P32]

Costs of Data

The "Costs of Data" barrier was mentioned by 5/35 (14%) respondents, with a mean effect score of 2.40 (median 2; range=2-3), and a total impact score of 12. Data requests are often associated with significant costs, and acquiring funding continues to be difficult for many researchers, especially when the data requests are often onerous and funding is typically provided only for a limited time span.

Facilitators

Some facilitators in navigating these challenges were identified by 17/35 (49%) respondents. Existing rapport with key contacts from data sources is a major facilitator to the success that some researchers have had (9/17, 53%). Although this has proven beneficial for those who had these existing networks, it does represent a barrier to those without them. Researchers also note

that they will rely on primary collection or use publicly available data when possible (5/17, 29%). However, using data sources with clearly outlined data access processes and existing infrastructure to support their data requests (eg, Population Data BC) is a facilitator for those who do attempt to access external datasets (3/17, 18%).

Opportunities

Opportunities for the data group were identified by 32/35 (91%) respondents. The following categories emerged: data access facilitation (14/32, 44%) and migration toward a single data platform (10/32, 31%).

Data Access Facilitation

It was suggested that a support unit or a central resource dedicated to data access would be highly beneficial as a centralized and focused support system does not seem to currently be in place. The hope is for the potential team to facilitate the entire data access process, from consultation to support with data request logistics (eg, data request forms).

Single Data Platform

Respondents would like to explore the opportunity of developing a single platform where existing data could be linked, and new data can be entered through single point of entry. It would have infrastructure built to collect a set of standardized variables from all patients and the capacity to be adapted for specific projects. This would limit data duplication through different prospective studies collecting the same variables. The possibility was also mentioned of having a patient portal in such a system to allow patients to contribute data on their own accord.

Data Sources and Management Tools

With respect to data sources, five respondents had no previous experience accessing external datasets. Other participants most frequently accessed datasets from the hospital clinical data warehouse (13/30, 43%), through Population Data BC (popData) [17] (9/30, 30%), or the Canadian Institute for Health Information [18] (7/30, 23%). Other datasets used include BC Perinatal Data Registry [19], BC Children's Hospital Biobank [20], Canadian Neonatal Network [21], Canadian Neonatal Follow-Up Network [22], or Edudata [23].

The most commonly used statistical and computing tools were SPSS [24] (20/35, 57%), and R [25] (19/35, 54%), others including SAS [26], STATA [27], MATLAB [28], and Python [29]. The most commonly used data management tools were REDCap [30] (28/35, 80%) and MS-Excel [31] (26/35, 74%), with additional tools with low usage (MS Access [32], custom databases, Dropbox [33], Dacima [34], and various survey tools).

Discussion

Principal Findings

Timely access to health and health-related data is crucial to advancing health care systems and stimulating innovation to improve quality of care [1-10]. BCCHRI houses a wide breadth of topics and relies on many different data sources. The most critical data needs identified by respondents, like improved access and facilitated data linkage, directly reflect the challenges currently faced; for example, the lengthy turnaround time and the opaque and highly variable data access processes. These factors are detrimental to current research endeavors, and often result in researchers refraining from using existing data, but rather collecting it again through a prospective study.

A need to create resources to facilitate and support data access and ultimately to move towards a single data platform that will allow comprehensive, linked, clean and processed clinical data, not isolated by discipline or disease, is strongly evident from this scan. An increased capacity for data linkage also improves the site's ability to participate in and contribute to national and international projects and registries. Furthermore, there is an apparent lack of awareness of the datasets available, and how to gain access to them. Most researchers will use the bigger, more centralized resources such as the hospital clinical data warehouse or popData, which have better defined processes and points of contact, rather than the smaller isolated dataset with no clear shop front. This challenge of having documentation for such processes and methods to gather and link disparate sources of data are echoed in the literature [2,3,6,9,10]. This highlights the need for a centralized source of information, which could take the form of a repository or a data navigator role, to connect researchers with these isolated datasets, thereby enhancing their utilization and maximizing the value of the data.

REDCap usage is prevalent, probably due to its ease of accessibility, not only at BCCH, but also at many different sites across Canada (allows for easy collaboration), the ease of Research Ethics Board approval for its use, and low cost. The use of these tools is essential to streamline and standardize data management and analysis practices. They emphasize how critical it is to have support systems broadly available to our community and to have central access to, and support for, specialized statistical software such as SPSS and R.

These opportunities are real and would bring great benefit to both researchers and patients by increasing the value of the data they contribute [1-12]. However, there are logistical and administrative challenges that are difficult to overcome [2,4,5,6,9,10,12]. The interpretation of the privacy legislation mandating data access mechanisms is at the discretion of each individual data custodian and steward and can be hard to harmonize. This is consistent with other reports that note strong variation in the interpretation of privacy legislation, which lead to variable data access processes and inconsistencies in access time [2,3,6,7,9,10].

Furthermore, data governance needs to be clearly established, particularly when applying data linkage and integration between existing data sources, to define clear rules and oversight for the data access platforms and mechanisms. This is consistent with the findings from the 2015 Accessing Health and Health-Related Data in Canada report, which cited strong and clear governance models, a willingness to enable appropriate use of data, recognizing that risk cannot always be completely eliminated, and establishing explicit guidelines for privacy risk assessment as principles for success at "best practice" institutions [5]. While obstacles do exist, creating a system that allows for timely data access while simultaneously protecting and respecting

confidentiality is feasible and has been demonstrated in "best practices" entities such as the Farr Institute in Scotland and the Wales Secure Anonymized Information Linkage Databank [10]. We can learn from these institutions' successes in mitigating barriers to data access now that needs in this area have been identified and prioritized.

Limitations of the Environmental Scan

Limitations to our study include the small sample size, as it was conducted at a single center and only 35 researchers of the entire research community participated, which limits generalizability. However, since the data group was formed as an open forum that any BCCH researcher could join when their research includes a strong data component, we believe that our respondent sample includes most of the knowledge and expertise related to data usage and access in our community. This in-depth work, although at the level of a single institution, has implications far beyond it, as the patient population that passes through BCCH is representative of patients across the entire province, and BCCH is a prominent partner in many national initiatives and international data networks to improve research in health care. This allows the results of this study to propagate beyond this institution alone. Also, based on the supporting literature, these themes are common among many institutions globally. This work represents a systematic way of identifying and prioritizing barriers and opportunities to data access and usage, which can be shared and reflected upon among different provinces and health authorities. As such, this work has played a part in motivating the changes made to privacy review processes at the Provincial Health Services Authority (PHSA), which introduced a new Privacy Advisor position that works directly with PHSA researchers and staff to identify privacy and security risks. This new role is intended to streamline the privacy review process while also ensuring that research conducted in PHSA institutions is carefully reviewed for privacy considerations. Environmental scans, such as ours, can demonstrate impact, which lies in policy and governance changes, as well as communicating these challenges, best practices and potential solutions among the research community.

In addition, as interviews were semistructured, a variable amount of data was captured for each participant. For example, the responses to open ended questions regarding barriers, facilitators and opportunities yielded varying levels of detail from each respondent. Additionally, participants' selection options changed as the scan progressed, as the lists provided to them grew during data collection. To prioritize and rank barrier items, we used a total impact score, which is derived from both the frequency of mention of a problem and the effect of the problem. While the frequency is objectively measured, the effect is determined by the interviewer based on the interviewee's comments. We note that both trends are similar despite a slight exaggeration of the Time and Awareness barriers, which shows that even though the effect is subjectively measured, it doesn't influence the total impact score considerably (Figure 1). Furthermore, for some metrics, only a subset of participants was able to contribute; for example, only those with previous experience requesting a dataset from a custodian would be able to contribute to the question related to previous data sources used.

Conclusion

In an era of increasing digitization of information and globalization, the demand and need for health and health-related data will continue to grow. By identifying the current state and needs of the data community onsite, this study enables us to focus our resources on combating the challenges having the greatest impact on researchers. The current state of BCCHRI parallels that of the national landscape, and by looking towards organizations that have been able to ensure protection of privacy while achieving efficient data access, the institute will be able to maximize their research capacity. Solutions do exist and acknowledging problem areas and taking action is the first step towards achieving the ultimate and shared goal between all stakeholders—to better health outcomes.

Acknowledgments

The authors wish to thank all participating members from the Clinical and Community Data, Analytics and Informatics group, Evidence to Innovation theme, BC Children's Hospital Research Institute for their willingness to participate in the interviews and feedback, as well as Dawn Mount and Ivan Cepeda for administrative support and thoughtful comments, and Angelica Leon for her insights regarding template analysis. Funding was provided by the Evidence to Innovation theme at BC Children's Hospital Research Institute.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Environmental scan interview questions.

[PDF File (Adobe PDF File), 48KB - medinform_v6i2e32_app1.pdf]

Multimedia Appendix 2

List of available datasets provided to investigators.

[PDF File (Adobe PDF File), 33KB - medinform_v6i2e32_app2.pdf]

Multimedia Appendix 3

Final analysis template.

[PDF File (Adobe PDF File), 11KB - medinform_v6i2e32_app3.pdf]

References

1. 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]
2. Canadian Institute for Health Information. Ottawa, ON; 2013. Better Information for Improved Health: A Vision for Health System Use of Data in Canada URL: https://www.cihi.ca/en/hsu_vision_report_en.pdf [accessed 2018-01-22] [WebCite Cache ID 6wh0gLwoh]
3. Canadian Health Information Management Association. 2016. Big Data and Data Analytics URL: https://www.echima.ca/uploaded/pdf/emails/0041.16_Big%20Data%20and%20Data%20Analytics.pdf [accessed 2018-01-21] [WebCite Cache ID 6wh0p9yUC]
4. Keen J, Calinescu R, Paige R, Rooksby J. Big data + politics = open data: The case of health care data in England. POI 2013 Jul 22;5(2):228-243. [doi: 10.1002/1944-2866.POI330]
5. Jee K, Kim G. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. Healthc Inform Res 2013 Jun;19(2):79-85 [FREE Full text] [doi: 10.4258/hir.2013.19.2.79] [Medline: 23882412]
6. van PWG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health 2014 Nov 05;14:1144 [FREE Full text] [doi: 10.1186/1471-2458-14-1144] [Medline: 25377061]
7. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PLoS One 2011;6(6):e21101 [FREE Full text] [doi: 10.1371/journal.pone.0021101] [Medline: 21738610]
8. Publishing Research Consortium. 2010. Access vs Importance: A global study assessing the importance of and ease of access to professional and academic information Phase I Results URL: http://www.publishingresearchconsortium.com/index.php/prc-documents/prc-research-projects/19-prc-access-vs-importance/file [accessed 2017-08-02] [WebCite Cache ID 6sPhX2U3s]
9. Pearce N, Smith AH. Data sharing: not as simple as it seems. Environ Health 2011 Dec 21;10:107 [FREE Full text] [doi: 10.1186/1476-069X-10-107] [Medline: 22188646]
10. Council of Canadian Academies. Ottawa, ON; 2015. Accessing Health and Health Related Data URL: http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/health-data/healthdatafullreporten.pdf [accessed 2017-08-01] [WebCite Cache ID 6sQ5w1KVw]
11. Cragin MH, Palmer CL, Carlson JR, Witt M. Data sharing, small science and institutional repositories. Philos Trans A Math Phys Eng Sci 2010 Sep 13;368(1926):4023-4038 [FREE Full text] [doi: 10.1098/rsta.2010.0165] [Medline: 20679120]
12. Faniel I, Zimmerman A. International Journal of Digital Curation. 2011 Mar 11. Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse URL: http://ijdc.net/index.php/ijdc/article/view/163/231 [accessed 2018-01-20] [WebCite Cache ID 6wh08zh9D]
13. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 2015 Apr;35(2):137-144. [doi: 10.1016/j.ijinfomgt.2014.10.007]
14. BC Children's Hospital Website. Our Unique Role URL: http://www.bcchildrens.ca/about/our-unique-role [accessed 2018-04-03] [WebCite Cache ID 6yOfHe7Xk]
15. King N. Essential guide to qualitative methods in organizational research. 2004. Using templates in the thematic analysis of text URL: https://www.researchgate.net/file.PostFileLoader.html?id=58640e5f40485448d21b27e8&assetKey=AS%3A444348253904897%401482952287371 [accessed 2018-01-20] [WebCite Cache ID 6wh0HGNe2]
16. Nielsen J. Usability Inspection Methods Internet. In: Nielsen Norman Group. New York, NY: John Wiley & Sons; 1994:-471.
17. Population Data BC. URL: https://www.popdata.bc.ca [accessed 2017-08-02] [WebCite Cache ID 6sQ48uF8h]
18. Canadian Institute for Health Information. URL: https://www.cihi.ca/en [accessed 2017-07-30] [WebCite Cache ID 6sQ4EWidw]
19. BC Perinatal Data Registry. URL: http://www.perinatalservicesbc.ca/health-professionals/data-surveillance/perinatal-data-registry [accessed 2017-08-01] [WebCite Cache ID 6sQ4JPHoZ]
20. BC Children's Hospital BioBank. URL: https://bcchr.ca/biobank [accessed 2017-07-31] [WebCite Cache ID 6sQr4Yo0o]
21. The Canadian Neonatal Network. URL: http://www.canadianneonatalnetwork.org/portal [accessed 2017-07-31] [WebCite Cache ID 6sQr8ur3u]
22. Canadian Neonatal Follow-Up Network. URL: http://www.cnfun.ca [accessed 2017-07-31] [WebCite Cache ID 6sQrBqc7s]
23. Edudata Canada. URL: https://edudata.ca/site/home/ [accessed 2017-07-31] [WebCite Cache ID 6sQ4JPHoZ]
24. IBM SPSS. URL: http://www.ibm.com/analytics/us/en/technology/spss/ [accessed 2017-07-31] [WebCite Cache ID 6sQrO0ohN]

25. R: The R Project for Statistical Computing. URL: https://www.r-project.org [accessed 2017-07-31] [WebCite Cache ID 6sQrOq463]
26. SAS: Analytics, Business Intelligence and Data Management. URL: https://www.sas.com/en_ca/home.html [accessed 2017-07-31] [WebCite Cache ID 6sQrPaBAy]
27. Stata: Data Analysis and Statistical Software. URL: http://www.stata.com/ [accessed 2017-08-01] [WebCite Cache ID 6sQrQggN2]
28. MATLAB - MathWorks. URL: https://www.mathworks.com/product/matlab.html?s_tid=hp_products_matlab [accessed 2017-08-01] [WebCite Cache ID 6sQrROqrV]
29. Python. URL: https://www.python.org/ [accessed 2017-08-01] [WebCite Cache ID 6sQrlo25w]
30. REDCap - Research Electronic Data Capture. URL: https://www.project-redcap.org/ [accessed 2017-08-01] [WebCite Cache ID 6sQrSTWGJ]
31. Microsoft Excel Spreadsheet Software. URL: https://products.office.com/en-ca/excel [accessed 2017-08-01] [WebCite Cache ID 6sQrTKhrc]
32. Microsoft Access: Database Software and Applications. URL: https://products.office.com/en-ca/access [accessed 2017-08-01] [WebCite Cache ID 6sQrUFPLM]
33. Dropbox. URL: https://www.dropbox.com [accessed 2017-08-01] [WebCite Cache ID 6sQrVMzyw]
34. DACIMA: Electronic Data Capture & Clinical Trial Management Software. URL: http://www.dacimasoftware.com/ [accessed 2017-08-01] [WebCite Cache ID 6sQrkijjU]

Abbreviations

BCCHRI: BC Children's Hospital Research Institute
BCCH: BC Children's Hospital
PHSA: Provincial Health Services Authority
PIA: Privacy Impact Assessment

Edited by G Eysenbach; submitted 16.08.17; peer-reviewed by F Shen, F Wang; comments to author 25.11.17; revised version received 02.02.18; accepted 10.04.18; published 14.05.18.

Please cite as:
Ho HKK, Görges M, Portales-Casamar E
Data Access and Usage Practices Across a Cohort of Researchers at a Large Tertiary Pediatric Hospital: Qualitative Survey Study
JMIR Med Inform 2018;6(2):e32
URL: http://medinform.jmir.org/2018/2/e32/
doi:10.2196/medinform.8724
PMID:29759958

Original Paper

Privacy-Preserving Predictive Modeling: Harmonization of Contextual Embeddings From Different Sources

Yingxiang Huang[1*], BA; Junghye Lee[2,3,4*], PhD; Shuang Wang[1], PhD; Jimeng Sun[5], PhD; Hongfang Liu[6], PhD; Xiaoqian Jiang[1], PhD

[1]Health Sciences, Department of Biomedical Informatics, University of California - San Diego, La Jolla, CA, United States
[2]School of Management Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic Of Korea
[3]Department of Biomedical Informatics, University of California - San Diego, La Jolla, CA, United States
[4]Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Republic Of Korea
[5]School of Computational Science and Engineering at College of Computing, Georgia Institute of Technology, Atlanta, GA, United States
[6]Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, United States
[*]these authors contributed equally

Corresponding Author:
Yingxiang Huang, BA
Health Sciences
Department of Biomedical Informatics
University of California - San Diego
9500 Gilman Drive
La Jolla, CA, 92093
United States
Phone: 1 858 534 2230
Email: yih108@eng.ucsd.edu

Abstract

Background: Data sharing has been a big challenge in biomedical informatics because of privacy concerns. Contextual embedding models have demonstrated a very strong representative capability to describe medical concepts (and their context), and they have shown promise as an alternative way to support deep-learning applications without the need to disclose original data. However, contextual embedding models acquired from individual hospitals cannot be directly combined because their embedding spaces are different, and naive pooling renders combined embeddings useless.

Objective: The aim of this study was to present a novel approach to address these issues and to promote sharing representation without sharing data. Without sacrificing privacy, we also aimed to build a global model from representations learned from local private data and synchronize information from multiple sources.

Methods: We propose a methodology that harmonizes different local contextual embeddings into a global model. We used Word2Vec to generate contextual embeddings from each source and Procrustes to fuse different vector models into one common space by using a list of corresponding pairs as anchor points. We performed prediction analysis with harmonized embeddings.

Results: We used sequential medical events extracted from the Medical Information Mart for Intensive Care III database to evaluate the proposed methodology in predicting the next likely diagnosis of a new patient using either structured data or unstructured data. Under different experimental scenarios, we confirmed that the global model built from harmonized local models achieves a more accurate prediction than local models and global models built from naive pooling.

Conclusions: Such aggregation of local models using our unique harmonization can serve as the proxy for a global model, combining information from a wide range of institutions and information sources. It allows information unique to a certain hospital to become available to other sites, increasing the fluidity of information flow in health care.

(JMIR Med Inform 2018;6(2):e33) doi:10.2196/medinform.9455

KEYWORDS

interoperability; contextual embedding; predictive models; patient data privacy

Introduction

Motivation

As large datasets from different areas ranging from genetics, microbiomes, nutrients, medicine, medical devices to the environment are being collected from large populations, it is believed that more efforts should be spent on reshaping the wealth of data and utilizing them to promote precision medicine [1]. The characterization of each person on a multidimensional level might lead to far more intricate diagnostic and prognostic groupings of populations and labeling of individuals [2]. Pertinent studies include finding relevant biomarkers, distinguishing patterns for rare diseases, discovering the combined effects of multiple genetic variants or epistasis, and researching the unique phenotype of diseases that only appears in certain demographics or ethnicities. All of them require a large sample size to avoid false positives and insignificant results [3,4].

To gather such large samples, there have been some efforts to share deidentified data such as clinical notes in compliance with the Health Insurance Portability and Accountability Act (HIPAA) [5]. However, permissions to access other's data in a central warehouse are still cumbersome to obtain, and deidentification efforts are either costly, error prone, or ineffective [6]. Human-based deidentification efforts cost over 5000 hours and US $500,000 on the Medical Information Mart for Intensive Care-III (MIMIC-III) dataset [7] which contains only about 50,000 patient visits and 100 million words [8] and produces error recall ranging from 0.63 to 0.94 [9]. Machine-assisted deidentification shows varying results from time savings of 13.85% to 21.5% to results showing no improvement in either quality or time saved [10]. Machine learning, algorithm-based, automated deidentification can be very useful, but state-of-the-art deep learning–based deidentification models for unstructured data is still incapable of reaching the level of privacy protection set by HIPAA safe harbor, which has roughly a 0.013% reidentification rate [8,11]. In the biomedical community, there is an urgent need for developing a new method to share information learned from local sources to generalize and scale up research effort.

Objective

Our objective to address the above challenges is to create a federated clinical analysis framework through the aggregation of local representations and models. Related studies have been published, focusing on not only simple analyses such as database queries with very specific inclusion or exclusion criteria, but also sophisticated algorithms for prediction analysis, including logistic regression [12,13], support vector machine [14,15], *k*-nearest neighborhood [16], Cox regression [17], and tensor factorization [18]. However, most studies involve restrictive assumptions originating from the requirement that data should be integrated in a matrix format, either common feature events assumption for horizontally partitioned data or common patient records assumption for vertically partitioned data. Both assumptions have limitations to reflect the situations in reality. For horizontally partitioned data, having common feature events is an unreasonable assumption as different hospitals may have different attributes because of different specialties. These attributes are often structured data such as International Classification of Diseases (ICD) code used for billing, or custom assigned code for prescriptions, lab tests, procedures, etc. Furthermore, different hospitals might have their own annotation systems for the same medical events because of the lack of a consistent and unambiguous terminology system. Similarly, for vertically partitioned data, having common patient records in different institutions is somewhat another unreasonable assumption as we might not expect all patients to be accurately linked together for the hospitals they visit. There is a need to develop a new model that is more realistic.

Recently, there has been considerable attention in the application of neural networks to represent medical concepts as multidimensional and continuous vectors [19,20]. A process called contextual embedding, commonly used in natural language processing, maps each word from a corpus of text to a hyperdimensional space where similar words in terms of meaning or distributed usage would be located nearby (eg, short cosine distance). In the realm of health care, given a corpus of patients' history in a structured form, where medical events such as diagnoses, prescriptions, and lab tests are ordered chronologically for each patient, contextual embedding can embed each of these medical events so that similar events are closer in the final acquired space. Unlike one-hot representation, which may not be able to make distinction between related concepts such as congestive heart failure and myocardial infarction, contextual embedding produces a closer distance for these two concepts than other unrelated concepts (ie, kidney failure). Models that utilize such representation have shown higher prediction performance than previous models that do not [19,20]. Furthermore, research into identifying named entity recognition [21], abbreviation expansion [22], predicting unplanned hospital readmission [23], and predicting disease risk that incorporates long- or short-term dependencies in the electronic health record (EHR) [24] are examples of areas that have improved results with the application of word embedding as the first step [25]. As more deep-learning models dive into the realm of clinical text, instead of just using structured data to make predictions, word embeddings is becoming the paramount prerequisite for these studies.

Existing contextual embedding models are often built upon EHR data from a single institution. Each of these separate sites may contain information that other sites lack. It would be ideal if a model was built on raw data aggregated from different hospital sites to compensate for the missing or sparse information each site may have, but because of privacy concerns and the current state of interoperability in health care, such aggregation is often infeasible. To address these problems, we propose that each hospital builds its own contextual embedding model, after which no patient-level information would remain in the acquired representations (ie, embeddings). Then, hospitals can share their own local models and subsequently, the wealth of information from their hospitals without violating patient privacy. Such aggregation of local models can serve as the proxy for a global model, combining information from a wide range of institutions and information sources.

As each model is trained separately and lies in different embedding spaces, it is difficult to analyze events from different hospitals together even though some events might be semantically related or even identical. In this paper, we propose a methodology that harmonizes different contextual embeddings into a global model. Code can be found in [26].

Methods

Temporal Clinical Pathway

For this paper, we will explore structured data such as lab tests, prescriptions, symptoms, conditions, and diagnoses. We will also explore unstructured data or clinical notes. For structured data, each code was given a prefix added to differentiate them according to their type: "l_" for lab tests, "c_" for conditions, "s_" for symptoms, "d_" for diagnoses, and "p_" for prescriptions. Each of these medical events in each patient's history was then put in chronological order to form his or her clinical pathway. An example is shown in Figure 1.

For unstructured data, Metamap [27] was used to first identify the medical concepts from free texts. Multiple words identified as a single medical concept were concatenated to form a single word. All words not mapped were omitted from the notes. No words were excluded with a cutoff score or selected for specific functionality. As long as Metamap was able to identify a word as a concept in the Metamap database, the word or words were kept. As done with the structured data, each of these medical concepts in each patient's history were then put in chronological order to form his or her clinical pathway.

Contextual Embedding

Given the clinical pathways of structured or unstructured data, we used contextual embedding to create continuous vectors for each medical event or concept, respectively. An example is shown for structured data in Figure 2 (The detailed information about the figure will be described in the following "Harmonization" section). For contextual embedding technique of this paper, we chose Word2Vec [28], which uses a neural network architecture to represent words of a large corpus as vectors. Unlike classical representation techniques such as one-hot representation, Word2Vec can effectively model words by considering the context in which the words are contained. Two architectures exist in this regard [28]—the continuous skip-gram model and the continuous bag-of-words (CBOW) model—depending on how the neural network is configured. Both architectures are essentially a three-layer network consisting of input, projection, and output layers. Providing the input as a sequence of a 1-of-M coding, where M is the vocabulary size, Word2Vec is capable of projecting them into a lower dimensional space while extracting their context. For this paper, we chose the skip-gram model for its accuracy. Experimental results for CBOW and GloVe (another embedding method) are shown in Multimedia Appendices 1-8, but skip-gram showed the best overall results. The model requires two parameters, size and window, defining the dimensionality of the final vector representation and maximum distance for contextual consideration, respectively [19,28].

There is one limitation to contextual embedding techniques such as Word2Vec and GloVe [29]. That is, because of random sampling in the training process, even repetitions on the same dataset result in embeddings of different orientations. This means that even if embeddings trained from the same dataset are pooled together naively, the medical events or concepts in one embedding would have an unreasonable relationship with events or concepts in the other embedding (eg, heart attack in one embedding would have the closest distance to elephantiasis in the other embedding).

Harmonization

Due to the limitation of contextual embeddings, two embeddings learned from two hospital sites would lie in different hyperdimensional spaces, which makes them difficult to be used together. Therefore, there is a need to harmonize them. This is regarded as a space alignment problem [30], and it can be solved by manifold learning with or without dimensionality reduction. Manifold learning can be classified into linear and nonlinear approaches based on the assumption of data structure. Here we adopt Procrustes [31], a linear method that composes of three affine operations (transformation, rotation, and scaling) for its simplicity and generalizability. The basic idea is very similar to automatic image alignment based on scale-invariant feature transform in computer vision, although in this case, we were dealing with high dimensional attributes in this contextual embedding harmonization.

Using Procrustes to fuse different vector models into one common space requires a list of corresponding pairs [30]. These are pairs of words that are the same events but may or may not be labeled differently in different institutions. With most hospitals using standardized terminology systems such as ICD, Ninth Revision (ICD-9) and ICD-10 for billing purposes, we can reasonably identify a list of codes referring to the same events in different hospital sites to serve as our "anchor pairs" for alignment. Using these common events, we derived an orthogonal matrix that transforms one contextual embedding into the space of another.

Figure 1. Example of a clinical pathway created from a patient's structured data.

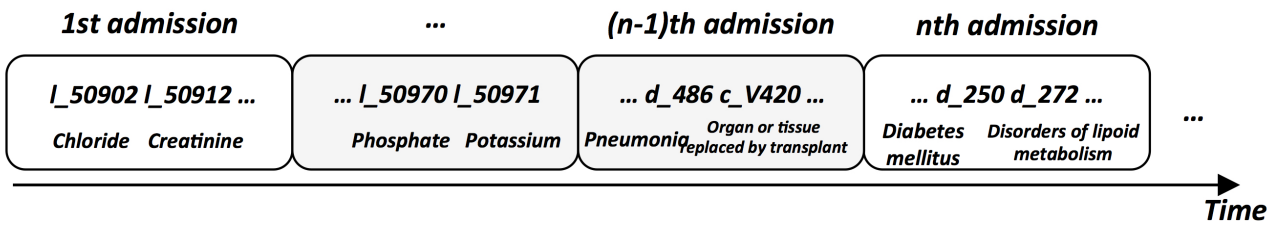
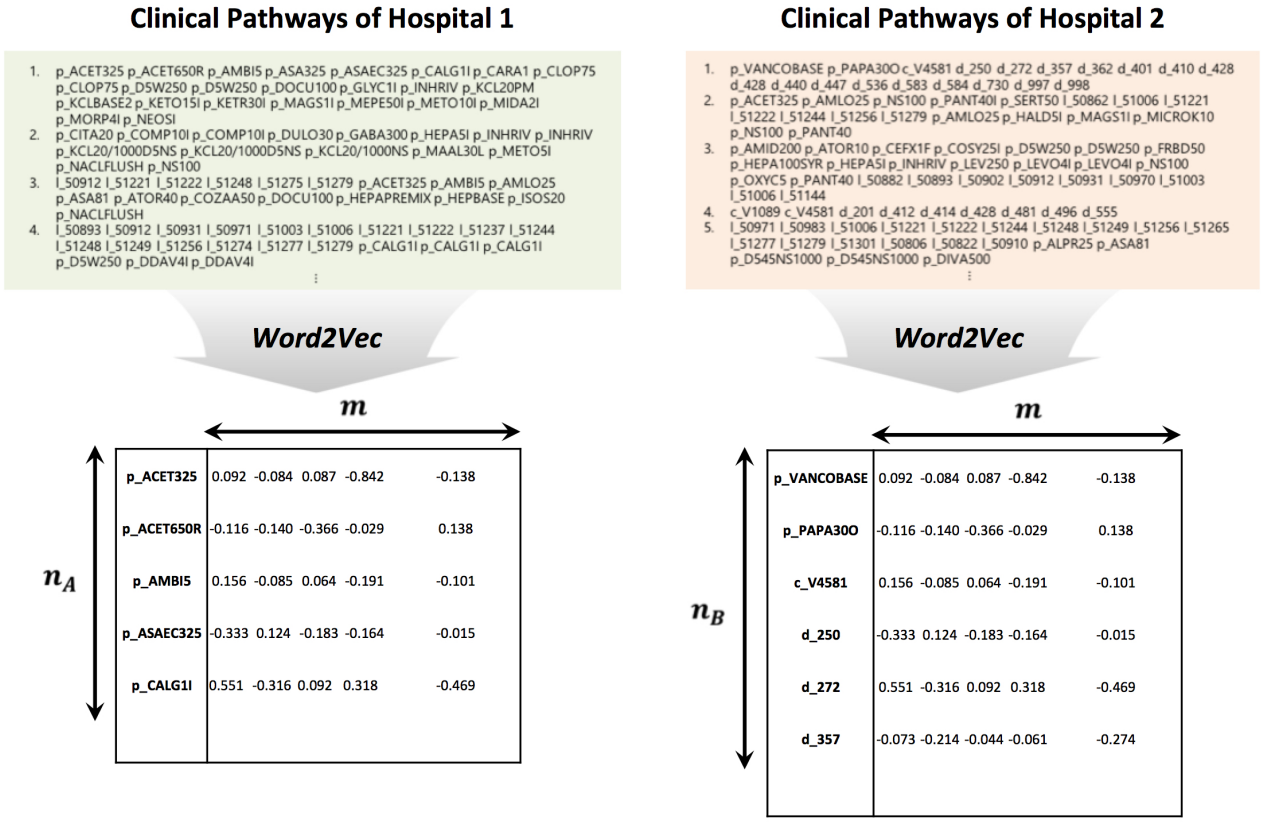


Figure 2. Example of two contextual embeddings created from two hospitals' structured data.



Taking the two contextual embeddings in Figure 2 as an example, the embeddings are shown in equation 1 where n_A and n_B are the number of contextual embeddings from two hospital sites, respectively, and m is the dimensionality of embeddings. Taking the corresponding anchor pairs X and Y, where X is a subset of A and Y is a subset of B, we can solve for orthogonal matrix Q and scalar k from the corresponding anchor pairs in equation 1. Applying Q and the scalar k, we can solve for $A^f$and B^f, which are the harmonized vector representation of A and B.

$$A \in R^{n_A \times m}$$

$$B \in R^{n_B \times m}$$

$$(1)\ Min_{Q,k} \| (X - 1_n \mu_X^T) - kQ(Y - 1_n \mu_Y^T) \|_F$$

$\mu_X^T$and μ_Y^Trepresent column-wise mean vectors of X and Y; n is the number of corresponding anchor pairs. Q is solved as shown in equation 2 using singular value decomposition:

$$(X - 1_n \mu_X^T)^T (Y - 1_n \mu_Y^T) = U\Sigma V^T$$

$$Q = UV^T$$

$$(2)\ k = trace(\Sigma) / trace(Y^T Y)$$

With the orthogonal matrix and scaling factor, Q and k, we can apply them onto one of the contextual embeddings to transform one into the space of the other as shown in equation 3:

$$A^f = A - 1_{nA} \mu_X^T$$

$$(3)\ B^f = kQ(B - 1_{nB} \mu_Y^T)$$

An example is illustrated in Figure 3. We can see that "hepatitis_c," "cirrhosis," "lungs," "myocardial," and "renal_failure" in the upper left and upper right part of Figure 3 are common in both local models. Using them as anchor pairs to derive the orthogonal matrix, we harmonized the two local models into a common one shown as the bottom part of Figure 3.

Patient Diagnosis Projection Similarity

To predict the next likely diagnosis of a new patient for structured data experiments, we used the patient-diagnosis projection similarity (PDPS) method [19]. To calculate PDPS, we first create a patient vector. In short, we normalize the summation of each vector representation of events in the clinical pathway of a patient, with each event vector multiplied by a time decay function (ie, $e^{-\lambda t}$with a time decay factor λ; see Figure 4). To calculate the probability of each diagnosis occurring as the next event, we calculate the cosine similarity between the patient vector and a diagnosis vector. The equation explaining this process is shown in equation 5, where the V_d is the contextual vector representation of diagnosis d i n the vector space, V_c is the vector contextual representation of a medical event in the clinical pathway of a patient, S, and thus equation 4 is the patient vector. The number of events from the last event of the clinical pathway is t_c.

$$(4)\ \Sigma_{c \in S} V_c e^{-\lambda tc}$$

$$(5)\ y(S, d) = cosine(V_d, \Sigma_{c \in S} V_c e^{-\lambda tc})$$

Figure 3. Example of Procrustes harmonization. Upper left: local embeddings of site 1. Upper right: local embeddings of site 2. Bottom: combined embeddings of two sites after harmonization.

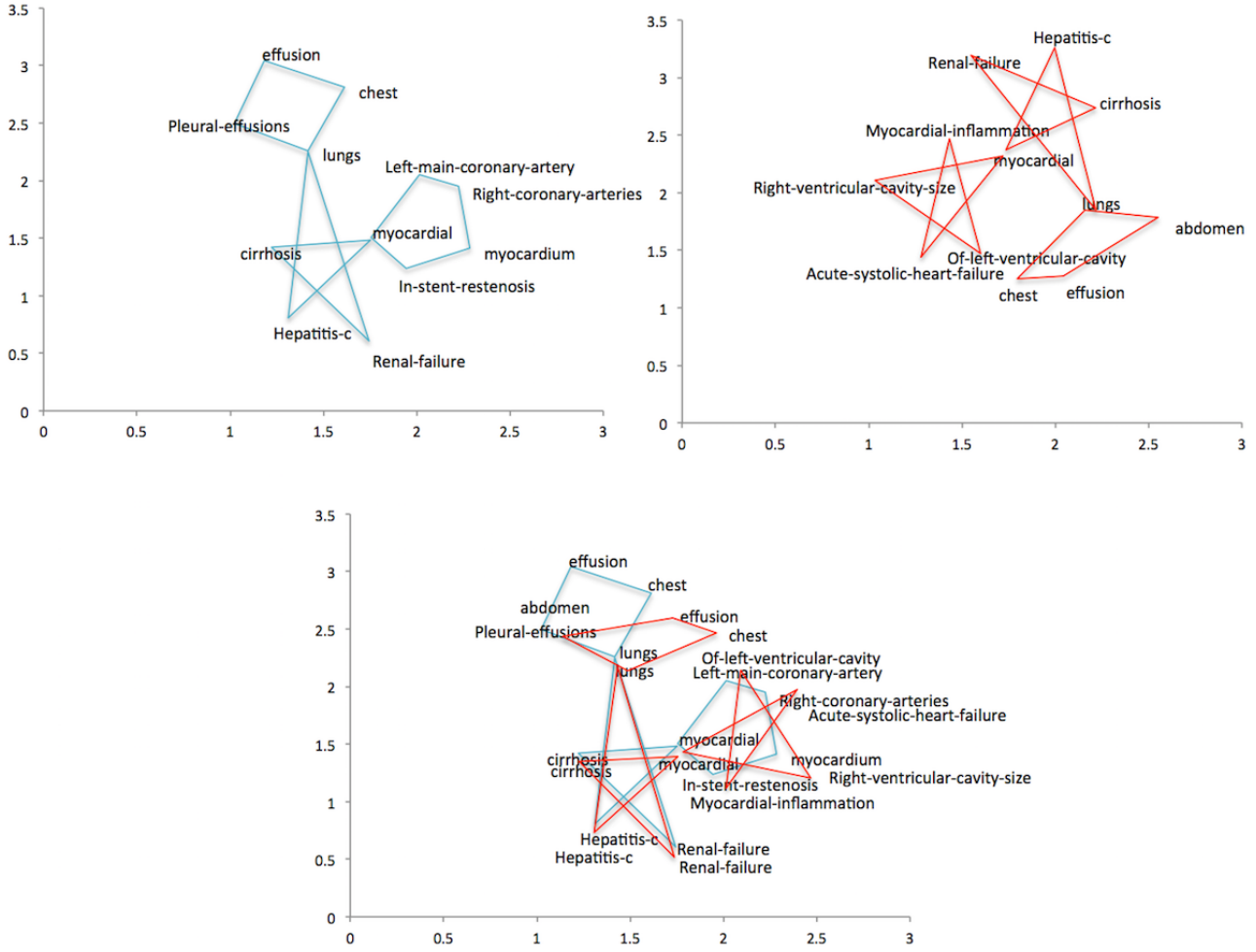
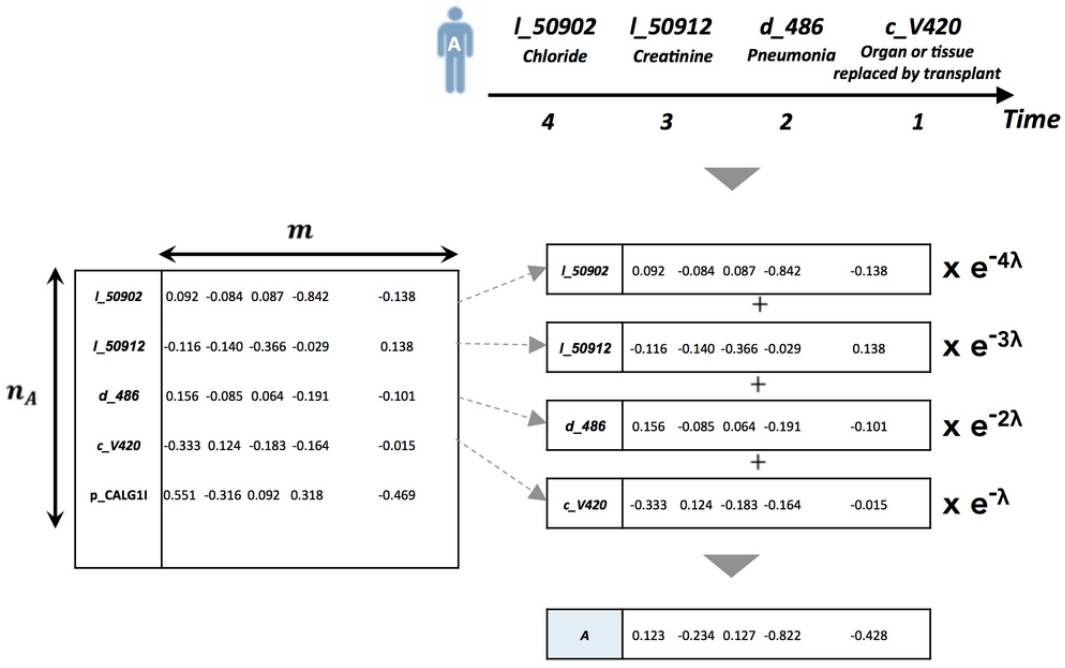


Figure 4. Example of creating a patient vector from event-level vector representation. The patient vector is a linear combination of the event vectors, weighted by a time decay function (e –λt with a time decay factor λ).



Results

Data Processing

The dataset we used was from MIMIC-III, a freely accessible critical care database [7]. We had two sets of experiments. One was conducted on structure data such as codes for diagnoses, prescriptions, lab tests, symptoms, and conditions. In MIMIC-III, these structured parts of the data include everything coded that were billable. ICD-9 code was used for symptoms, diagnoses, and conditions. Custom local codes were used for lab tests and prescriptions. Another experiment was conducted on unstructured data (ie, clinical notes). For structured data, ICD-9 codes for diagnoses in MIMIC-III were generalized to level 3. For example, a patient with "diabetes with ketoacidosis, type I (juvenile type) uncontrolled" (250.13) was generalized to diabetes mellitus (250) by reducing all ICD-9 code to three digits. Because our evaluation was based on the prediction accuracy, we excluded patients who only have one admission. We also excluded rare medical events that happened in less than 50 admissions for the structured data. In the end, we kept 5639 patient records for the experiment. From these records, we constructed the temporal clinical pathway for both structured and unstructured data. Ten-fold cross validation was implemented for all experiments, which randomly splits the dataset into ten folds with equal sizes, using nine folds for training and one fold for testing.

In each replicate, to simulate two different hospital sites, we divided the training patient records into two groups of patients randomly; we call these "local" sites. Experiments done on all training patients were used as a gold standard for comparison; we call this "global." From the training set, we created the "global" contextual embedding model using all patient records and the two "local" embedding models each using half of all patient records. The size and window parameters used to learn word embedding for structured data were 350 and 30, respectively. For unstructured data, the parameters were 350 and 100, respectively.

These two "local" embeddings were harmonized into a common model using Procrustes. As our two "local" hospital sites both came from splitting MIMIC-III [7], technically, the number of corresponding anchor pairs can account for almost the entirety of all the medical events. To create more realistic simulations, we tried different smaller fractions of all possible corresponding anchor pairs and changed the rest of the pairs artificially to be labeled differently so no events could be recognized by the other site except for the corresponding events. This was done to simulate the difference hospitals might have in their own terminology and the possibility that only a fraction of all their medical events codes are in common.

Structured Data Results

For structured data, the harmonization of the two "local" embeddings required common events to serve as corresponding anchor pairs. There were a total of approximately 2700 total unique events between the two sites, of which there were approximately 2500 common events. We used different percentages of all possible common events as corresponding pairs for different experimental scenarios, and the rest of the pairs were artificially labeled differently, where the word in the pair from one site was appended with suffix "m1," and the word from the other site was appended with "_m2."

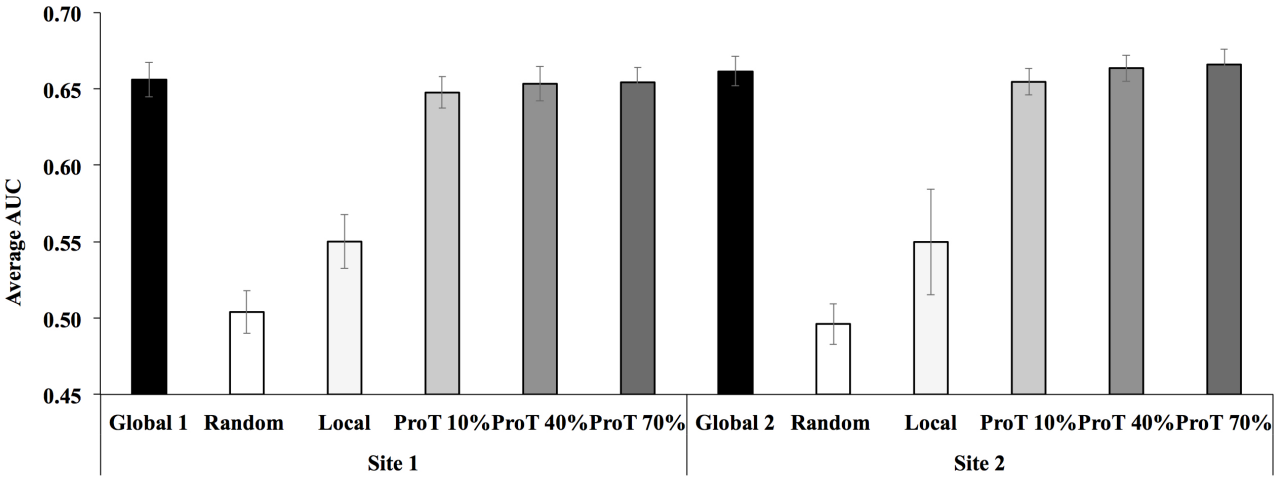
For all scenarios, we used PDPS to predict test patients' diagnoses in the final admission given all their records before the final admission. As an evaluation measure, we used the area under the receiver operating characteristic curve (AUC), for which 1 represents a perfect model, and 0.5 represents a worthless model. The average AUCs of a variety of diagnoses of different models were compared, and the benefit of harmonization is shown in the following three scenarios.

Incomplete Information

To evaluate the performance of the Procrustes harmonization, we first looked at how well missing medical events in one site can be compensated with the event vectors from another site. Oftentimes, small clinics or hospitals might not have encountered all medical events. In terms of PDPS prediction, missing events with no embeddings cannot be incorporated into the making of the patient vector, making prediction less accurate. Moreover, prediction simply cannot be made with PDPS for certain diagnoses if there are no embeddings for those diagnoses.

To resolve such a problem, we compensated the missing information using diagnosis vectors from another hospital. To test whether diagnosis vectors from another hospital can be used accurately to predict in another site, we first took the 40 most common diagnoses in MIMIC-III and randomly separated them into two sets of 20 diagnoses. The two "local" sites both originally had all 40 diagnoses, but we took one "local" site and deleted all instances of one set of 20 diagnoses (Multimedia Appendix 9, colored blue) and took the other "local" site and deleted the other set of 20 diagnoses (Multimedia Appendix 9, colored red). We trained and contextually embedded these two raw datasets separately, making two embedding models where each was missing a different set of 20 diagnoses. Figure 5, *Global 1*, shows that the average AUC of the 20 diagnoses that were missing from site 1 predicted using the global embedding model; Figure 5, *Global 2*, shows that the average AUC of the 20 diagnoses that were missing from site 2 predicted using the global model. These two act as the baselines. Then we explored what vectors added to the "local" embeddings could compensate for the missing diagnoses vectors. When we simply added random vectors for the 20 missing diagnoses to the respective sites and predicted using the "local" models, the average AUC for those missing diagnoses for both sites was close to 0.5 (Figure 5, *Random*). If we compensated vectors for the missing diagnoses with vectors in the other site without Procrustes harmonizing the embeddings of two sites first, the average AUC only improved to approximately 0.55 (Figure 5, *Local*). However with harmonization, compensating the missing diagnoses vectors with vectors from the other site returned the AUC to the level of the global model (Figure 5 Procrustes Transformed, *ProT*). We also tested whether using different percentage of corresponding anchor pairs for Procrustes would alter the AUC. Figure 5 shows that increasing the percentage of corresponding pairs increases the AUC in a very negligible manner.

Figure 5. Average area under the receiver operating characteristic curve (AUC) of several different scenarios. Site1 had no vectors for a set of 20 diagnoses, whereas site 2 had no vectors for another set of 20 diagnoses. Using global model, we predicted on the 20 diagnoses missing from site 1 (Global 1) and the 20 diagnoses missing from site 2 (Global 2). The missing vectors were compensated with either random vectors (random), untransformed vectors from the other site (local), or Procrustes harmonized vectors from the other site that were harmonized using different percentage of corresponding pairs (ProT 10%, 40%, and 70%).



Split Patient History

In another scenario, it is conceivable that patients may go to different hospitals, leaving parts of their patient history in one hospital while other parts in other hospitals. One can simply predict future events of these patients using part of their clinical pathway at each site, but a more accurate prediction can be made from his or her entire clinical pathway. However, obtaining the entire clinical pathway is not easy. First, it might be time-consuming or even infeasible to release patient history across hospital sites. Second, even if the entire patient history is in one site, the events in a patient's clinical pathway may be coded differently from site to site, leading to some events being unrecognizable by a model built solely on one site and unusable for prediction. To solve these two problems, hospitals can first share their own contextual embeddings and combine them into a common space using Procrustes. Then, for all the patients who have history in multiple hospitals, each local clinical pathway can be made into a local patient vector, effectively rendering the history unidentifiable. Finally, every local patient vector can be summed and normalized to obtain an approximation of the global patient vector. Then prediction can be conducted using the approximated global patient vectors and diagnoses vectors in each "local" hospital. The following experiment shows that the initial harmonization step is required to obtain significant prediction results.

For this task, we divided the raw MIMIC-III training set into three "local" sites then trained three "local" embedding models. For medical events in each "local" site 1, 2, and 3, suffixes "_m1," "_m2," and "_m3" were added to the end, respectively, simulating that each "local" site used their own coding system. To simulate test patients who have records in three "local" sites, we divided the clinical pathway of each test patient into three sections, where each section was appended with suffix "_m1," "_m2," or "_m3" to designate which section belonged to which site. The average AUC of the 80 most common diagnoses from MIMIC-III (Multimedia Appendix 9) was evaluated based on PDPS for different models and is shown in Figure 6. The average AUCs calculated with "local" site 1, 2, and 3 embedding models (*Original of Local 1, 2, and 3*) dropped significantly compared with the global model (*Global*). This was because each "local" model could only use one-third of the information of each test patient for prediction. If we summed and normalized the three local test patient vectors together without harmonizing the embeddings first, the AUC did not improve (*Original of Combined 1, 2, and 3*). However, when the three contextual embeddings were harmonized with Procrustes (*ProT*) first, summing and normalizing the local test patient vector together improved the AUC closer to the AUC of prediction made by the global model (*ProT of Combined 1, 2, and 3*). Figure 6 shows three "combined" results because each local site had its own diagnosis vectors that PDPS and subsequent AUC were calculated based on. We tested different percentages of corresponding pairs of 10%, 40%, and 70% for Procrustes harmonization. Similar to the previous experiment, Figure 6 shows that increasing the percentage of corresponding pairs had positive but negligible effect.

Hospitals With Different Sizes

Another possible scenario is that hospitals have different sizes. One hospital might be much smaller than the other. The smaller hospital might not be able to predict diseases of new patients accurately based on its existing patient history because small hospitals have limited or skewed data. In this case, our alignment can help the small hospital overcome such a limitation by incorporating information from the larger hospital. To test this scenario, we split MIMIC-III raw data into two sites with imbalanced ratios of hospital sizes that varied from 80% and 20%, 90% and 10%, to 95% and 5%. We also used different ratios of corresponding anchor pairs from 40%, 70%, to 100% of all possible corresponding pairs. After harmonization, we introduced one more simple task called fusion to boost the prediction performance in the small hospital. If an event was included in the anchor pairs, we took the weighted average depending on the size of the hospital to combine the two event vectors into one vector. If the event was not included, we found the nearest neighborhood (ie, k=1 where k is the number of

nearest neighborhoods) from the other site and averaged itself and the nearest neighborhood. After harmonization and fusion, the average AUC of the most common 80 diagnoses for each combination was calculated. Figure 7 shows AUC results of large and small hospitals with or without harmonization and different ratios of corresponding pairs. A larger difference in hospital size results in a larger difference in AUC, meaning the more information there is, the more accurate the model is. Furthermore, Figure 7 shows that a small hospital can improve its prediction performance without compromising the integrity of the larger hospitals through harmonization and fusion with data from large hospitals, which is an amenable feature.

Unstructured Data Results

Our harmonization method can be extended to unstructured clinical notes. Using the same method as structured data, we built a clinical pathway for unstructured data, except we used medical concepts extracted from Metamap [27]. Then we built a global embedding model and two "local" embedding models. Finally, we harmonized the two "local" model with Procrustes. There were approximately 150,000 unique medical concepts extracted with Metamap, of which approximately 72,000 were common in both sites that could be used as anchor pairs. For anchor pairs, we used the top 10% most common occurring corresponding pairs. These were common concepts such as "admission," "alter," "recalls," etc. The most common occurring anchor pairs were chosen because they were the most likely to have a similar neighborhood and structure relative to other concepts in each "local" hyperdimensional space, giving us the most reliable transformation matrix Q for equation 1. For some of the experiments, we had to create patient vectors. We used the same method as the method for structured data, but we omitted the time decay factor (λ). The following experiments were conducted to demonstrate the benefit of harmonized local embedding models.

Concept Unique Identifier Group Distances

For unstructured data, the clinical pathways used to train Word2Vec model consist of Metamap concepts [27]. Each for these concepts belongs to a concept unique identifier (CUI) created by the Unified Medical Language System, and each CUI may contain many concepts. For example, in the top image of Figure 8, there are many words that belong to the CUI C0392747, which is about "changing." We calculated the average pairwise cosine similarity among all members within a CUI group using concept vectors from the global model and found the average similarity for all CUI groups to be approximately 0.42. For the local models, within each CUI group, some words were learned from site1 local model (designated with the suffix "_m1"), whereas some words were learned from site 2 local model (designated with the suffix "_m2"). For common words that appeared in both sites, we randomly assigned them with either vectors from site 1 "local" model or vectors from site 2 "local" model. Again, we calculated the average pairwise cosine similarity among all members within each CUI group. If site 1 and site 2 models were not harmonized with Procrustes, the average similarity dropped to approximately 0.23, but the average similarity returned to approximately 0.42 if the local models were harmonized. Furthermore, if we limit the words in both sites to only include top 10% of the most common words, the average similarity after harmonization further increase to approximately 0.62, whereas the similarity without harmonization remains approximately 0.24. In addition, experiment was repeated with two real datasets. We used the MIMIC-III dataset to build site 1 "local" model and i2b2 dataset [32-36] to build site 2 "local" model. The results showed similar pairwise cosine similarity as the artificial datasets results.

Figure 6. Average area under the receiver operating characteristic curve (AUC) of the 80 most common diagnoses. Local 1, 2, and 3 show local sites using local embedding models either unharmonized (Original) or harmonized (Procrustes Transformed, ProT) and using only the part of the clinical pathway of test patients in the respective hospital. Combined 1, 2, and 3 show all three local test patient vectors combined, where each local vector is made with locally learned event vectors either unharmonized (Original) or harmonized (ProT).

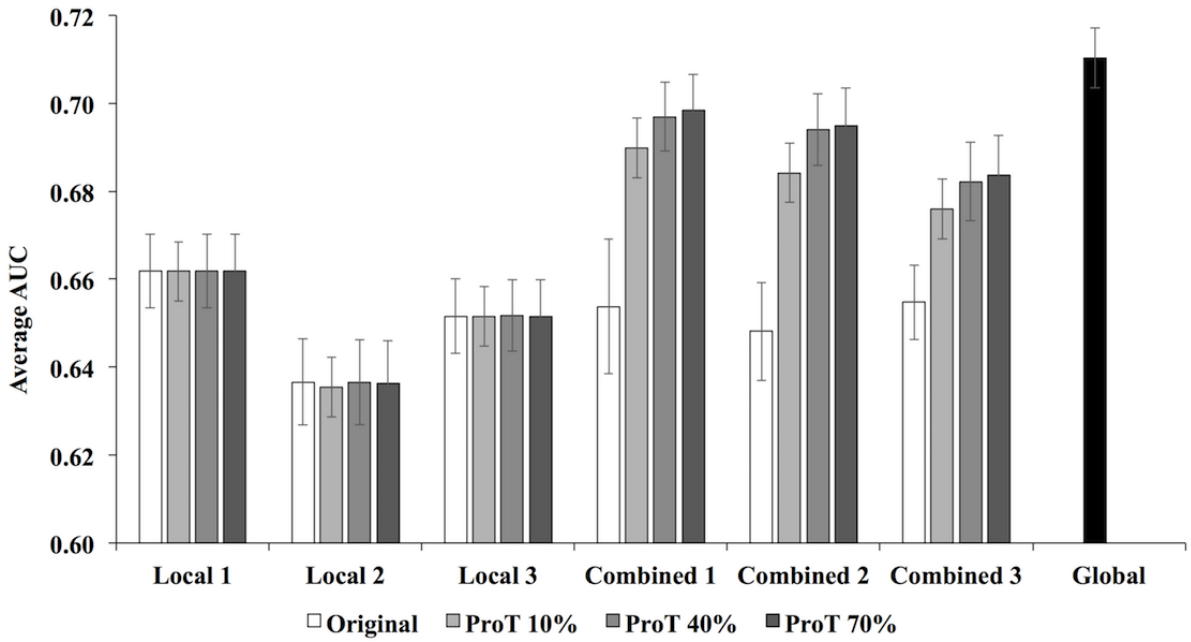


Figure 7. Average area under the receiver operating characteristic curve (AUC) of 80 most common diseases in Medical Information Mart for Intensive Care III (MIMIC-III) for hospital of different sizes harmonized using different percentage of corresponding pairs: upper left: 80% versus 20%, upper right: 90% versus 10%, bottom: 95% versus 5%.

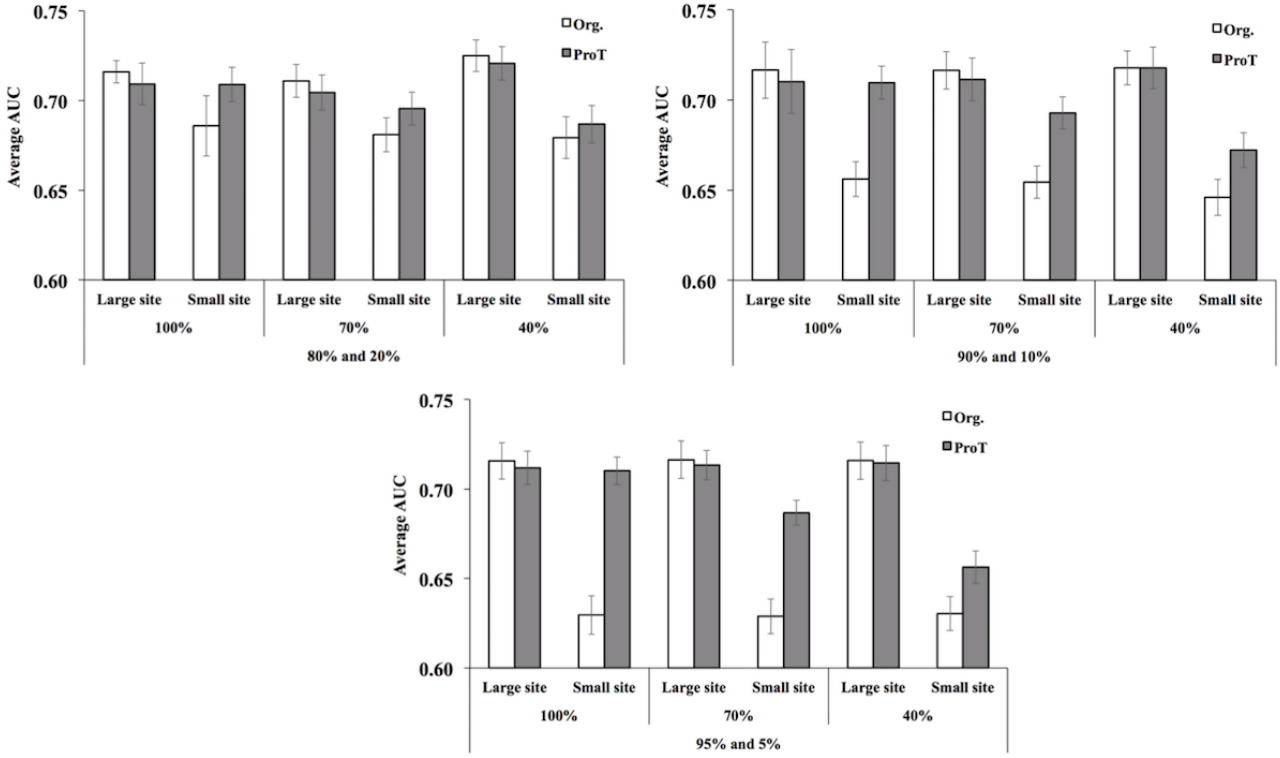


Figure 8. (a) Example of medical concepts belonging to a concept unique identifier (CUI) groups from a global model. (b) Red words were found from site 1 "local" model, magenta words were found from site 2 "local" model.

C0392747 (changing)
['modify', 'modified', 'modifying', 'modification', 'modifications', 'change', 'changes', 'changed', 'changing', 'alter', 'alters', 'altered', 'altering']
['modify_m1', 'modified_m1', 'modifications_m1', 'changed_m1', 'altered_m1', 'altering_m1', 'modifying_m2', 'modification_m2', 'change_m2', 'changing_m2', 'changes_m2', 'alter_m2', 'alters_m2']

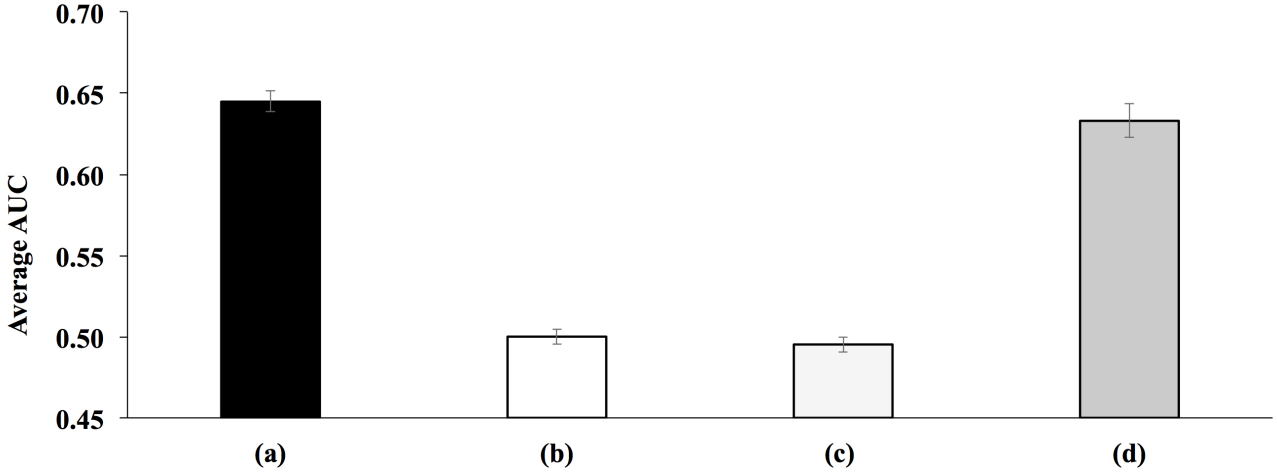
This demonstrates that words that are part of the same CUI group but learned from separate local sites can be combined and have distances restored to the global level with harmonization at a concept embedding level. Next, we will explore harmonization at the patient level.

Patient Similarity

For the following experiments, we attempted to explore patient similarity. We created patient vectors with the "global" embedding model and found pairs of patients who were the most similar in terms of the cosine similarity. We then separated these patients into two local sites and trained embedding models separately to evaluate most similar patients' retrieval from the other site. After making patient vectors for the two "local" sites, we took all patients from one site and calculated the rank of their most similar patients previously found using the global model, which was then in the other "local" site. Without Procrustes harmonizing the embedding models, the average rank of the most similar patient previously found in the global model dropped to approximately 1035.89 in the other site. However, with harmonization, the average rank for the most similar patient previously found rose back to approximately 1.35 in the other site. This demonstrates that it's difficult to find the most similar patient across site without harmonizing the embeddings of the sites first.

Even though we were able to restore the most similar patients that were found in the global model, we still needed to demonstrate that the most similar patients found after harmonization is relevant in terms of prediction. Therefore, we designed the following experiment to see if missing information can be compensated with information from other sites that is harmonized. To predict for future diagnosis in this experiment, PDPS was not used because it required diagnoses vectors, which, for the structured data, were ICD-9 diagnosis codes learned during contextual embedding. However, contextual embedding for unstructured data was done on clinical notes that did not contain any ICD-9 codes.

Figure 9. Average area under the receiver operating characteristic curve (AUC) over 80 most common diagnoses with each taking turn acting as the diagnosis of interest. Site 1 contained patients with diagnosis of interest, whereas site 2 did not. (a) Patient vectors of new test patients were created with embeddings 1, and training patients vectors of site 1 were used to find most similar patients for new test patients and to predict the diagnosis of interest. (b) Patient vectors of new test patients were created with embeddings 2, and training patients vectors of site 2 were used. (c) Patient vectors of new test patients were created with embeddings 2, and training patients vectors of site 1 were used. (d) Embeddings of site 1 and site 2 were harmonized before creating patient vectors. Then, patient vectors of new test patients were created with embeddings 2, and training patients vectors of site 1 were used to find most similar patients for new test patients.



Instead, we used the most similar training patients to each test patient to predict test a patient's future diagnosis. To find the most similar patients for each test patient, we calculated the average cosine similarity between a test patient to the training patients. Then, we used the training patients whose cosine similarities were one SD above the average as the most similar patients. Finally, prediction and subsequent AUC were calculated by probabilities from voting using the most similar training patients and their true ICD-9 diagnoses from the structured data as labels.

After learning contextual embeddings for the two "local" sites, we again created training patient vectors for the two sites. Then, in site 2, we deleted all patients with a certain diagnosis of interest but retained patients with this diagnosis in site 1. This experiment was done for the 80 most common diagnoses in MIMIC-III, with each taking turn acting as the diagnosis of interest. The average AUC over these 80 diagnoses is shown in Figure 9. Given a set of test patients, their patient vectors were created using either embeddings from site 1 or embeddings from site 2 depending on which hospital they were admitted to. The first column of Figure 9 shows the result where new test patients were admitted to site 1, and their patient vectors were created with site 1 embeddings. We predicted whether these new patients would develop the diagnosis of interest based on the most similar training patients in site 1 and obtained reasonable results. However, when new test patients were admitted to site 2, and we used the most similar training patients in site 2 to predict the probability of developing the diagnosis of interest, the result was no better than guessing as shown in the second column of Figure 9, because site 2 did not contain patients with the diagnosis of interest. Similarly, when new test patients were admitted to site 2, and we used the most similar patients found from site 1 to form a prediction, the result was no better than guessing as shown in the third column of Figure 9 because the embedding space was not harmonized, and not enough relevant similar patients were found. Finally, we harmonized the "local" embeddings between site 1 and site 2 first and created training patients vectors from them. When new test patients were admitted to site 2 and created their patient vectors with the harmonized "local" embedding of site 2, we could then use site 1 to find reasonable most similar patients and obtain significant AUC as shown in the fourth column of Figure 9. This shows that if new patients were admitted to a hospital and found a lack of relevant most similar patients to reasonably make accurate predictions on a diagnosis of interest, patient record from another hospital could compensate and provide relevant most similar patients. However, such compensation could only be achieved if contextual embeddings were harmonized between the hospitals.

Discussion

Principal Findings

This paper serves as a proof-of-concept that contextual embedding models, which are becoming bedrocks to deep learning analysis in place of one-hot representations, can be harmonized and subsequently synchronize information from different hospital sites for better prediction capability without sacrificing privacy. However, one limitation of our work is that all experiments were conducted on a single MIMIC-III database. The underlying structure of the simulated local models may be similar, making it easy to approximate the global model from combining harmonized models. However, we argue that every hospital will have similar structures and relationships for medical events or concepts related to diagnoses that are common and widespread. Using events related to these common diagnoses, we can nevertheless derive a reasonable transformation matrix to apply to the rest of the data even if we extend the method beyond the MIMIC-III database. Ultimately, harmonization can bring knowledge specialized in each hospital into the same space. This is a major benefit because once embeddings are created for each medical event or concept, it is difficult to add new event or concepts in relation to the existing embeddings without training the model again. With

harmonization, we can leverage embeddings learned from another source and add new vectors. At the moment, what this process fails to address are instances when two hospitals have conflicting embeddings regarding an event or concept. This method does not alleviate the issue but simply leaves both embeddings. The fusion method mentioned in the experiment conducted on hospitals with different sizes somewhat explored the issue. However, to truly create a global model where these two embeddings harmonize into one, further work is required. The current method works best to incorporate new information that hospitals are missing, whether it is missing diagnoses or parts of patient's clinical pathway.

The portability of event and patient vectors is another major strength of the harmonization method. With event and patient vectors rendered to vectors of numbers, privacy is preserved, yet information is still conveyed to hospitals involved in the harmonization. Instead of preserving privacy through deidentification and encryption, we take more of a machine-learning approach, where we tackle privacy protection and the sharing of data simultaneously. Currently, the way patient vectors are created is relatively naive, especially for unstructured data. However, the explosion of deep neural networks, such as recurrent and convolutional neural networks, can create more sophisticated patient vectors. What we have shown is the need of harmonization to analyze patients vectors learned at different sites together. It would be interesting to see if harmonization can be applied to deep learning in a distributed manner.

Finally, we have shown in a limited way the extension of the harmonization method onto three sites, but we also see that even with harmonization, the prediction results do not reach the global level. Future work can explore the possibility of harmonizing more sites, where the number of corresponding pairs diminishes as the number of sites increases.

Conclusions

Contextual embedding models are extremely useful in health care modeling because of their representativeness and applicability to downstream machine-learning models. With patient privacy being a paramount concern, it is nontrivial to directly share medical records in both structured and unstructured form. The emergence of contextual embedding in health care allows for a new way to share models without sharing data. We proposed an innovative framework to combine locally trained embeddings into embeddings in a global sense. Utilizing our unique harmonization, more accurate analyses can be made with the accumulated knowledge acquired from local sources. Such a technique can allow for information unique to a certain hospital to become available to other sites, increasing the fluidity of information flow in health care. Our demonstration is on Word2Vec, but it is widely applicable to other contextual embedding models, including the most recent Med2Vec [20] and Graph2Vec [37].

Acknowledgments

This research was supported in part by the National Institute of Health under award number R01GM114612, R01GM118574, R00HG008175, U01EB023685, and NLM T15 grant NIH/NLM T15LM011271 to the University of California, San Diego. Finally, it was also supported in part by grant U01TR002062 and OHNLP U01.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CBOW replicate for Figure 5.

[JPG File, 367KB - medinform_v6i2e33_app1.jpg]

Multimedia Appendix 2

GloVe replicate for Figure 5.

[JPG File, 376KB - medinform_v6i2e33_app2.jpg]

Multimedia Appendix 3

CBOW replicate for Figure 6.

[JPG File, 399KB - medinform_v6i2e33_app3.jpg]

Multimedia Appendix 4

GloVe replicate for Figure 6.

[JPG File, 381KB - medinform_v6i2e33_app4.jpg]

Multimedia Appendix 5

CBOW replicate for Figure 7.

[JPG File, 344KB - medinform_v6i2e33_app5.jpg]

Multimedia Appendix 6

GloVe replicate for Figure 7.

[JPG File, 324KB - medinform_v6i2e33_app6.jpg]

Multimedia Appendix 7

CBOW replicate for Figure 9.

[JPG File, 195KB - medinform_v6i2e33_app7.jpg]

Multimedia Appendix 8

GloVe replicate for Figure 9.

[JPG File, 114KB - medinform_v6i2e33_app8.jpg]

Multimedia Appendix 9

List of 80 most common diagnoses used for prediction. Diagnoses that are colored blue are diagnoses that were deleted from Site 1 in Section 3.2.1, while diagnoses that are colored red are diagnoses that were deleted from Site 2.

[PDF File (Adobe PDF File), 38KB - medinform_v6i2e33_app9.pdf]

References

1. Mandl KD, Kohane IS. Federalist principles for healthcare data networks. Nat Biotechnol 2015 Apr;33(4):360-363 [FREE Full text] [doi: 10.1038/nbt.3180] [Medline: 25850061]
2. Kohane IS, Drazen JM, Campion EW. A glimpse of the next 100 years in medicine. N Engl J Med 2012 Dec 27;367(26):2538-2539. [doi: 10.1056/NEJMe1213371] [Medline: 23268669]
3. Patten IS, Rana S, Shahul S, Rowe GC, Jang C, Liu L, et al. Cardiac angiogenic imbalance leads to peripartum cardiomyopathy. Nature 2012 May 9;485(7398):333-338 [FREE Full text] [doi: 10.1038/nature11040] [Medline: 22596155]
4. Berndt SI, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet 2013 May;45(5):501-512 [FREE Full text] [doi: 10.1038/ng.2606] [Medline: 23563607]
5. Office for Civil Rights, HHS. Standards for privacy of individually identifiable health information. Final rule. Fed Regist 2002 Aug 14;67(157):53181-53273. [Medline: 12180470]
6. Douglass MM, Cliffford GD, Reisner A, Long WJ, Moody GB, Mark RG. De-identification algorithm for free-text nursing notes. 2005 Presented at: Computers in Cardiology; September 25-28, 2005; Lyon, France. [doi: 10.1109/CIC.2005.1588104]
7. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
8. Dernoncourt F, Lee JY, Uzuner O, Szolovits P. De-identification of patient notes with recurrent neural networks. J Am Med Inform Assoc 2017 May 1;24(3):596-606. [doi: 10.1093/jamia/ocw156] [Medline: 28040687]
9. Neamatullah I, Douglass MM, Lehman LW, Reisner A, Villarroel M, Long WJ, et al. Automated de-identification of free-text medical records. BMC Med Inform Decis Mak 2008 Jul 24;8:32 [FREE Full text] [doi: 10.1186/1472-6947-8-32] [Medline: 18652655]
10. South BR, Mowery D, Suo Y, Leng J, Ferrández Ó, Meystre SM, et al. Evaluating the effects of machine pre-annotation and an interactive annotation interface on manual de-identification of clinical text. J Biomed Inform 2014 Aug;50:162-172 [FREE Full text] [doi: 10.1016/j.jbi.2014.05.002] [Medline: 24859155]
11. 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]
12. Wu Y, Jiang X, Kim J, Ohno-Machado L. Grid Binary LOgistic REgression (GLORE): building shared models without sharing data. J Am Med Inform Assoc 2012;19(5):758-764 [FREE Full text] [doi: 10.1136/amiajnl-2012-000862] [Medline: 22511014]
13. Jiang W, Li P, Wang S, Wu Y, Xue M, Ohno-Machado L, et al. WebGLORE: a web service for Grid LOgistic REgression. Bioinformatics 2013 Dec 15;29(24):3238-3240 [FREE Full text] [doi: 10.1093/bioinformatics/btt559] [Medline: 24072732]

14. Yu H, Jiang X, Vaidya J. Privacy-preserving SVM using nonlinear kernels on horizontally partitioned data. 2006 Presented at: SAC '06 Proceedings of the 2006 ACM symposium on Applied computing; April 23-27, 2006; Dijon, France. [doi: 10.1145/1141277.1141415]
15. Que J, Jiang X, Ohno-Machado L. A collaborative framework for distributed privacy-preserving support vector machine learning. AMIA Annu Symp Proc 2012;2012:1350-1359 [FREE Full text] [Medline: 23304414]
16. Shaneck M, Kim Y, Kumar V. Privacy preserving nearest neighbor search. 2006 Presented at: Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06); 18-22 December 2006; Hong Kong, China. [doi: 10.1109/ICDMW.2006.133]
17. Lu C, Wang S, Ji Z, Wu Y, Xiong L, Jiang X, et al. WebDISCO: a web service for distributed cox model learning without patient-level data sharing. J Am Med Inform Assoc 2015 Nov;22(6):1212-1219 [FREE Full text] [doi: 10.1093/jamia/ocv083] [Medline: 26159465]
18. Kim Y, Sun J, Yu H, Jiang X. Federated tensor factorization for computational phenotyping. KDD 2017 Aug;2017:887-895 [FREE Full text] [doi: 10.1145/3097983.3098118] [Medline: 29071165]
19. Farhan W, Wang Z, Huang Y, Wang S, Wang F, Jiang X. A predictive model for medical events based on contextual embedding of temporal sequences. JMIR Med Inform 2016 Nov 25;4(4):e39 [FREE Full text] [doi: 10.2196/medinform.5977] [Medline: 27888170]
20. Choi E, Bahadori MT, Searles E, Coffey C, Thompson M, Bost J, et al. Multi-layer Representation Learning for Medical Concepts. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.: ACM Press; 2016 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16; August 13-17, 2016; San Francisco, California. [doi: 10.1145/2939672.2939823]
21. 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]
22. Liu Y, Ge T, Mathews KS, Ji H, McGuinness DL. Exploiting Task-Oriented Resources to Learn Word Embeddings for Clinical Abbreviation Expansion. In: Proceedings of the 2015 Workshop on Biomedical Natural Language Processing (BioNLP 2015). USA: Association for Computational Linguistics; 2015 Presented at: 2015 Workshop on Biomedical Natural Language Processing (BioNLP 2015); July 30, 2015; Beijing, China.
23. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. $mathtt {Deepr}$: a convolutional net for medical records. IEEE J Biomed Health Inform 2017 Dec;21(1):22-30. [doi: 10.1109/JBHI.2016.2633963] [Medline: 27913366]
24. Arxiv. Exploiting Convolutional Neural Network for Risk Prediction with Medical Feature Embedding URL: https://arxiv.org/pdf/1701.07474.pdf [accessed 2018-04-03] [WebCite Cache ID 6yPoARTc0]
25. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform 2017;1:1. [doi: 10.1109/JBHI.2017.2767063]
26. GitHub. URL: https://github.com/easonfg/harmonization
27. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21 [FREE Full text] [Medline: 11825149]
28. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. 2013. URL: http://llcao.net/cu-deeplearning15/presentation/Distributed_Representation_of_Words.pdf [WebCite Cache ID 6z9se0RPa]
29. Pennington J, Socher R, Manning CD. GloVe: Global Vectors for Word Representation. URL: https://nlp.stanford.edu/pubs/glove.pdf [WebCite Cache ID 6yPoc21eg]
30. Choo J, Bohn S, Nakamura GC, White AM, Park H. Heterogeneous Data Fusion via Space Alignment Using Nonmetric Multidimensional Scaling. 2012 Presented at: Proceedings of the 2012 SIAM International Conference on Data Mining; April 26-28, 2012; Anaheim, California. [doi: 10.1137/1.9781611972825.16]
31. Gower JC. Generalized procrustes analysis. Psychometrika 1975 Mar;40(1):33. [doi: 10.1007/BF02291478]
32. Stubbs A, Uzuner Ö. Annotating longitudinal clinical narratives for de-identification: the 2014 i2b2/UTHealth corpus. J Biomed Inform 2015 Dec;58 Suppl:S20-S29 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.020] [Medline: 26319540]
33. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: overview of 2014 i2b2/UTHealth shared task Track 1. J Biomed Inform 2015 Dec;58 Suppl:S11-S19 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.007] [Medline: 26225918]
34. Stubbs A, Uzuner Ö. Annotating risk factors for heart disease in clinical narratives for diabetic patients. J Biomed Inform 2015 Dec;58 Suppl:S78-S91 [FREE Full text] [doi: 10.1016/j.jbi.2015.05.009] [Medline: 26004790]
35. Stubbs A, Kotfila C, Xu H, Uzuner Ö. Identifying risk factors for heart disease over time: overview of 2014 i2b2/UTHealth shared task Track 2. J Biomed Inform 2015 Dec;58 Suppl:S67-S77 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.001] [Medline: 26210362]
36. Sun W, Rumshisky A, Uzuner O. Annotating temporal information in clinical narratives. J Biomed Inform 2013 Dec;46 Suppl:S5-12 [FREE Full text] [doi: 10.1016/j.jbi.2013.07.004] [Medline: 23872518]
37. Narayanan A, Chandramohan M, Venkatesan R, Chen L, Liu Y, Jaiswal S. MLG Workshop. 2017. graph2vec: Learning Distributed Representations of Graphs URL: http://www.mlgworkshop.org/2017/paper/MLG2017_paper_21.pdf [WebCite Cache ID 6z9smNJfu]

Abbreviations

AUC: area under the receiver operating characteristic curve
CBOW: continuous bag-of-words
CUI: concept unique identifier
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
ICD: International Classification of Diseases
MIMIC-III: Medical Information Mart for Intensive Care-III
ProT: Procrustes Transformed

Edited by G Eysenbach; submitted 20.11.17; peer-reviewed by R Zhang, J Bian, P Boisrond, Y Wang; comments to author 04.01.18; revised version received 01.03.18; accepted 09.03.18; published 16.05.18.

Please cite as:
Huang Y, Lee J, Wang S, Sun J, Liu H, Jiang X
Privacy-Preserving Predictive Modeling: Harmonization of Contextual Embeddings From Different Sources
JMIR Med Inform 2018;6(2):e33
URL: http://medinform.jmir.org/2018/2/e33/
doi:10.2196/medinform.9455
PMID:29769172

Original Paper

Predicting the Reasons of Customer Complaints: A First Step Toward Anticipating Quality Issues of In Vitro Diagnostics Assays with Machine Learning

Stephane Aris-Brosou[1], PhD; James Kim[2], MBA; Li Li[2], MD, PhD; Hui Liu[2], PhD

[1]Department of Biology, University Of Ottawa, Ottawa, ON, Canada

[2]Ortho Clinical Diagnostics, Raritan, NJ, United States

Corresponding Author:
Stephane Aris-Brosou, PhD
Department of Biology
University Of Ottawa
30 Marie Curie Pvt.
Ottawa, ON, K1N 6N5
Canada
Phone: 1 613 562 5800
Fax: 1 613 562 5486
Email: sarisbro@uottawa.ca

Abstract

Background: Vendors in the health care industry produce diagnostic systems that, through a secured connection, allow them to monitor performance almost in real time. However, challenges exist in analyzing and interpreting large volumes of noisy quality control (QC) data. As a result, some QC shifts may not be detected early enough by the vendor, but lead a customer to complain.

Objective: The aim of this study was to hypothesize that a more proactive response could be designed by utilizing the collected QC data more efficiently. Our aim is therefore to help prevent customer complaints by predicting them based on the QC data collected by in vitro diagnostic systems.

Methods: QC data from five select in vitro diagnostic assays were combined with the corresponding database of customer complaints over a period of 90 days. A subset of these data over the last 45 days was also analyzed to assess how the length of the training period affects predictions. We defined a set of features used to train two classifiers, one based on decision trees and the other based on adaptive boosting, and assessed model performance by cross-validation.

Results: The cross-validations showed classification error rates close to zero for some assays with adaptive boosting when predicting the potential cause of customer complaints. Performance was improved by shortening the training period when the volume of complaints increased. Denoising filters that reduced the number of categories to predict further improved performance, as their application simplified the prediction problem.

Conclusions: This novel approach to predicting customer complaints based on QC data may allow the diagnostic industry, the expected end user of our approach, to proactively identify potential product quality issues and fix these before receiving customer complaints. This represents a new step in the direction of using big data toward product quality improvement.

(JMIR Med Inform 2018;6(2):e34) doi:10.2196/medinform.9960

KEYWORDS

post market surveillance; QC chemistry results; complaint data; CART; adaptive boosting

Introduction

Connected and so-called smart meters and other tools have transformed virtually every industry by enabling new functions and capabilities such as continuous monitoring, control, optimization, and autonomy [1]. This is particularly true in the health care industry, which deployed analytical systems ranging from electronic health records (EHRs) to clinical decision support systems [2]. Connected systems also include in vitro diagnostic (IVD) analyzers, which work with different assays that measure a number of markers in patients' blood samples

such as sodium or potassium, as well as other biomarkers such as troponin—which altogether are called "assays." Being connected, their manufacturers can monitor the analyzers' output in real time through encrypted, two-way interactive connections. As such, manufacturers can potentially quickly detect issues and act promptly to resolve the problem.

However, the sheer amount of data generated by these connected systems is such that big data analytics are required [3]. For this, a number of platforms have been developed, going from statistical tools such as R [4], to dedicated business intelligence and data mining tools. These platforms can then generate queries, reports, and perform online analytics processing, as well as data mining [2]. These aggregated data can then be used to perform one of three kinds of analytics: (1) descriptive analytics that permit the visualization of the data; (2) predictive analytics that try and predict the future of a system from its past behavior; and (3) prescriptive analytics that make recommendations about the best way to resolve a particular issue [5]. However, different health analytics contexts may require different approaches, as in the case of quality control (QC) data logged by analyzers.

As QC data are routinely used to monitor the performance of IVD and identify signals that may indicate a performance change, a number of approaches have been developed. These range from panels of experts that submit monthly reports [6], to automated systems that resort to summary statistics computed over temporal windows [7-11]. Although simple linear models can be used to monitor these complex systems [12], machine-learning algorithms have already proved capable of generating highly accurate predictions [13,14]. However, past approaches mostly have explored simple tools such as decision trees and other standard classifiers [15] and have not (1) Explored more sophisticated algorithms such as adaptive boosting [16] and (2) In the context of noisy and moderately large dataset—that are, hence, not always amenable to deep learning as recently deployed in the context of EHRs [17]. One aspect that has rarely been integrated into the analysis of QC data is its relation with customer data: when a shift in performance of a test assay is identified, what is its impact on the user (customer)? Will this trigger a complaint about QC? If the complaint is specific, such as "QC high" or "accuracy low," can we learn something about the quality of the data from the combination of those specific complaints?

The objective here is therefore to integrate these two kinds of data, QC data and customer complaints, to be able to predict specific QC issues, while accounting for intrinsic issues pertaining to customer data. Indeed, customer complaint databases have at least three inherent limitations that need to be considered when designing a prediction tool. First, complaint databases may contain inaccurate, incomplete, untimely, or unverified information [18]. Second, incidence may be under [19,20] or overreported [21]. For instance, certain advertising or regulatory actions may result in increased reporting [22], which could ultimately result in an overwhelming important signal with noise. Third, despite the best efforts of complaint handling professionals, errors while curating complaints (eg, misclassification of complaints) occur [23]. However, it is possible that by focusing solely on errors directly related to QC, or even by binning particular errors into larger categories (eg, "QC high" and "QC shift high" in the same category), it might alleviate some of these reporting issues.

Here, based on a particular connected IVD analyzer, we show that integrating QC data with a database of customer complaints can be used to predict which type of issues customers complain about. We hypothesized that connected systems can be utilized more efficiently by utilizing the collected QC data more efficiently and more specifically by resorting to machine-learning algorithms. We show that it is possible to identify product issues more proactively, which makes it possible to act on these before they trigger a customer complaint. We further show that some filtering of the complaint data (denoising) improves the accuracy of issues prediction. This work represents a first step toward meeting the recent plan from the US Food and Drug Administration (FDA) to leverage on big data to improve device performance and health care [24].

Methods

Data Collection

e-Connectivity Data

Data were collected using the e-Connectivity application's chemistry results, manufactured by Ortho Clinical Diagnostics (Raritan, New Jersey). This feature allows the manufacturer to pull information remotely from equipment installed at customer sites, which are themselves distributed throughout the world. The data retrieved in this study were generated by Ortho Clinical Diagnostics' VITROS analyzers of the 5,1 FS series, the 5600, 4600, 3600, or ECi/ECiQ Systems, that all log the same kind of information through e-Connectivity. Only QC data were extracted to avoid complications linked to patients' data (identifiability, variability, etc).

The e-Connectivity data contain information relative to the assay, serial numbers reflecting its origin, the measured concentrations, as well as some information relative to the analyzer itself (see Table 1 and Multimedia Appendix 1 for a full description of the e-Connectivity variables). We focused on five assays, here recoded as "assay A" to "assay E." The data pulled ranged from March 16, 2016 00:00:20 EST to June 14, 2016 23:38:51 EST, a total of 90.98 days, and contained 824,885 QC logs across the five assays. To assess the effect of the training period, we constructed a second set of data limited to the last 45 days of this 90-day set.

Customer Data

The corresponding customer complaint data were obtained by querying the product complaint database of the same manufacturer for the same time window as the QC data. Customer data contained information with respect to the assay for which an issue is reported, the call area (error code), and other information related to the assay (see Table 1 and Multimedia Appendix 2 for a full description of the customer variables; Multimedia Appendix 3 list the call areas reported over the five assays employed here). These data contained a total of 7999 logs. Across the five assays tested here, a total of

99 call areas were found. The goal here is to predict these call areas from the QC data.

Records Matching

The only fields that are shared between QC and customer data are assay name, J numbers, and lot numbers (Table 1). As each analyzer has a unique J number, we used this shared information to match QC samples with customer data. Although this approach works in most cases, there are instances when the same customer processes multiple samples, potentially from multiple analyzers, but logs only one call. Thus, the data that will be used to train the predictive algorithms are, in essence, noisy.

Predictive Classifiers

Feature Definitions

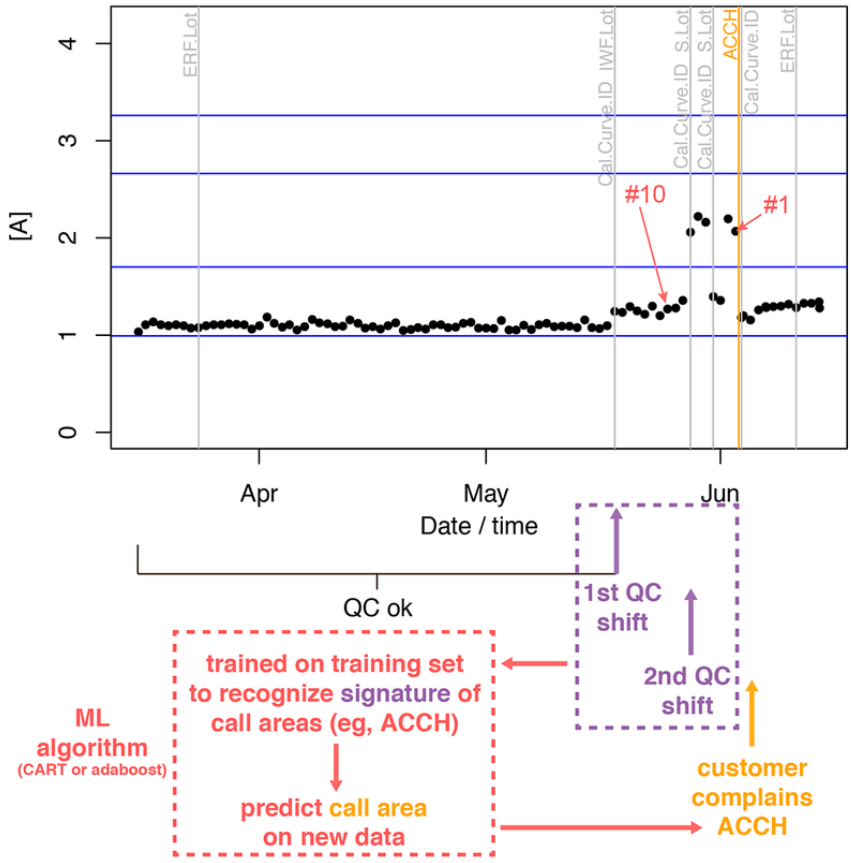
To find predictors of customer complaints based on QC data, we need to define operational variables, which are called features. These features were defined by inspecting a typical log of the system (Figure 1). From this, two types of features were defined, based on (1) concentration readings and (2) maintenance events (eg, change of calibration).

Table 1. List of the fields logged by e-Connectivity (that includes quality control, QC data) and by the customer complaint system. Corresponding abbreviations are shown.

Data and Abbreviation	Short description
e-Connectivity	
Assay	Abbreviation of assay name (recoded here)
J Number	Unique identifier assigned to each analyzer placed
F Concentration	Concentration of solute (assay); QC[a] result
Units	Unit of measured concentration (mmol/L)
F Concentration (SI)	Concentration of solute (assay); QC result
Units SI	Unit of measured concentration (SI)
Reagent Lot Number	Reagent lot number
S Gen	Manufacturing generation number
S Lot	Manufacturing lot number
ERF Lot	Electrolyte reference fluid lot
IWF Lot	Immuno-wash fluid lot number
Control Lot Number	Performance verifier lot number
Cal Curve ID	Calibration curve ID
Result ID	Unique identifier (encrypted) of QC result
Sample Name	Unique identifier (encrypted) of sample name
Time Metering	Time stamp of concentration log through e-Connectivity
Total Dilution	Dilution factor
Operator Dilution	Operator requested dilution
Body Fluid	Fluid type (serum, plasma, or urine)
Customer	
Create Audit Date	Time stamp of when complaint was placed
Call Subject	Same as assay in e-Connectivity
Call Area	Classification of concern or problem of the product or the analyzer-generated condition
Resolution	Term describing how the complaint was resolved
Complaint Number	Unique identifier of each complaint
Customer Number	Unique identifier of each customer
J Number	Analyzer serial number
Lot number	Reagent lot number
Region	Geographic region where complaint was placed
Call Status	Current call status of complaint (closed or open)
Problem description	Free-text field describing the complaint

[a]QC: quality control.

Figure 1. Feature definitions based on a typical sample logged in e-Connectivity. Assay concentrations (here for assay A) are plotted as a function of time. Horizontal blue lines show the modes of the density of sample means (our estimated verifiers). Vertical gray lines show timing of maintenance activities (change of calibration curves, etc). The orange vertical line shows when the customer placed a call—for "accuracy high" (ACCH; indicates the measured concentration is suspected of being higher than the actual value) in this example. The concentration reading just before this call ("#1") and 10 e-Connectivity logs before it ("#10") are indicated in red. Our machine-learning (ML) algorithms (in red) aim at learning the signatures (in purple) of call areas (orange) from a training set, to be able to identify those call areas, before a customer complains.



Concentration readings departing from expected values can be thought of as the prime trigger of customer complaints. Obviously, their absolute value with no other context has no predictive value (as long as it is not outside of the biological range) for QC data, and therefore, we should focus on departure from verifiers, which are known concentration readings produced during manufacturing. However, these verifiers are not logged by e-Connectivity and are only available as PDF files, which cannot be easily parsed. Customers may also choose to use QC material that is manufactured by a third-party, which further complicates the retrieval of verifier information. As a workaround, we calculated mean concentrations by samples, estimated the density kernel of these sample means, and determined the location of all the modes (Multimedia Appendix 4). We assumed that each mode corresponds to a verifier: the closest mode to each QC reading logged in e-Connectivity was assumed to be the verifier concentration. We then used the relative differences between concentration readings and estimated verifier. On the basis of this, we defined four features according to concentration readings just before customer call (log "#1" in Figure 1) and of the average logs two, five, and ten time points before the call (orange vertical bar in Figure 1). Because variability of QC logs may also signal issues, we defined three more features by SDs of the two, five, and ten concentration logs before customer call.

Customers may notice suboptimal performance of a machine and decide to try and resolve the issue on their own and place a call for assistance only if they cannot resolve the issue. We therefore defined features based on different maintenance events logged by the system (six in total): change of S Gen, S Lot, ERF Lot, IWF Lot, Control Lot Number, and Cal Curve ID. We considered both the timing of the last event before the call and the number of such events before the call. This led us to define 12 additional features based on maintenance events, for a grand total of 19 features (Table 2). A twentieth feature was defined as the time it takes for a customer to call since the last e-Connectivity reading (at "#1" in Figure 1).

Because the use of only "positive samples" (samples that led to a customer call) to train our algorithms would bias any prediction toward overpredicting calls, we also defined features for "negative samples." These are QC samples that did not generate any customer complaints. If *n* calls were logged for a given assay (among the 7999 logs in total), we drew *n* such negative samples. We calculated the features as above by drawing a cutoff time at random (from a uniform distribution limited by start and end time of QC logs for a given sample) that plays the role of a customer call in the positive samples. In this case, the call area (error code) is "OK"—giving then a total of 100 call areas that we want to predict.

Machine-Learning Algorithms

These 20 features were used as predictors during the training of machine-learning algorithms, whose goal was to classify (predict) the qualitative nature of problem represented by each call area. Two such algorithms were used here: a simple one, based on decision trees [25], and a more sophisticated one that recently proved very successful in one of our applications [26].

Decision trees represent one of the simplest type of classifier, with Classification and Regression Trees (CART) being one of the most basic algorithms. We employed the algorithm implemented in the tree library [27] version 1.0-37 in R version 3.2.3 [4]. Unlike CART, adaptive boosting relies on an iterated process that proposes boundaries in the space of predictors, each giving rise to a weak classifier; the final classifier then combines these different weak classifiers, emphasizing misclassifications, to create a final strong classifier [16]. The adabag library version 4.1 [28] was used. To avoid overfitting with both algorithms, each dataset (the 90- and the 45-day sets) was split into two subsets, where two-thirds of the data were used as a training set and the remaining one-third used to test performances (compute misclassification or error rate from the confusion table). Because of the many ways to split the data in a 2:1 ratio, we repeated this cross-validation exercise for 2500 random such 2:1 splits of the data, for both classifiers. Such a cross-validation experiment can also be seen as a means to prevent overfitting the data with a complex model.

Data Denoising

Over the 99 call areas employed so far, some are not directly related to QC, and those related to QC might share some characteristics. Both issues can create some noise, which can easily be filtered out of the data. We therefore created two filters, one that removes all non-QC related call areas (essentially, all error codes starting with a "Z" in Multimedia Appendix 3, as they are related to a misconfiguration of the analyzer) and one that bins some call areas. The first filter reduced the complaint data from 7999 to 572 logs and from 99 to 21 call areas by eliminating error codes unrelated to QC. The second filter, binning all QC high (QC high, QC Drift High, QC Shift High) together and all QC low (QC low, QC Drift Low, QC Shift Low) together, further reduced the number of call areas from 21 to 17. Applied both to the 90- and the 45-day data, these filtering steps led to four additional datasets. Our expectation was that these denoising steps would improve the performance of our classifiers, as reducing the number of categories from 99 to 17 simplifies the classification problem. The R code developed in this study is available from GitHub (sarisbro account); the QC data we used are proprietary, contain no patients records, but the variables used are listed in Table 1.

Table 2. List of the features used in the predictive modeling. Note that a "cutoff" represents the time when a customer calls in the case of "positive samples" (when there is an actual complaint), or the time drawn at random in the case of "negative samples" (see Methods).

Feature name	Definition
MostRecentConcentration	Assay concentration reading just before cutoff
TwoMostRecentConcentrationMean	Mean concentration for the two readings before cutoff
FiveMostRecentConcentrationMean	Mean concentration for the five readings before cutoff
TenMostRecentConcentrationMean	Mean concentration for the ten readings before cutoff
TwoMostRecentConcentrationSD	SD of concentration for the two readings before cutoff
FiveMostRecentConcentrationSD	SD of concentration for the five readings before cutoff
TenMostRecentConcentrationSD	SD of concentration for the ten readings before cutoff
NbPriorSGenChange	Number of S Gen changes before cutoff (since start of QC sample)
NbPriorSLotChange	Number of S Lot changes before cutoff
NbPriorERFLotChange	Number of ERF Lot changes before cutoff
NbPriorIWFLotChange	Number of IWF Lot changes before cutoff
NbPriorContLotNumChange	Number of Control Lot Number changes before cutoff
NbPriorCalCurveChange	Number of Calibration Curve changes before cutoff
TimeSinceLastSGenChange	Time elapsed since last S Gen change before cutoff
TimeSinceLastSLotChange	Time elapsed since last S Lot change before cutoff
TimeSinceLastERFLotChange	Time elapsed since last ERF Lot change before cutoff
TimeSinceLastIWFLotChange	Time elapsed since last IWF Lot change before cutoff
TimeSinceLastContLotNumChange	Time elapsed since last Control Lot Number change before cutoff
TimeSinceLastCalCurveChange	Time elapsed since last Calibration Curve change before cutoff
TimeToComplain	Time elapsed since last e-Connectivity log before cutoff

Results

Very Low Error Rates Even With Noisy Data

To predict which call areas are used when a customer complains only using QC data (Figure 1), we implemented two machine-learning algorithms that we ran on five different assays. As expected, CART showed error rates that were higher than those obtained with adaptive boosting, but both algorithms did much better than chance, with median error rates as small as 8% (Multimedia Appendix 5). Over the 90-day sample, each assay had triggered different numbers of complaints (assay A: 200, assay B: 835, assay C: 227, assay D: 182, assay E: 410, Multimedia Appendix 6), so that we expected that assays with larger number of complaints would have larger predictive power, but that was not the case (t_3=1.027, P=.38). Instead, the temporal dynamics of customer complaints, which increased in the second half of the 90-day period (Figure 2), affected error rates (Multimedia Appendix 5): in particular, the first quartile of the empirical cumulative distribution of customer complaints was a strong predictor of the error rate (adaptive boosting: t_3=4.103, P=.03). This suggests that it is easier to predict a call area (the type of a problem) for assays that quickly generate complaints.

Importance of Timing and Variability in Predicting Call Type

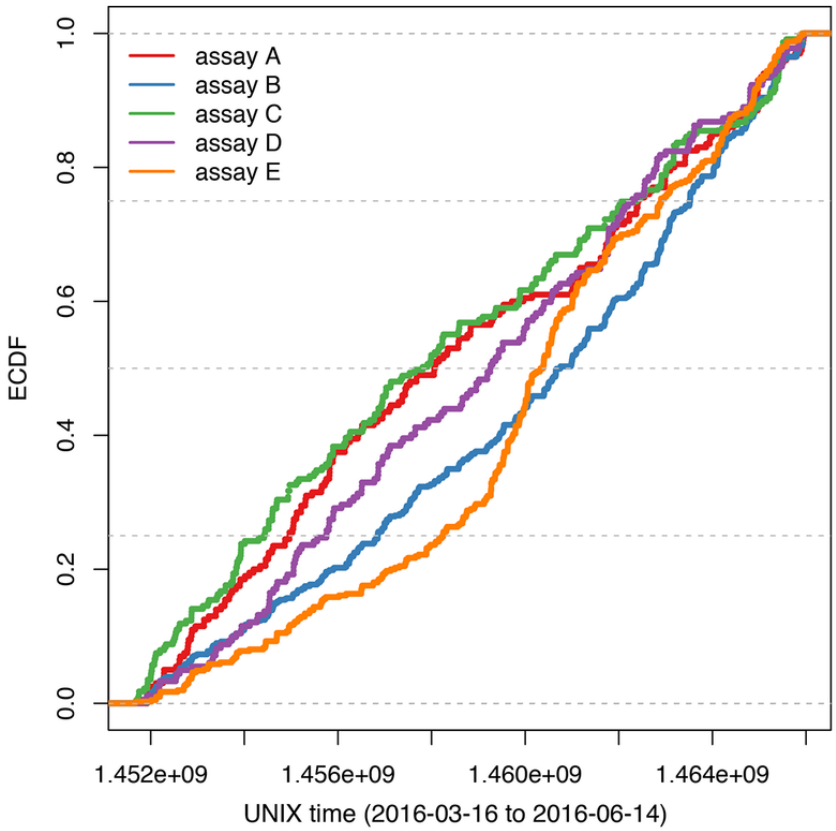
Adaptive boosting computes a measure of importance for each feature. Multimedia Appendix 6 shows that time to complain was the most important feature across all five assays tested. The second most important features were mostly those involving the timing of maintenance events, followed by the variability of concentrations (SDs) of the QC material. Unexpectedly, the actual concentration means (last two, five, or ten) were systematically the least important features for predicting call areas.

Misclassification Increases When Time to Complain Is Ignored

The previous results included time to complain as a feature; again, this is the time lag between the last QC reading by the system and the time when a customer placed a complaint (Figure 1). This is unrealistic, as in a real application, we would not know when a customer is going to complain. As a result, we assessed the impact of removing the time to complain feature from our classifiers. Both CART and adaptive boosting were affected by this removal, even if all five assays still had median error rates < 50% and as low as 20% with adaptive boosting (Multimedia Appendix 5). This increase in error rate after removal of this feature shows that time to complain is an important determinant of a complaint, which in turn suggests that customers are quick to complain after detecting a QC shift.

Note, however, that this removal of the most important feature did not affect the relative importance of the other features: those involved in the timing of maintenance events and those describing the variability of concentrations (SDs) were still the most important predictors (Multimedia Appendix 7).

Figure 2. Empirical cumulative distribution function (ECDF) of customer complaints. The ECDF was plotted for the five assays considered. The horizontal gray bars represent the first, second, and third quartiles. Each assay is color-coded as shown (inset).



Shorter Datasets Increase Accuracy

The results above suggest that the rate of complaint may affect performance. But it is unclear if longer training periods can benefit the performance of our algorithms. To test this, we subset the 90-day data to its last 45 days. When all the features were used to train the algorithms, all classification error rates decreased (Multimedia Appendix 5). A consistent pattern was observed when time to complain was also removed from the feature list (Multimedia Appendix 5). This suggests that the statistical process underlying call areas is nonstationary in time (ie, is time-heterogeneous). This hypothesis was supported by the change in error rate of assay E, which was the worst performer with the 90-day data but became one of the best one with the 45-day data, where a sharp increase in customer complaints can be observed at the beginning of this period (Figure 2, and Multimedia Appendix 6). It is therefore possible that training periods might have to be adjusted as calls are coming in: small number of complaints may require longer training periods, whereas an increase in complaint volume may necessitate reducing the training period in real time. On the other hand, it is also clear from Multimedia Appendix 6 that by shortening the training period, the number of call types was also reduced, so that the algorithms needed to predict fewer categories, which also contributed to lowering error rates. Therefore, shorter datasets may increase accuracy, but at the cost of being less general in the type of calls that can be predicted.

Data Denoising Increases Accuracy

In an attempt to denoise the customer data, we first removed non-QC related complaints and trained our classifiers on both the 90- and the 45-day datasets. This led to decreased error rates over all five assays (Multimedia Appendix 8), with some assays benefiting better than others (see assay D vs B) and to similar most important features (Multimedia Appendix 9). Note that for assay C, the small volume of complaints as observed in Figure 3 led to difficulties in training both classifiers on the 45-day data, and results for this assay at this shorter time frame are therefore absent. A closer examination of the confusions tables in this case suggests that no pattern exists in how errors are generated: some assays such as B can fail to predict almost 16% of accuracy high (indicating that the measured concentration is suspected of being higher than the actual value) call areas, whereas others such as E may have a bias in overpredicting QC high (Multimedia Appendix 10). Similarly, binning the QC-high or QC-low data on the QC-only complaints led to further improvements, leading in some cases to classifiers with a zero error rate (eg, see assay A in Figure 3).

In this case, where data are denoised by binning and by only considering QC-only data, the most important features for the classifier based on adaptive boosting remain TimeToComplain for both the 90- and the 45-day datasets (Figure 4). When this feature is removed, timing of events and variability of QC logs remain the most critical factors in determining call areas of customer complaints.

Figure 3. Distribution of prediction error rates for the binned quality check (QC)–only data. Error rates are shown as derived from the cross-validation analyses, where the data were split 2500 times (see Methods). Results are shown for both classifiers, Classification and Regression Trees (CART; broken lines) and adaptive boosting (solid lines), over the five assays considered for the 90-day data with all features (a) or with TimeToComplain removed (b) and likewise for the 45-day data with (c) or not (d) all features. Each assay is color-coded as shown.

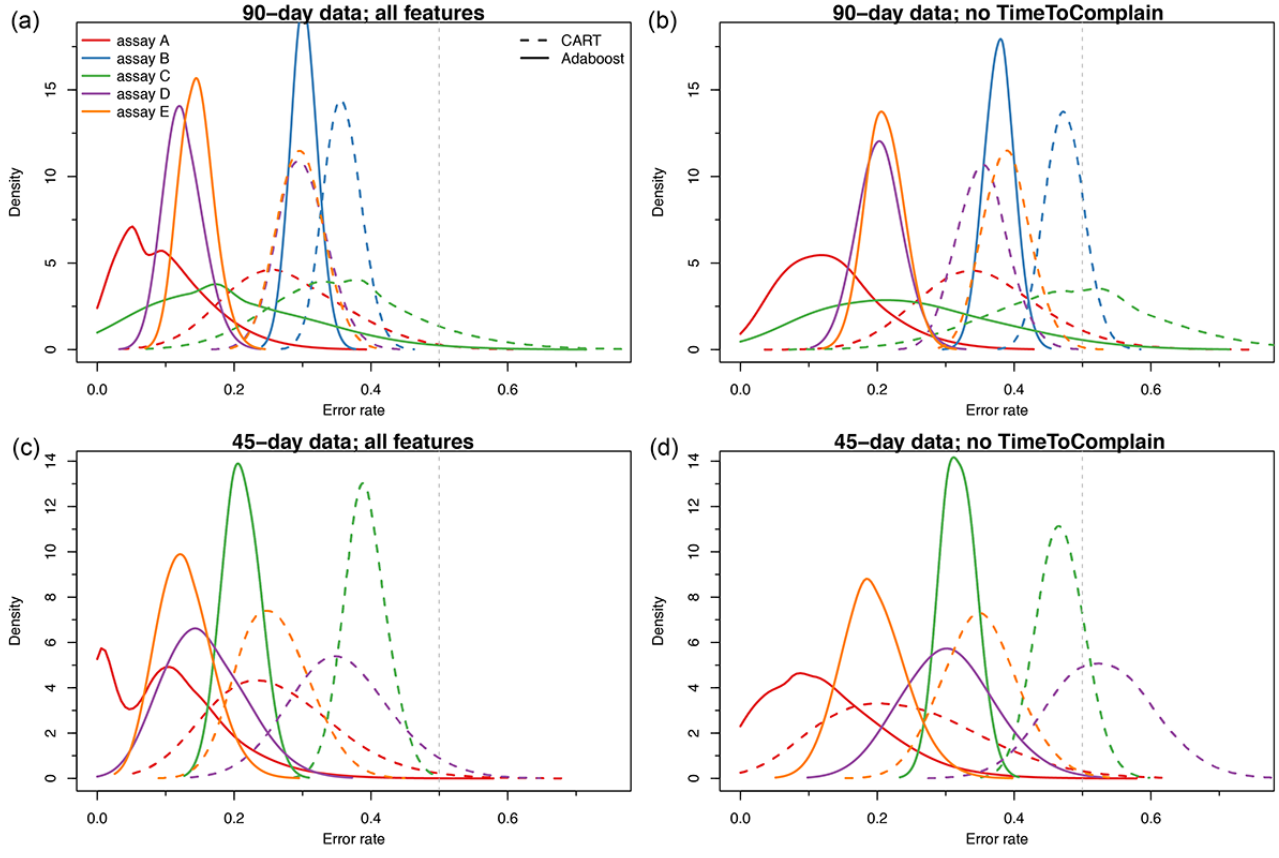
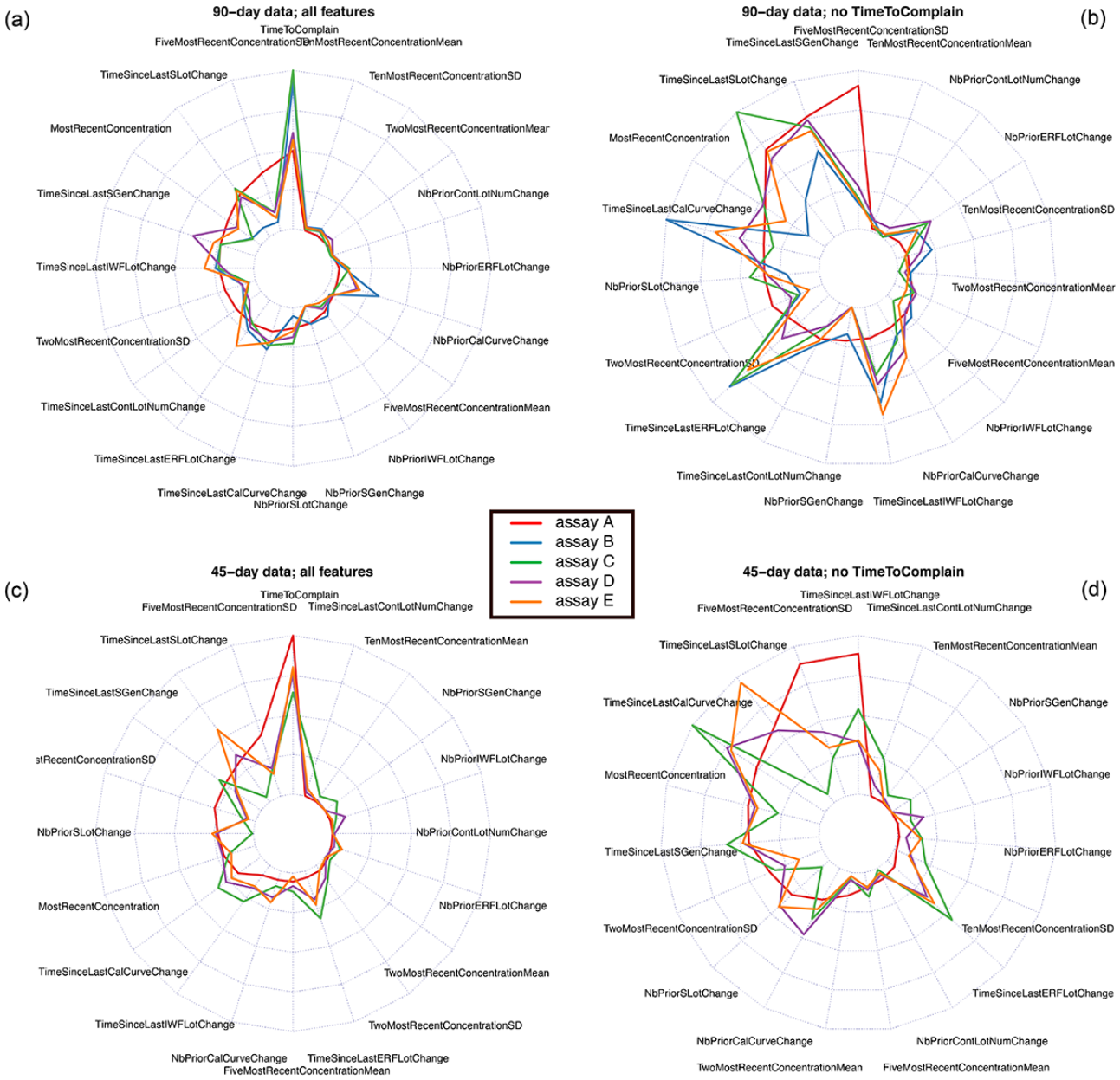


Figure 4. Feature importance under adaptive boosting for the binned quality control (QC)–only data. Importance of the features are shown as radar charts, over the five assays considered. Each assay is color-coded as shown. Top panels are for the whole 90-day datasets, whereas the bottom panels are for the 45-day datasets. Left panels include all feature; right panels exclude TimeToComplain from the models.



Discussion

Principal Findings

Traditionally, failure prediction in industrial applications aims at predicting *when* a particular system is likely to fail [29,30]. Here, we addressed a different question, one not directly related to the timing of failure, but one that focused on which type of failure can be predicted based on customer complaints (Figure 1). This is critical in the health care industry as it can point as to where along the process (from assay production to delivery, to storage, to use, to service on the analyzers) a product issue occurred, and hence, to take remedial steps to avoid further costs—and customer complaints. For this prediction of call areas, we compared two machine-learning algorithms, one based on decision trees (CART) and a more sophisticated one, adaptive boosting, that combines weak classifiers to produce a strong one. We showed that median errors rates can be as low as 8%—while still being as low as 20% in more realistic settings, where it is unknown when a customer is going to complain—and very close to 0% after denoising of the customer data (Figure 2). Note that not knowing when a customer complains does not seem to affect performance order on the five assays tested here.

One of our challenges here is that a complaint is a symptom of an actual product issue. When an issue occurs, the customer may complain, or not. The customer may wait to have several incidences of same issue before complaining, or may choose not to complain because he or she is busy or stopped complaining when it is a recurrent problem. It is also possible that a customer complains when there is no product-related issue. As a result, the complaint database that we used is intrinsically noisy, but (1) This database represents the best data available and (2) The manufacturer's goal is to improve

customer satisfaction by being able to identify issues before (or even without) a complaint call is placed.

To achieve this goal, we resorted to machine learning. As in any machine-learning application—except maybe with some deep-learning applications as those trained directly on images at the pixel level [31]—a key element is the identification and definition of the features used to train a classifier [32,33]. Rather than selecting features in an *ad-hoc* manner, as is often the case with EHRs [34], we took inspiration from standard recordings logged by a connected system to identify features that can easily be extracted from the data and that are also likely to reflect QC shifts (Figure 1). This led us to identify two kinds of features: those based on concentration readings and those based on the timing of maintenance events. In the context of this particular manufacturer in the health care industry (Ortho Clinical Diagnostics), we showed that timing of events represent the most important features in predicting a call type in a realistic setting (were the time when a customer complains is unknown), followed by variability in concentration readings (Figure 4). A future improvement of our approach could attempt to perform unsupervised feature learning, as done in deep learning [17]. This might circumvent the following difficulties: different data types (eg, patient health instead of QC data), equipment (eg, Bio-Rad [Hercules, California] rather than Ortho Clinical Diagnostics), or logging system (eg, Bio-Rad's UnityConnect vs e-Connectivity), which might require the definition of alternative features. However, it is likely that (1) All connectivity solutions log similar chemistry end points (concentrations, timing of service, etc) and that (2) Sophisticated machine-learning algorithms such as adaptive boosting will still produce quite impressive results. Here, we did not evaluate other algorithms such as support vector machines [14], neural networks [13] or deep learning [17], or others, as most of these approaches have the same goals and can behave equally well [35].

Limitations

Some additional questions and limitations remain, however. First, we extracted data for a period of 90 days and showed that the length of this period could affect performances. Indeed, shorter training periods seem to improve prediction performances when complaint rate is high. If complaint volume does affect performance, the length of the period used for analytics should be optimized in real time. This point was not addressed here and will require further investigation, in particular, to better understand the link between the volume of customer complaints for specific call areas, the features that become the most important, and how prediction performances are affected (Figure 4). Second, we only focused on five assays and showed that our general approach seems to deliver similar performances across those particular assays. However, this need not be the case, and a systematic survey should be undertaken. Third, we employed only one particular system here, the VITROS System, manufactured by Ortho Clinical Diagnostics. However, it is not immediately clear whether our approach can be ported to other systems, be they distributed by the same or by other manufacturers. Yet, it may be expected that most systems from most manufacturers will log QC data in a similar way, which can be interpreted in the same way as here (see feature definitions). Fourth, we were limited in our analysis of QC data by not having access to actual verifiers from the manufacturer. This forced us to resort to changes in the measured concentrations, rather than simply checking departures between measurements and verifiers. However, obtaining these data was in our case challenging, as these data were only available as individual PDF files for each performance verifier lot (there were hundreds of lots). Obtaining information about these verifiers would help train our predictive tools. Fifth, we exclusively focused on QC data used for monitoring health care systems, not on patients' health. This was done to avoid complications linked to obtaining consent forms from patients in hospitals that are themselves scattered around the world. Eventually, health care analytics should also monitor individual patients and hence, help physicians in their diagnosis. Sixth, call areas, which we aimed at predicting, are used by a manufacturer to identify an issue with the product or with the analyzer in the complaint handling process: they are not the root cause of the issue, which can only be determined through what is known as a root cause investigation (RCI). RCIs are, however, very time consuming to conduct, especially on analyzers that are distributed globally, so that most of the time, the actual cause of a reported issue may not be known. However, knowing which issue may arise (ie, our prediction of call areas) instead of the actual cause can help manufacturers to initiate targeted RCIs more proactively. Finally, we have only presented one side of the health care analytics in predicting call areas, not *when* failures occur. An integrated solution should put together both questions, possibly by merging our approach with traditional time series methods [29,30].

In the future, a more agnostic approach with respect to feature definition may be required: indeed, the features that are based on concentration readings all depend, to some extent, on the exact time when a customer complained. This time is unknown when performing real-time analytics. To circumvent this limitation, it might be better to implement a sliding window, defined over a time period $T=[t_0, t_1]$, and use time $t=t_1$ as the cutoff point to define features that are based on the timing of events.

Conclusions

Although the approach we described will require further validation and testing, the ultimate goal is to implement this kind of predictive tool into the global monitoring system of IVD analyzers to help manufacturers be more proactive in detecting quality issues of the various assays they marketed around the world. This may help them pinpoint where in the manufacturing process issues are likely to originate—eg, if only a particular lot number is globally generating the same call area, a manufacturing problem specific to this lot can be identified. As such, we might one day be able to develop *automated analytics* or systems that can not only identify when and how failures will occur but also automatically take remediation steps to resolve these issues, in real time, without the intervention of any human being [5].

In the meantime, the US FDA is planning to use big data to guide regulatory decisions [24]. Consequently, medical companies will soon have to harness all the data logged by their

instruments and use these data to their full potential to further improve the health care system. Our contribution here is a first step in this direction, laying ground to predicting call areas, and hence, enabling manufacturers, the expected end user of our approach, to be more proactive in postmarket surveillance. We predict that by combining our machine-learning approach with traditional time series analysis, we will eventually be able to predict when a customer will complain, in addition to what he or she will complain about. This work paves the way to developing an automated tool to anticipating customer complaints and identifying product quality issues through connected systems.

Acknowledgments

The authors would like to thank the Center for Advanced Computing and Compute Ontario for providing them with computing time; Jennifer Paine, Ian Wells, and the Safety Risk Management and Surveillance (SRMS) team for their support; Tom Balland and Craig Ritson for their help accessing the data; and Greg Munro, Jeanette Owejan, and Mike Torpey for discussions, as well as two anonymous reviewers for providing them with constructive comments. This work was supported by the Natural Sciences Research Council of Canada (SAB) and was part of Ortho Clinical Diagnostics SRMS project #2016-0825 (SAB, JK, LL, HL). This work was completed while SAB was being hosted by Yutaka Watanuki, at the University of Hokkaido in Hakodate, thanks to an Invitational Fellowship from the Japanese Society for the Promotion of Science.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed description of the e-Connectivity features.

[PDF File (Adobe PDF File), 48KB - medinform_v6i2e34_app1.pdf]

Multimedia Appendix 2

Detailed description of the customer features.

[PDF File (Adobe PDF File), 38KB - medinform_v6i2e34_app2.pdf]

Multimedia Appendix 3

Description of the 99 error codes reported by the analyzers over the five assays.

[PDF File (Adobe PDF File), 127KB - medinform_v6i2e34_app3.pdf]

Multimedia Appendix 4

Distribution of mean concentration reading per sample for the same assay. For each sample in the e-Connectivity data, the mean of all concentration readings was taken, and their distribution over the entire e-Connectivity 90-day data set was plotted. This distribution is multimodal; modes were estimated and are shown as vertical red dotted lines.

[PDF File (Adobe PDF File), 9KB - medinform_v6i2e34_app4.pdf]

Multimedia Appendix 5

Distribution of prediction error rates for the unfiltered customer data. Error rates are shown as derived from the cross-validation analyses, where the data were split 2500 times (see Methods). Results are shown for both classifiers, CART (broken lines) and adaptive boosting (solid lines), over the five assays considered, for the 90-day data with all features (a) or with TimeToComplain removed (b), and likewise for the 45-day data with (c) or not (d) all features. Each assay is color-coded as shown.

[PDF File (Adobe PDF File), 237KB - medinform_v6i2e34_app5.pdf]

Multimedia Appendix 6

Distribution of call areas for each assay. Distributions are shown for the whole 90-day data sets (a) and the 45-day data set (b). Each assay is color-coded as shown. Non-QC related call areas were filtered out.

[PDF File (Adobe PDF File), 163KB - medinform_v6i2e34_app6.pdf]

Multimedia Appendix 7

Feature importance under adaptive boosting for the unfiltered customer data. Importance of the features are shown as radar charts, over the five assays considered. Each assay is color-coded as shown. Top panels are for the whole 90-day data sets, while the bottom panels are for the 45-day data sets. Left panels include all feature, right panels exclude TimeToComplain from the models.

[PDF File (Adobe PDF File), 254KB - medinform_v6i2e34_app7.pdf]

Multimedia Appendix 8

Distribution of prediction error rates for the QC-only customer data. Error rates are shown as derived from the cross-validation analyses, where the data were split 2500 times (see Methods). Results are shown for both classifiers, CART (broken lines) and adaptive boosting (solid lines), over the five assays considered, for the 90-day data with all features (a) or with TimeToComplain removed (b), and likewise for the 45-day data with (c) or not (d) all features. Each assay is color-coded as shown.

[PDF File (Adobe PDF File), 233KB - medinform_v6i2e34_app8.pdf]

Multimedia Appendix 9

Feature importance under adaptive boosting for the QC-only customer data. Importance of the features are shown as radar charts, over the five assays considered. Each assay is color-coded as shown. Top panels are for the whole 90-day data sets, while the bottom panels are for the 45-day data sets. Left panels include all feature, right panels exclude TimeToComplain from the models.

[PDF File (Adobe PDF File), 248KB - medinform_v6i2e34_app9.pdf]

Multimedia Appendix 10

Examples of confusion tables obtained during cross-validation on the 90-day data, filtered for quality control (QC)–only call areas (data not binned by QC level). Numbers on the diagonal show accurate predictions; false predictions are below the diagonal, whereas missed predictions are above.

[PDF File (Adobe PDF File), 65KB - medinform_v6i2e34_app10.pdf]

References

1. Porter M, Heppelmann J. Harvard Business Review. 2014. How smart, connected products are transforming competition URL: https://hbr.org/2015/10/how-smart-connected-products-are-transforming-companies.pdf [WebCite Cache ID 6z8So28JG]
2. Raghupathi W, Raghupathi V. An overview of health analytics. J Health Med Informat 2013;04(03):2. [doi: 10.4172/2157-7420.1000132]
3. Alexandru A, Alexandru C, Coardos D, Tudora E. Healthcare, big data and cloud computing. URL: https://pdfs.semanticscholar.org/cdc8/deed19d2269db56e86d77f74ff7bd54b10c8.pdf [WebCite Cache ID 6z8SzRsWr]
4. R-project. 2015. R: A Language and Environment for Statistical Computing URL: https://www.R-project.org/ [accessed 2018-04-19] [WebCite Cache ID 6ymbONZ08]
5. Davenport TH. The Wall Street Journal. 2015. The rise of automated analytics URL: http://www.tomdavenport.com/wp-content/uploads/The-Rise-of-Automated-Analytics.pdf [WebCite Cache ID 6z8T4ZuW2]
6. Vasikaran S. Anatomy and history of an external quality assessment program for interpretative comments in clinical biochemistry. Clin Biochem 2015 May;48(7-8):467-471. [doi: 10.1016/j.clinbiochem.2014.12.014] [Medline: 25543064]
7. Wilson A, Roberts W, Pavlov I, Fontenot J, Jackson B. Patient result median monitoring for clinical laboratory quality control. Clin Chim Acta 2011 Jul 15;412(15-16):1441-1446. [doi: 10.1016/j.cca.2011.04.024] [Medline: 21549689]
8. Van Houcke S, Stepman H, Thienpont L, Fiers T, Stove V, Couck P, et al. Long-term stability of laboratory tests and practical implications for quality management. Clin Chem Lab Med 2013 Jun;51(6):1227-1231. [doi: 10.1515/cclm-2012-0820] [Medline: 23337057]
9. Fleming J, Katayev A. Changing the paradigm of laboratory quality control through implementation of real-time test results monitoring: for patients by patients. Clin Biochem 2015 May;48(7-8):508-513. [doi: 10.1016/j.clinbiochem.2014.12.016] [Medline: 25549976]
10. Goossens K, Van Uytfanghe K, Twomey PJ, Thienpont LM, Participating Laboratories. Monitoring laboratory data across manufacturers and laboratories--a prerequisite to make “Big Data” work. Clin Chim Acta 2015 May 20;445:12-18. [doi: 10.1016/j.cca.2015.03.003] [Medline: 25771106]
11. Jørgensen LM, Hansen S, Petersen P, Sölétormos G. Median of patient results as a tool for assessment of analytical stability. Clin Chim Acta 2015 Jun 15;446:186-191. [doi: 10.1016/j.cca.2015.04.024] [Medline: 25920692]

12. Ferreira CE, França CN, Correr C, Zucker M, Andriolo A, Scartezini M. Clinical correlation between a point-of-care testing system and laboratory automation for lipid profile. Clin Chim Acta 2015 Jun 15;446:263-266. [doi: 10.1016/j.cca.2015.04.036] [Medline: 25952166]
13. Ghavami P. An Investigation of Applications of Artificial Neural Networks in Medical Prognostics. Washington: University of Washington; 2012.
14. Lidbury B, Richardson A, Badrick T. Assessment of machine-learning techniques on large pathology data sets to address assay redundancy in routine liver function test profiles. Diagnosis (Berl) 2015 Feb 01;2(1):41-51. [doi: 10.1515/dx-2014-0063] [Medline: 29540013]
15. Judson R, Elloumi F, Setzer R, Li Z, Shah I. A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. BMC Bioinformatics 2008 May 19;9(1):241 [FREE Full text] [doi: 10.1186/1471-2105-9-241] [Medline: 18489778]
16. Freund Y, Schapire R. Experiments with a new boosting algorithm. In: Machine Learning: Proceedings of the Thirteenth International Conference. San Francisco, CA: ICML; 1996:148-156.
17. Miotto R, Li L, Kidd B, Dudley J. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 Dec 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
18. DiBardino D, McElhinney D, Kaza A, Mayer J. Analysis of the US Food and Drug Administration Manufacturer and User Facility Device Experience database for adverse events involving Amplatzer septal occluder devices and comparison with the Society of Thoracic Surgery congenital cardiac surgery database. J Thorac Cardiovasc Surg 2009 Jun;137(6):1334-1341 [FREE Full text] [doi: 10.1016/j.jtcvs.2009.02.032] [Medline: 19464444]
19. Noble D, Pronovost P. Underreporting of patient safety incidents reduces health care's ability to quantify and accurately measure harm reduction. J Patient Saf 2010 Dec;6(4):247-250. [Medline: 21500613]
20. Varallo F, Guimarães Synara de Oliveira P, Abjaude S, Mastroianni P. [Causes for the underreporting of adverse drug events by health professionals: a systematic review]. Rev Esc Enferm USP 2014 Aug;48(4):739-747 [FREE Full text] [Medline: 25338257]
21. de Boissieu P, Kanagaratnam L, Abou TM, Roux M, Dramé M, Trenque T. Notoriety bias in a database of spontaneous reports: the example of osteonecrosis of the jaw under bisphosphonate therapy in the French national pharmacovigilance database. Pharmacoepidemiol Drug Saf 2014 Sep;23(9):989-992. [doi: 10.1002/pds.3622] [Medline: 24737486]
22. Motola D, Vargiu A, Leone R, Conforti A, Moretti U, Vaccheri A, et al. Influence of regulatory measures on the rate of spontaneous adverse drug reaction reporting in Italy. Drug Saf 2008;31(7):609-616. [Medline: 18558794]
23. Reader TW, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. BMJ Qual Saf 2014;23(8):678-689. [doi: 10.1136/bmjqs-2013-002437]
24. Food and Drug Administration. 2017. Regulatory science priorities (FY 2017) URL: https://www.fda.gov/downloads/MedicalDevices/ScienceandResearch/UCM521503.pdf [WebCite Cache ID 6ymdBGozJ]
25. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: Springer; 2013:6.
26. Shoji A, Aris-Brosou S, Culina A, Fayet A, Kirk H, Padget O, et al. Breeding phenology and winter activity predict subsequent breeding success in a trans-global migratory seabird. Biol Lett 2015 Oct;11(10) [FREE Full text] [doi: 10.1098/rsbl.2015.0671] [Medline: 26510674]
27. Ripley B. CRAN.R-project. 2016. tree: Classification and Regression Trees URL: http://cran.r-project.org/web/packages/tree/index.html [accessed 2018-05-03] [WebCite Cache ID 6z8Tx8g1O]
28. Alfaro E, Gámez M, García N. adabag: an R package for classification with boosting and bagging. J Stat Soft 2013;54(2):35. [doi: 10.18637/jss.v054.i02]
29. García F, Pedregal D, Roberts C. Time series methods applied to failure prediction and detection. Reliab Eng Syst Safe 2010 Jun;95(6):698-703. [doi: 10.1016/j.ress.2009.10.009]
30. Moura MD, Zio E, Lins I, Droguett E. Failure and reliability prediction by support vector machines regression of time series data. Reliab Eng Syst Safe 2011 Nov;96(11):1527-1534. [doi: 10.1016/j.ress.2011.06.006]
31. Esteva A, Kuprel B, Novoa R, Ko J, Swetter S, Blau H, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
32. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013 Aug;35(8):1798-1828. [doi: 10.1109/TPAMI.2013.50] [Medline: 23787338]
33. Jordan M, Mitchell T. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
34. Jensen P, Jensen L, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 02;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
35. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008 Feb;77(2):81-97. [doi: 10.1016/j.ijmedinf.2006.11.006]

Abbreviations

CART: Classification and Regression Trees

EHR: electronic health record
FDA: Food and Drug Administration
IVD: in vitro diagnostic
QC: quality control
RCI: root cause investigation

Edited by G Eysenbach; submitted 27.01.18; peer-reviewed by J Green, N Unciano; comments to author 14.03.18; revised version received 27.03.18; accepted 27.03.18; published 15.05.18.

Please cite as:
Aris-Brosou S, Kim J, Li L, Liu H
Predicting the Reasons of Customer Complaints: A First Step Toward Anticipating Quality Issues of In Vitro Diagnostics Assays with Machine Learning
JMIR Med Inform 2018;6(2):e34
URL: http://medinform.jmir.org/2018/2/e34/
doi:10.2196/medinform.9960
PMID:29764796

Original Paper

Finding Meaning in Medication Reconciliation Using Electronic Health Records: Qualitative Analysis in Safety Net Primary and Specialty Care

George Yaccoub Matta[1,2], BA; Elaine C Khoong[1], MS, MD; Courtney R Lyles[1,2], PhD; Dean Schillinger[1,2], MD; Neda Ratanawongsa[1,2], MPH, MD

[1]Division of General Internal Medicine, University of California, San Francisco, San Francisco, CA, United States

[2]UCSF Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, San Francisco, CA, United States

Corresponding Author:
Neda Ratanawongsa, MPH, MD
Division of General Internal Medicine
University of California, San Francisco
1001 Potrero, Box 1364
San Francisco, CA, 94143
United States
Phone: 1 4152063188
Email: neda.ratanawongsa@ucsf.edu

Abstract

Background: Safety net health systems face barriers to effective ambulatory medication reconciliation for vulnerable populations. Although some electronic health record (EHR) systems offer safety advantages, EHR use may affect the quality of patient-provider communication.

Objective: This mixed-methods observational study aimed to develop a conceptual framework of how clinicians balance the demands and risks of EHR and communication tasks during medication reconciliation discussions in a safety net system.

Methods: This study occurred 3 to 16 (median 9) months after new EHR implementation in five academic public hospital clinics. We video recorded visits between English-/Spanish-speaking patients and their primary/specialty care clinicians. We analyzed the proportion of medications addressed and coded time spent on nonverbal tasks during medication reconciliation as "multitasking EHR use," "silent EHR use," "non-EHR multitasking," and "focused patient-clinician talk." Finally, we analyzed communication patterns to develop a conceptual framework.

Results: We examined 35 visits (17%, 6/35 Spanish) between 25 patients (mean age 57, SD 11 years; 44%, 11/25 women; 48%, 12/25 Hispanic; and 20%, 5/25 with limited health literacy) and 25 clinicians (48%, 12/25 primary care). Patients had listed a median of 7 (IQR 5-12) relevant medications, and clinicians addressed a median of 3 (interquartile range [IQR] 1-5) medications. The median duration of medication reconciliation was 2.1 (IQR 1.0-4.2) minutes, comprising a median of 10% (IQR 3%-17%) of visit time. Multitasking EHR use occurred in 47% (IQR 26%-70%) of the medication reconciliation time. Silent EHR use and non-EHR multitasking occurred a smaller proportion of medication reconciliation time, with a median of 0% for both. Focused clinician-patient talk occurred a median of 24% (IQR 0-39%) of medication reconciliation time. Five communication patterns with EHR medication reconciliation were observed: (1) typical EHR multitasking for medication reconciliation, (2) dynamic EHR use to negotiate medication discrepancies, (3) focused patient-clinician talk for medication counseling and addressing patient concerns, (4) responding to patient concerns while maintaining EHR use, and (5) using EHRs to engage patients during medication reconciliation. We developed a conceptual diagram representing the dilemma of the multitasking clinician during medication reconciliation.

Conclusions: Safety net visits involve multitasking EHR use during almost half of medication reconciliation time. The multitasking clinician balances the cognitive and emotional demands posed by incoming information from multiple sources, attempts to synthesize and act on this information through EHR and communication tasks, and adopts strategies of silent EHR use and focused patient-clinician talk that may help mitigate the risks of multitasking. Future studies should explore diverse patient perspectives about clinician EHR multitasking, clinical outcomes related to EHR multitasking, and human factors and systems engineering interventions to improve the safety of EHR use during the complex process of medication reconciliation.

(JMIR Med Inform 2018;6(2):e10167) doi:10.2196/10167

KEYWORDS

medication reconciliation; electronic health records; physician-patient relations; patient safety; communication

Introduction

Clinicians in US safety net clinics—federally funded clinics serving socioeconomically disadvantaged populations [1]—face unique barriers to conducting effective ambulatory visit medication reconciliation. During medication reconciliation, as defined by the US Joint Commission National Patient Safety Goals, a clinician or care team member "compares the medications a patient should be using (and is actually using) to the new medications that are ordered for the patient and resolves any discrepancies" [2]. Although a requirement for safe transition of care, medication reconciliation is also an important patient-centered process for revealing patients' knowledge, concerns, and behaviors around their medications that should inform treatment decision making and can affect adherence [3]. Limited evidence exists for the most effective interventions to integrate medication reconciliation into the workflow of ambulatory care [4]. Meanwhile, safety net patients with limited health literacy and limited English proficiency experience communication barriers that increase their risk of incorrectly reconciled medications and medication error [5-10].

Safety net systems could receive incentives to facilitate electronic health record (EHR) implementation costs by meeting metrics for medication reconciliation for EHR "meaningful use," defined by the US Centers for Medicare and Medicaid Services (CMS) as "the process of identifying the most accurate list of all medications that the patient is taking, including name, dosage, frequency, and route, by comparing the medical record to an external list of medications obtained from a patient, hospital, or other provider" [11]. Although EHR use may improve patient-clinician communication if used to engage patients [12], EHR use may worsen communication by reducing eye contact, increasing silence and clinician multitasking, and shifting talk away from patient-centered topics [13-15]. Thus, EHR use may enhance or decrease patient-centered interviewing important to effective medication reconciliation.

Clinician multitasking—performing two or more tasks simultaneously [16]—may also affect medication reconciliation. Common examples of clinician EHR multitasking include eliciting a history while entering data (voluntary multitasking) or listening to a patient's question while ordering a prescription (externally prompted multitasking) [16-17]. Multitasking may increase risk of errors, either in communication with patients, such as missing cues, or in completing EHR tasks, such as documentation or computerized order entry [16,18-20]. Technology-induced errors may arise if EHRs increase the clinician's cognitive burden because of inadequate EHR design and development, problematic implementation and customization, or negative impacts on sociotechnical work processes [18-19]. If clinicians cope with this cognitive burden by using EHRs in silence, patients may be less satisfied [13,21]. However, delaying EHR use until later may not only risk errors because of potential memory lapses, but also increase clinicians' EHR workload, stress, and burnout [22,23].

To our knowledge, no study has examined ambulatory safety net communication during medication reconciliation using newer EHRs certified for meaningful use. In a prior study, we found that safety net clinicians spent 30.5% of visits multitasking on EHRs, silently used EHRs 4.6% of visit time, and used 33.1% of visit time for focused patient-clinician talk [17].

We conducted a mixed-methods study using observations of real-world safety net ambulatory encounters to develop a conceptual framework of how clinicians balance the demands and risks of EHR and communication tasks during medication reconciliation discussions.

Methods

Study Design and Participants

We conducted this observational study in five primary and specialty care safety net clinics that had recently transitioned (range 3-16 months, median 9) from a "basic" EHR to a CMS-certified "fully functional" EHR [24]. In this ambulatory EHR, medication information was potentially documented in multiple areas:

1. "Current medications list" in the patient EHR chart and "current medications" section of visit notes. The active medication list in a patient's electronic record is automatically imported into a visit note on the day of a visit for clinicians to update. Clinicians can check a box labeled "verified medications," placing a phrase in the note "medication list reviewed and verified with the patient" and meeting the "meaningful use" CMS metric. At the time of this study, refill data from pharmacy claims was not available.
2. History of present illness. Clinicians narratively document the patient-reported medication concerns or behaviors.
3. Assessments and plan. Clinicians may narratively document references to patient concerns or behaviors influencing their decision making.
4. Treatment and orders. Rather than resolving discrepancies in the current medication list, some clinicians may adjust medications in the treatment section or type instructions for patients.

Eligible patients included English- or Spanish-speaking adults (age >18 years) with at least one of three chronic medical conditions (diabetes, congestive heart failure, or rheumatoid arthritis) who received primary care in the adult internal medicine or family medicine clinic *and* subspecialty care at a diabetes, cardiology, or rheumatology clinic [15,17,25]. Eligible clinicians included physicians, nurse practitioners, fellows, and residents. All participants provided written informed consent, and the Institutional Review Board of the University of California, San Francisco, approved the study.

Data Collection

For participating dyads, we collected the following data from one clinician-patient visit [15,17,25]:

1. Structured previsit and postvisit interviews with patients to collect sociodemographic and medical data. Postvisit patient interviews occurred in person or via telephone. Native Spanish speakers translated and back-translated Spanish interview items into English.
2. Online questionnaires with providers to collect sociodemographic data.
3. Visit note written by clinician in the EHR.
4. Video recording of patient and provider visit.

Data Measurements

Sociodemographic Information

We used previously validated self-report screening questions to determine patients' English proficiency and health literacy [26-28]. We categorized Spanish-speaking patients who reported English proficiency less than "very well" as having limited English proficiency [26]. We categorized patients who were "somewhat," "a little bit," or "not at all" "confident filling out medical forms by yourself" as having limited health literacy [27,28].

Electronic Health Record Visit Documentation

We reviewed the EHR note corresponding to the videotaped interactions for the current medication list, any narrative text in any note section referring to medications, and the medications listed in the treatment section.

Number of Relevant Medications and Proportion Addressed

From the EHR visit note, we abstracted the number of medications on the EHR list or in the text of the visit note. Primary care providers are expected by national standards and local hospital policy to reconcile all medications, including any over-the-counter or nutritional supplements [3,11]. The local hospital policy specifies that specialty care medication reconciliation is required for "all medications related to their specialty, including those that may have drug or disease interactions" [29]. Thus, in addition to the total number of medications, we also created a category of "relevant medications." Classified by a physician investigator (NR), we included the following medications (listed in the note and discussed during the visit) as relevant:

1. Primary care: all medications, including any over-the-counter or nutritional supplements.
2. Cardiology: all antihypertensive, diuretic, antiplatelet, lipid-lowering, antiarrhythmic, or pulmonary hypertension medications.
3. Rheumatology: all immunosuppressant or analgesic medications or medications to mitigate those regimens' side effects (eg, folic acid with methotrexate or bisphosphonate with prednisone).
4. Diabetes: all oral or injectable hypoglycemic medications, antihypertensive medications, lipid-lowering medications, and aspirin. We excluded glucose monitoring supplies.

We considered a medication explicitly addressed if the patient or clinician specifically discussed its current or past use. We then summed the number of medications from the EHR medication list that were explicitly addressed during the visit, compared with both the number of relevant medications and the number of total medications.

Meaningful Use Indicator

We abstracted this if notes contained the phrase "medication list reviewed and reconciled with the patient."

Medication Reconciliation Duration

We classified visit time as related to medication reconciliation during segments when clinicians or patients demonstrated behaviors to compare patients' current medication-taking behaviors with the clinicians' available medication lists or to elicit and respond to patients' medication-related concerns and beliefs. We did not include discussions of newly prescribed or newly recommended medications. This duration included both verbal statements and nonverbal behaviors. Patient verbal statements included those elicited by clinicians' questions or volunteered independently. Clinician verbal statements included those elicited by patients' questions or volunteered independently. Nonverbal behaviors included clinicians' visual inspections of patient medication bottles or paper medication lists as well as data entry or review using the EHR. We then calculated total visit minutes and proportion of visit time spent on medication reconciliation.

Analysis

Electronic Health Record Use and Non-Electronic Health Record Behaviors During Medication Reconciliation

We categorized each segment of medication reconciliation time as mutually exclusive categories to describe whether clinicians conducted EHR-related or non-EHR-related tasks [17]: multitasking EHR use (clinician or patient speaking while clinician EHR use), silent EHR use (≥3 seconds silence), non-EHR multitasking (eg, paper chart, glucometer use), silent non-EHR use (≥3 seconds silence), and focused clinician-patient talk (no multitasking on EHR or non-EHR use).

Analysis of Communication and Electronic Health Record Use Patterns

We then conducted qualitative analysis to uncover patterns of communication and EHR use during medication reconciliation. We used a grounded theory approach, by which a theory to explain a phenomenon is derived from the data itself [30]. Two investigators (GYM and NR) independently analyzed a subset of videos, generating codes and negotiating discrepancies to create a preliminary coding template. One investigator (GYM) coded each remaining video with this template. Application of the template for coding and modifications to the coding template were arrived at by consensus. We conducted qualitative analysis using ATLAS.ti version 7.5.15 (Scientific Software Development GmbH, Berlin, Germany). We chose representative quotes highlighting these patterns. From these codes and patterns (Multimedia Appendix 1), we developed a conceptual framework of how clinicians balance the demands

and risks of EHR and communication tasks during medication reconciliation discussions.

Results

Visits and Participants

We recorded 35 visits (17 primary and 18 specialty care) between 25 patients and 25 clinicians. Table 1 shows patient, clinician, and relationship characteristics. Patients were mean 57 (SD 11) years in age, 44% (11/25) were women, 48% (12/25) were Hispanic/Latino, and 20% (5/25) had limited health literacy. The majority reported that their health was "poor" (40%, 10/25) or "fair" (20%, 5/25). Among clinicians, most were women (72%, 18/25) and 48% (12/25) were primary care physicians (PCPs). Among the 35 visits, 51% (18/35) were in primary care, and 40% (14/35) reported receiving care from the clinicians for more than 5 years. The median visit length was 20.6 (interquartile range [IQR] 16.7-32.2) minutes, and 17% (6/35) were in Spanish.

Task Performance and Medications Addressed During Medication Reconciliation

Table 2 describes the summary characteristics of medication reconciliation during each visit. The median duration of medication reconciliation was 2.1 minutes (interquartile range [IQR] 1.0-4.2 minutes), and medication reconciliation comprised a median of 10% (IQR 3%-17%) of visit time. EHR multitasking comprised a median of 47% (IQR 26%-70%) of medication reconciliation time. Silent EHR use and non-EHR multitasking occurred a smaller proportion of medication reconciliation time, with a median of 0% for both (IQR 0%-6% and IQR 0%-13%, respectively). Silent non-EHR tasks were not performed during medication reconciliation. Focused clinician-patient talk occurred a median of 24% (IQR 0%-39%) of medication reconciliation time. The median for total medications was 13 (IQR 9-17), with a median of 7 (IQR 5-12) relevant medications. The median number of addressed medications was 2 (IQR 1-5).

Figure 1 depicts duration of medication reconciliation and the proportion of activities occurring during medication reconciliation for each visit, with primary care visits labeled as "P" and specialty care visits were labeled as "S." The end of each bar is labeled with the number of medications explicitly addressed out of all relevant medications, with an asterisk indicating if the meaningful use medication reconciliation box was checked in the visit note. For specialty care encounters, the total of all medications is listed in parentheses. Among 35 visits, 29 (83%) involved a medication reconciliation discussion, almost always interspersed with other content and tasks throughout the visit, rather than in a single, uninterrupted segment. Clinicians multitasked on EHRs during medication reconciliation in 28 (80%) visits, with EHR multitasking occurring during the entirety of medication reconciliation in 5 (14%) of the visits. Silent EHR use occurred in 12 (34%) visits. The meaningful use medication reconciliation box was checked in 19 (54%) visit notes.

Communication and Electronic Health Record Use Patterns During Medication Reconciliation

Five sets of communication patterns with EHR medication reconciliation were observed. Within a given visit, multiple patterns could have been observed.

Pattern 1: Typical Electronic Health Record Multitasking for Medication Reconciliation

In the most common pattern, clinicians reviewed and added information in the EHR current medication lists while talking with patients about their medications, as demonstrated in visit S11 (female rheumatologist, female patient, Spanish-concordant; 100% EHR multitasking; 1 minute):

> *[Clinician sits facing the EHR current medication list, with the patient adjacent to the monitor, facing the clinician.]*
>
> *Clinician: "So you're taking Enbrel every week?"*
>
> *Patient: "Yes."*
>
> *Clinician: [clicks box checked, scrolls] "Okay, and you're still taking hydroxychloroquine—"*
>
> *Patient: "Yes."*
>
> *Clinician: "—once a day."*
>
> *Patient: "It's once now?"*
>
> *Clinician: [shifts gaze to patient for <1 second, then back to EHR] "You're taking it twice?"*
>
> *Patient: "Yeah, it was like that."*
>
> *Clinician: [clicks on medication in list, clicks to change frequency] "Okay, okay."*

Table 1. Patient, clinician, and visit characteristics in a study of electronic health record use in safety net primary and specialty care medication reconciliation.

Characteristics	Value
Patients (n=25)	
Age (years), mean (SD)	56.8 (11.0)
Gender (female), n (%)	11 (44)
Race/ethnicity, n (%)	
Hispanic	12 (48)
Asian	6 (24)
Caucasian	4 (16)
African-American	2 (8)
Multiethnic	1 (4)
Language, n (%)	
Primary language Spanish	10 (40)
Limited English proficiency	6 (24)
Education, n (%)	
≤8th grade education	2 (8)
Some high school or high school graduate/General Education Diploma	7 (28)
Some college or college graduate	16 (64)
Limited health literacy, n (%)	5 (20)
Income (≤US $20,000/year), n (%)	23 (92)
“Poor” or “fair” quality of life, n (%)	18 (60)
Clinicians (n=25)	
Age (years), mean (SD)	44.9 (11.9)
Gender (female), n (%)	14 (67)
Clinic, n (%)	
Primary care clinic	14 (56)
Diabetes clinic	5 (20)
Cardiology clinic	3 (12)
Rheumatology clinic	3 (12)
Role, n (%)	
Physician	21 (84)
Nurse practitioner or physician assistant	4 (16)
Years since professional degree, mean (SD)	15.7 (11.3)
Visits (n=35)	
Relationship length years at baseline, n (%)	
<1 year	2 (6)
1-5 years	19 (54)
>5 years	14 (40)
Language during visit, n (%)	
English	29 (83)
Spanish	5 (14)
Spanish interpreter	1 (3)
Visit length (minutes), median (interquartile range)	20.6 (16.7-32.2)

Table 2. Characteristics of medication reconciliation during safety net primary and specialty care visits (n=35).

Medication reconciliation characteristics	Value
Medication reconciliation duration (minutes), median (IQR[a])	2.1 (1.0-4.4)
% of medication reconciliation time spent performing activities, median (IQR)	
Multitasking EHR[b] use	47 (26-70)
Silent EHR use	0 (0-6)
Non-EHR multitasking	0 (0-13)
Focused patient-clinician talk	24 (0-39)
Number of total medications[c], median (IQR)	13 (9-17)
Number of relevant medications[c], median (IQR)	7 (5-12)
Number of relevant medications addressed[d], median (IQR)	2 (1-5)

[a]IQR: interquartile range.

[b]EHR: electronic health record.

[c]The total medications included all listed in the patient's note or discussed during the visit encounter. All medications were categorized as relevant for primary care encounters. For specialty care encounters, relevant medications were related to the clinician's specialty and those with drug or disease interactions.

[d]Medications were categorized as "addressed" if the patient or clinician specifically discussed its current use.

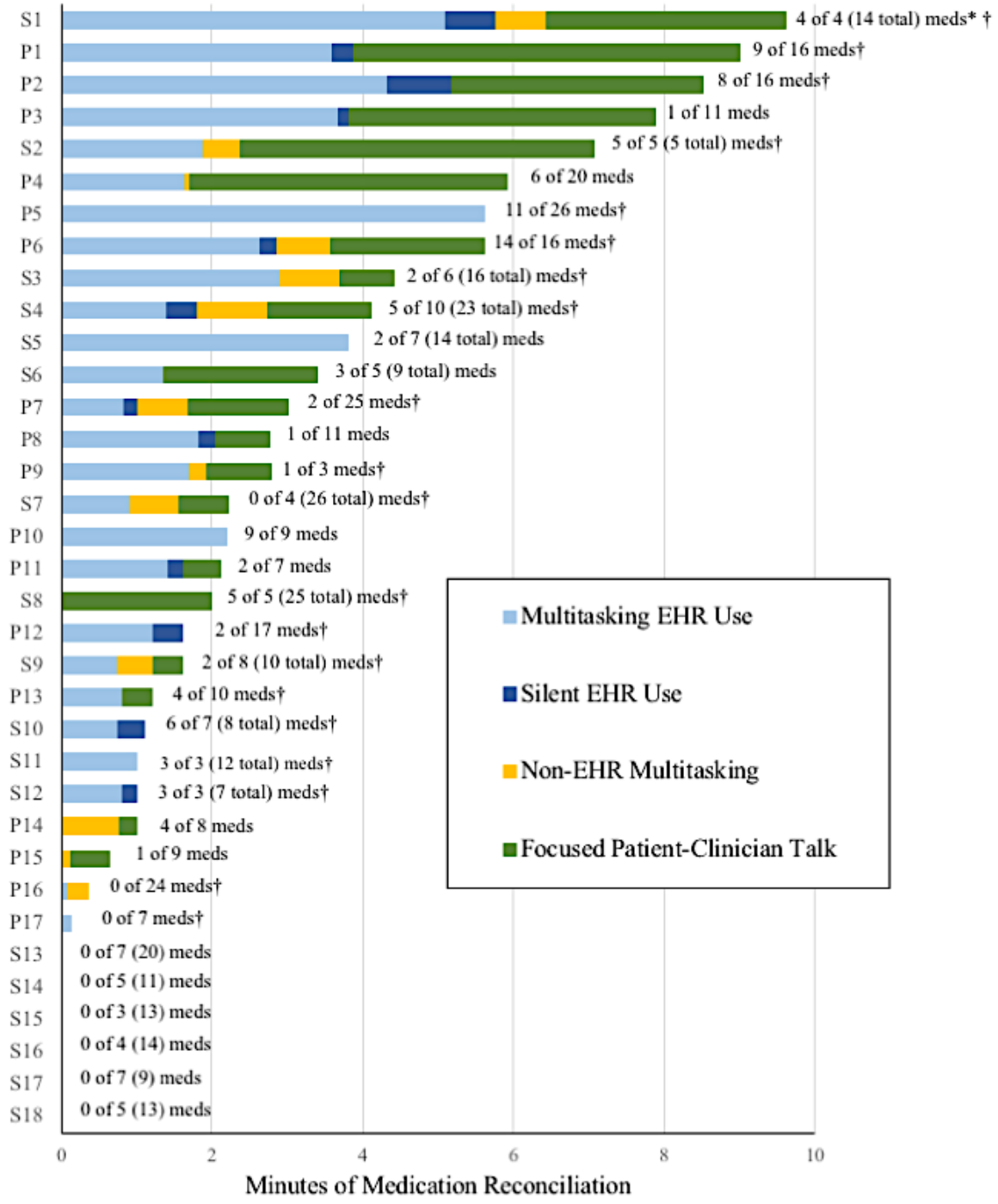
Occasionally this also included interspersed EHR multitasking with non-EHR multitasking, such as looking at pill bottles or paper medication lists to update the EHR current medication list. In encounter P11 (male PCP, male patient, English-concordant; 68% EHR multitasking, 9% silent EHR, 24% focused clinician-patient talk; 2.1 minutes) a clinician used the EHR and pill bottles to check a recently uninsured patient's medications:

> *[Clinician sits facing EHR screen. Patient sits adjacent to the screen, facing clinician, and takes pill bottles out of bag.]*
>
> *Clinician: [holds bottle in hand, looks at label for 1 second] "Are you taking these—" [puts down bottle, looks at EHR current medication list] "—every day or only once in a while?"*
>
> *Patient: [takes out other pill bottles] "Once in a while."*
>
> *Clinician: [begins typing into medication list] "Okay."*
>
> *Patient: [shakes bottle] "Uh, I'm down to one here."*
>
> *Clinician: [glances at pill bottle for <1 second, then to EHR] "Okay, would you say all of these you're taking just once in a while?" [looks at bottle for <1 second then to EHR]*
>
> *Patient: "Mhm. This one's gone." [shakes bottle]*
>
> *Clinician: [looks at bottle for 1 second, back to EHR] "Okay."*

Pattern 2: Dynamic Electronic Health Record Use Beyond the Current Medication List

To negotiate medication discrepancies, clinicians often navigated beyond the current medication list, using multiple sections within the note and the entire electronic chart, including past visit notes, notes from other settings, and test results. Clinicians multitasked, navigating and reading EHR sections while eliciting or listening to information from the patient. Clinicians also interspersed EHR multitasking with silent EHR use to concentrate on reviewing information or completing tasks. Patients often offered social talk breaking this silent EHR use and triggering clinician multitasking.

Figure 1. Multitasking, silent electronic health record (EHR) use, and number of medications explicitly addressed during safety net medication reconciliation (N=35). *Primary care encounters are labeled with a "P" and specialty care encounters with an "S." The number following each line indicates the number of "addressed" medications out of the total number of "relevant" medications. Medications were categorized as "addressed" if the patient or clinician specifically discussed its current use. For primary care encounters, all medications listed in the patient's note or discussed during the visit encounter were categorized as "relevant." For specialty care encounters, medications related to the clinician's specialty or with drug or disease interactions were categorized as "relevant"; the total number of all medications is listed in parentheses for these specialty encounters. † means clinicians clicked on a box labeled "verified medications" to indicate that medication reconciliation was performed.



During visit P1 (female PCP, male patient, English-concordant; 40% EHR multitasking, 3% silent EHR use, 50% focused patient-clinician talk; 3.6 minutes), the clinician spent 3 minutes attempting to verify one medication, using multiple EHR sources of information (the medication list, three past clinician notes from the current and previous EHR), pill bottles, and patient history. Silent EHR use occurred when the clinician reviewed a previous visit note:

> *[Clinician sits facing EHR screen with hand on mouse. EHR screen is angled toward patient who sits facing clinician. Pill bottles are on the desk.]*

Clinician: [looks at pill bottle] "Okay...You're taking this one, the torsemide?" [shows patient pill bottle while clicking into the EHR current medication list]

Patient: [looks at pill bottle] "Yeah. Once a day."

Clinician: [scrolling through medication list] "Is that a new one?"

Patient: "No."

[Clinician looks at pill bottle, then back to medication list.]

Patient: [looking at clinician] "Why? It's not on my list?"

Clinician: [begins typing in medication list search function] "It's not, actually."

Patient: [putting away backpack] "Then maybe I shouldn't take it. It's an old one."

[Clinician looks down at pill bottle for one second, then back to medication list. Clinician exits out of medication list to History of Patient Illnesses section.]

Clinician: [clicking into list of previous clinician notes] "Hold on. When was the last time you saw these cardiac doctors?"

Patient: [looking at clinician while clinician scrolls down list of past visits] "Um, like two months ago. Yeah, two months ago." [Clinician clicks into a cardiology visit note.] "I think I've got another appointment coming up."

Clinician: [scrolls through note, pill bottle still in hand] "Okay, hold on, I'm going to try to read their note here."

[5 seconds of silence as clinician looks at note. Patient looks down at hands.]

Patient: "How many years did it take for you to become a doctor like you are?"

[2 minutes of EHR multitasking; clinician clicks and scrolls through five different EHR sections while engaging in social talk with the patient.]

Clinician: [scrolls through note] "Okay here's the deal. I can't find...umm..." [highlights information in note] "...I cannot find this guy—" [shakes pill bottle] "—in the notes from cardiology." [turns to patient] "Okay, so we're going to stop this, because this medicine acts the same as this one." [picks up another pill bottle and shows patient] "So we don't want two medicines doing the same thing. So you're not going to use that anymore."

Pattern 3: Focused Patient-Clinician Talk for Medication Counseling and Addressing Patient Concerns

Clinicians demonstrated different ways of interspersing focused patient-clinician talk with multitasking or silent EHR use. For example, clinicians used brief periods of focused talk to address patients' medication or health concerns arising during medication reconciliation. In visit P2 (female PCP, male patient, Spanish-concordant; 51% EHR multitasking, 10% silent EHR use, 39% focused patient-clinician talk; 8.5 minutes), a clinician stopped multitasking on the EHR to address the patient's worry about a side effect:

[Clinician sits with body toward the patient while clicking off medications in the EHR current medication list and examining pill bottles for 4 seconds.]

Patient: "For example, the one for the heart, they told me it could have a side effect that feels like arthritis." [points to chest] "That's what they told me."

Clinician: [slightly turns head from screen, looks at patient] "Well, those medications for gout and arthritis aren't really related to the heart. Don't worry about that."

[Patient hands clinician pill bottle. Clinician looks at it for <1 second and looks back to patient.]

Clinician: "And no—I don't believe—It is true that gout could be a secondary effect of the medicine, but it's not—" [turns head to EHR, pauses for 1 second] "Well, that's not true." [turns head back to patient] "The furosemide, the diuretic, could possibly—"

Patient: [hands clinician another pill bottle] "This one?"

Clinician: "Yeah, this one."

Patient: "I've been taking this one for 25 years."

Clinician: "But it's not that the number of years affects the gout, it's just that in the moment that you're taking it you have greater risk of gout."

Patient: [takes out other pill bottles] "Okay."

Clinician: "Thank you for bringing these, I'm going to review these medications with the list in the computer." [turns head to EHR screen]

The clinician then interspersed silent EHR use and EHR multitasking using pill bottles, updating the current medication list. While EHR multitasking, the clinician found out from the patient that he was no longer receiving colchicine for gout due to insurance limitations. After completing reconciliation of all pill bottles, the clinician readdressed the patient's concern through focused talk:

[Clinician's body oriented toward patient with full eye contact.]

Clinician: "So about the gout...It's true that this—" [holds up pill bottle] "—has a side effect, and although your gout may be better controlled without this medicine, the rest of your body would be worse off without it—" [laughs] "—because you need this furosemide, and any medicine that removes water affects gout. So, I would like to continue the same with this medication—" [points to pill bottle] "—and I would like to increase—" [picks up other pill bottle] "—this medicine, the allopurinol, to prevent more gout attacks. What do you think of this plan? You agree, too?"

Pattern 4: Responding to Patient Concerns While Maintaining Electronic Health Record Use

Patients often revealed concerns about their medications and nonmedication topics during medication reconciliation, sometimes without clinician elicitation. Some clinicians responded by continuing to multitask, offering expressions of empathy or exploring patients' concerns while continuing to use the EHR.

In visit P12 (female PCP, female patient, English-concordant; 76% EHR multitasking, 24% silent EHR; 1.6 minutes), the clinician addresses the patient's concern verbally while continuing EHR tasks:

> *[Patient sits in a wheelchair adjacent to EHR monitor, facing the clinician. The clinician is walking toward EHR after sanitizing hands].*
>
> *Patient: [concerned tone] "I was going to ask you, with my Ativan—" [clinician sits and faces EHR] "—if you can increase it."*
>
> *Clinician: [scrolls through History of Patient Illnesses] "Tell me about that. Tell me why you want to increase it."*
>
> *Patient: [looking at clinician] "Because...I'm under a lot of stress..."*
>
> *Clinician: [scrolls through EHR list of visits, clicks into past visit note, inquisitive tone] "Mhm."*
>
> *Patient: "And what I was taking; it's not working for me."*
>
> *Clinician: "Mm-hm." [past visit note loads] "Tell me what else we're doing to help." [concerned tone] "I know you're under a lot of stress." [scrolls through note] "We've talked about, before, worrying about the side effects of the Ativan and I think there may be better medicines for you to take to deal with the stress."*

At times, clinicians did not respond to the patients' concerns, continuing their multitasking or silent EHR use. In visit P5 (male PCP, female patient, English-concordant; 100% EHR multitasking; 5.6 minutes), the patient describes pain and depression which the clinician does not address during the visit:

> *[Clinician sits facing EHR screen, hand on mouse. Patient sits adjacent to monitor, facing clinician. Clinician has been EHR multitasking; discussing the patient's medication-taking behaviors for 1 minute.]*
>
> *Clinician: "...And, the other ones were the um, medication for your stomach."*
>
> *Patient: [looks away from clinician] "Yes, and also the other one is a cough syrup." [concerned tone] "Yeah, sometimes I have—"*
>
> *Clinician: [clicks medication in EHR current medication list, flat tone] "Okay."*
>
> *Patient: [looks back at clinician] "—tremendous pain."*
>
> *Clinician: [gaze on EHR, flat tone] "Okay, and are you still taking the Duloxetine? It's like an antidepressant." [glances <1 second at patient then back to EHR]*
>
> *Patient: [gaze on clinician, concerned tone] "Yes. And when I take that in the morning, it makes me, you know." [puts hand in a fist]*
>
> *Clinician: [clicks in EHR current medication list, monotone] "Okay, you take that in the morning."*
>
> *Patient: [looking at clinician] "—I stop to cry."*
>
> *Clinician: [scrolls in medication list, flat tone] "Uh huh."*
>
> *Patient: [looking at clinician] "I stop to cry."*
>
> *Clinician: [shifts gaze to patient, flat tone] "You stop to cry?"*
>
> *Patient: "Yes. When I take that for the uh—" [clinician nods, shifts gaze back to EHR] "—when I stop to take the medication for the depression, I get so sensitive. So sensitive." [motions hand toward heart] "When I take my two pills in the morning, I am strong." [laughs]*
>
> *Clinician: [scrolling in medication list, flat tone] "Okay."*
>
> *Patient: "Yeah, I found out because I take notes also for my medication."*
>
> *[Clinician nods, maintaining gaze on EHR.]*

Pattern 5: Clinicians Using Electronic Health Records to Engage Patients During Medication Reconciliation

In two encounters, clinicians with high levels of EHR use engaged patients, through screen sharing, transparent disclosure of EHR tasks, and shifting bodily orientation toward their patients.

In visit P10 (female PCP, male patient, English-concordant; 100% EHR multitasking; 2.2 minutes), the clinician shared her screen to review all nine of the patient's medications, with the patient sitting next to her reviewing the EHR list actively:

> *Clinician: [clicks into current medication list] "So your meds..."*
>
> *Patient: [looks at screen and reads] "Sildenafil, five tablets three times a day. Yes."*
>
> *Clinician: [clicks to check off medication] "You're taking furosemide..."*
>
> *Patient: [looks at clinician, then screen] "Which one is that?"*
>
> *Clinician: "That's the water pill."*
>
> *Patient: [looks at screen] "Uh, only if I need it. If I have swollen ankles."*
>
> *Clinician: [clicks to pull up text box] "How often are you taking it?"*
>
> *Patient: [turns head from screen to clinician] "I haven't taken it in probably 6 months."*
>
> *Clinician: [quickly turns toward the patient, makes eye contact, smiles, and turns back to screen] "Wow!" [turns back to the screen and types] "I'm going to*

leave it on your list but I'm going to put 'as needed.' How about that?"

Patient: [looks at screen] "Yeah. 'As needed' sounds good."

In visit P6 (female PCP, male patient, English-concordant; 47% EHR multitasking, 4% silent EHR tasks, 13% non-EHR multitasking, 36% focused patient-clinician talk; 5.6 minutes), the clinician screen shares while multitasking and transparently explains the need to use EHR silently:

[EHR in triangle between clinician and patient. Clinician's body is angled half toward patient and half toward EHR]

Clinician: [looking directly at patient] "I want to go over your medicines. Did you bring your box thing?"

Patient: "I did not bring it."

Clinician: "OK" [glances at EHR med list, then back to patient] "Would you recognize the names?"

Patient: "Yeah! Yeah!"

Clinician: "OK let's look through them." [turns screen to share with patient] "Now I'm going to show you this thing. Can you see it?"

Clinician: [looking at screen with patient] "The furosemide—how are you taking that one?" [turns gaze to patient]

Patient: "I'm taking the one 80 pill in the morning and then at night."

Clinician: "And how late do you take the second one?"

Patient: "The second one I take it around dinner time."

Clinician: "OK good, because you don't want to be up all night peeing..."

Patient: "Yeah..."

[They troubleshoot timing of the diuretic medication with focused talk]

Conceptual Diagram: Multitasking Clinicians Balancing the Demands and Risks of Electronic Health Records and Communication Tasks During Medication Reconciliation

Based on this analysis, we developed a conceptual diagram representing the multitasking clinician balancing the cognitive and emotional demands posed by incoming information from multiple sources, attempting to synthesize and act on this information through EHR and communication tasks, and adopting strategies that may help mitigate the risks of this multitasking (Figure 2).

Because most clinicians multitask during medication reconciliation (pattern 1), this complex process is represented by solid black arrows demonstrating the input and output of information (1) between the clinician and the EHR, (2) between the clinician and the patient, and (3) between the clinician and the patient's medication bottles or paper medication lists. To complete medication reconciliation, the clinician searches across the EHR chart to find, read, and process complex information entered by multiple members of clinical care teams (pattern 2). Meanwhile, clinicians navigate into multiple sections of the visit note to enter data relevant to medication reconciliation (pattern 2). At the same time, the clinician hears and processes complex patient histories about patients' medication-taking behaviors and concerns, mixed with overt and subtle clues that offer empathic opportunities.

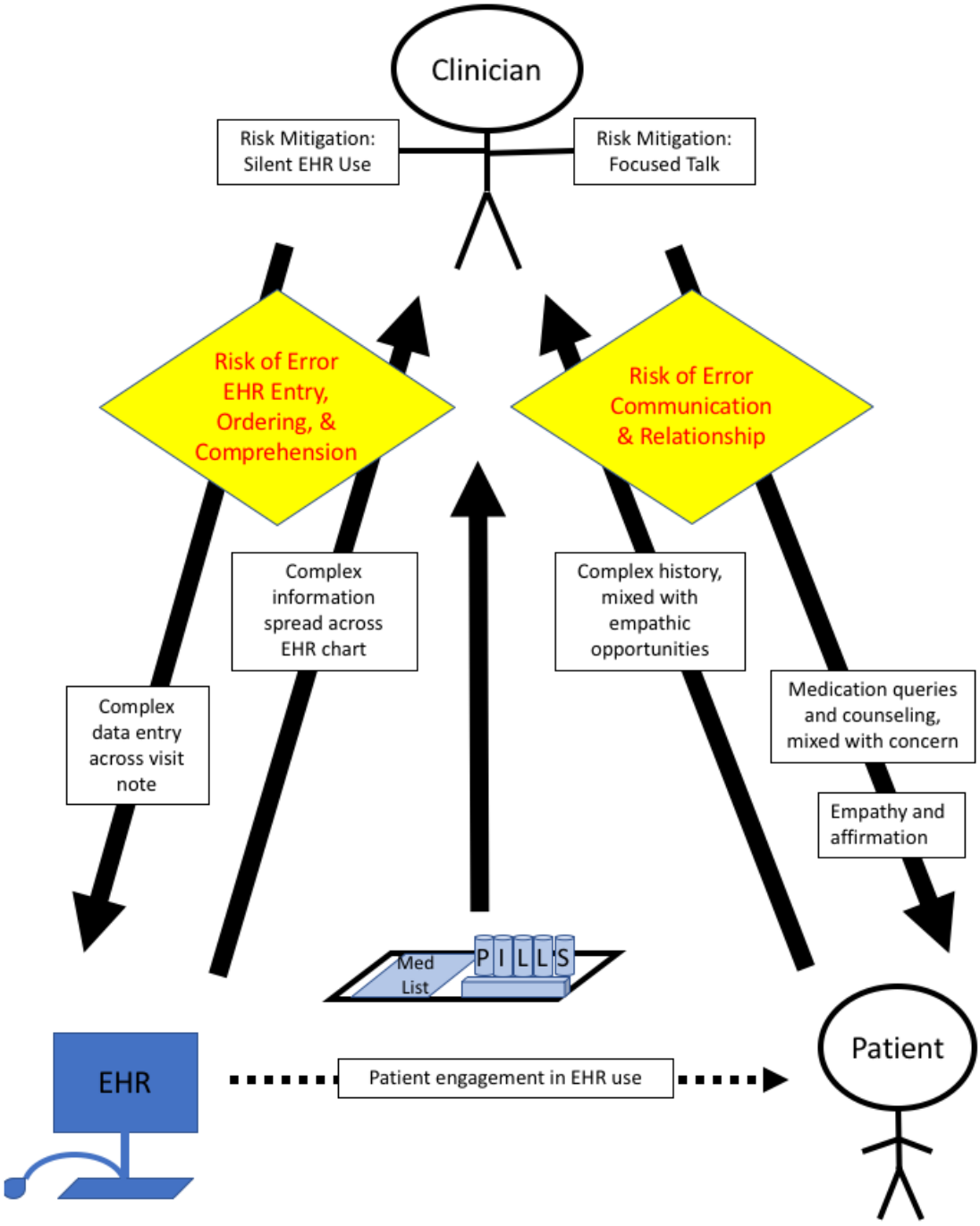
The clinician usually responds by eliciting more information and providing counseling, expressing concerns over medication discrepancies and suboptimal adherence (patterns 3 or 4). Sometimes, the clinician takes advantage of empathic opportunities to offer support and affirmation (patterns 3 and 5). Sometimes, the clinician does not recognize the empathic opportunities—potentially distracted by EHR tasks—or chooses to let the opportunities pass to continue completing medication reconciliation tasks (pattern 4).

The multitasking clinician has risks for error on two fronts, represented by the yellow diamonds. The clinician may make errors in EHR entry, ordering, and comprehension or errors in patient-clinician communication. Of note, communication errors may lead not only to risks to the patient-clinician relationship, but also errors in diagnosis and management if clues about patient symptoms or behaviors are not caught and addressed.

Consciously or unconsciously, clinicians may mitigate these errors by focusing on one interaction at a time. To mitigate the risk of EHR-related errors, clinicians may take periods of silence to focus solely on the EHR, investigating medication discrepancies with more complex and dynamic EHR use. To mitigate the risk of communication errors and relational damage, clinicians may cease EHR multitasking to focused patient-clinician talk.

This diagram's focus is on the multitasking clinician. However, on rare occasions, patients actively engaged with the EHR use by watching their clinicians use the EHR or after receiving explicit invitation by clinicians to join the process. Potentially, this relationship-centered EHR use offers a third option for risk mitigation, allowing clinicians to feel more comfortable using the EHR with the patients at their side.

Figure 2. Conceptual diagram: multitasking clinicians balancing the demands and risks of electronic health records (EHRs) and communication tasks during medication reconciliation.



Discussion

In this safety net study, patient-clinician visits exhibited interesting variations in the depth of medication reconciliation discussion and the patterns of EHR use to support that discussion. The absolute number and proportion of total medications addressed bore little relationship with the length of time, and rather than occurring in a "medication reconciliation block," these discussions were interspersed with other content, including biomedical, psychosocial, and social talk. Overall, clinicians asked simple closed or open-ended questions about the patient's medication-taking behaviors overall, with deeper investigation about only a subset of medications. Some clinicians employed the more detailed, patient-centered interviewing recommended for learning about a patient's medication beliefs and burdens, but not about each medication on a given list. Of note, the longest medication reconciliation discussion lasted almost 10 minutes, addressing all four of the most relevant medications to that specialist, while leaving untouched the 10 other medications on the patient's list. In a cross-sectional analysis, one cannot determine if the depth and length of these medication reconciliation discussions were affected by EHR use or the clinician's pre-EHR approach to medication reconciliation, uninfluenced by the meaningful use requirement. Given that many relationships were more than 5 years, some clinicians may know some of this information from past visits.

However, our results suggest that a comprehensive, patient-centered medication reconciliation interview with medically and psychosocially complex patients may be time-consuming for a safety net primary care or specialty clinician to conduct on their own.

Safety net patients experience barriers to patient-clinician communication and have higher medication reconciliation needs [5-10]. Limited health literacy is associated with poorer ability to interpret medication labels and their instructions, poorer ability to demonstrate taking medications, and inability to identify medications leading to a higher number of unreconciled medications [6-8]. Limited English proficiency, which often interacts with limited health literacy, is also associated with nonadherence to newly prescribed medications, errors in demonstrating how to measure doses of medications, and poorer knowledge of both chronic medications and medication changes on hospital discharge [31-35]. Our sample was also chronically ill with poor or fair quality of health and a high medication burden, also shown to be a risk factor for not taking medications on the current medication list [9,10]. Thus, medication reconciliation for safety net clinicians may be more important, but also more complex.

Meanwhile, multitasking EHR use comprised almost half of medication reconciliation time, a higher proportion of time than in the visits as a whole [17]. Many clinicians may not believe they are multitasking when updating the EHR medication list while talking to patients because both tasks are concordant with the goals of reconciling medications [36]. However, research in cognitive psychology suggests that the act of reading or entering computer data while listening or talking with another person may increase both the risk of errors and the time required to complete each task [16]. Although this risk may be lower when a patient is affirming the accuracy of the current medication list, this risk increases when the clinician is conducting more complex cognitive steps required to negotiate medication discrepancies.

Clinical multitasking predated EHRs, with clinicians reviewing paper charts or patient pill bottles while interviewing patients. EHRs have the potential to reduce errors overall [37-39] by reducing the cognitive difficulties in this work by synthesizing and organizing information in accessible, usable formats, supporting clinical decision making, and offering new information that was previously unavailable, such as medication refill data. However, research is increasingly recognizing the risk of technology-induced errors arising from a technology's design and development, implementation and customization, and resultant human-computer interactions and sociotechnical work processes [19,20].

Our study suggests that clinician multitasking is associated with important risks of errors in communication, which pose not only dangers to the relationship but also the accurate diagnosis and management of the patient's medical and psychosocial needs. As seen in pattern 4, sometimes patients express medication and nonmedication concerns that would require deeper exploration. These concerns may signal suboptimal adherence, an undiagnosed or undertreated medical or psychosocial condition, and a need for empathy and reassurance from the clinician.

As the conceptual diagram depicts, multitasking clinicians using an EHR for medication reconciliation must cope with the cognitive and emotional burdens of this work while managing many other tasks. When patients disclose nonadherence or offer their concerns about medications or other topics, through overt or more subtle clues, clinicians have a series of cognitively challenging tasks. First, they must recognize the cues, which may be more challenging during EHR multitasking when their gaze and attention is not focused on the patient. Second, if they recognize the clues, clinicians must choose whether to respond through further exploration at that time (with focused talk or continued multitasking) or by deferring the exploration until after completing their medication reconciliation tasks. If clinicians respond at the time, they risk making a mistake, both in the EHR tasks and in their communication. If they do not respond at the time, they risk missing important patient engagement opportunities, to the detriment of both patient satisfaction and patient care. This study offered examples of all those patterns and the transitions across them, but this study cannot reveal how many of the clinicians' choices were intentional to mitigate risk or subconscious. We also do not know what these patients would have preferred and what they felt about their clinicians' communication and EHR use behaviors in those moments.

Finally, a few clinicians in this sample exemplified behaviors of harnessing EHRs to further relationship-centered communication [40]. In addition to sharing the EHR screen through inclusive positioning [41], these clinicians used body language, eye contact, affective tone of voice, and empathy statements to elicit and respond to patients' concerns. In those cases, patient-centered medication reconciliation addressed the majority of these patients' medications in less than 6 minutes. This kind of EHR use likely has many facilitators, including the clinicians' communication style, their existing relationships with their patients, their computer and EHR proficiency, and the encounter room positioning. However, these examples lend support for recent efforts to teach real-time EHR use during patient-clinician encounters [42], since shifting all EHR documentation to nonvisit times or other team members may not be possible or sustainable for most clinicians. Clinicians need additional training and support on how to transition intentionally between multitasking, silent EHR use, and focused clinician-patient talk when appropriate to the situation, using strategies to communicate these transitions transparently to their patients.

This analysis also adds to the growing literature about newer generation EHRs in the United States under the meaningful use incentives programs, particularly in a safety net primary and specialty care setting. This study adds to the call for clinicians, health systems, and policymakers to redefine medication reconciliation to acknowledge the multiple levels of medication reconciliation that incorporates the patient's full perspective, including eliciting opportunities to describe and reduce the physical, emotional, and economic burdens of medication regimens [3]. Better measures of high quality medication reconciliation, incorporating the patient's and clinician's

perspectives, are needed. Mandates and incentives to promote medication reconciliation are insufficient for promoting high quality medication reconciliation and may increase the risk of technology-induced errors associated with currently designed and implemented EHRs. Health information technology should be designed and developed, using human factors and systems engineering framework, to facilitate engagement across the appropriate members of the health system care team, community pharmacists, family members and caregivers, and the patient.

This study's limitations should be considered. First, our results may be affected by volunteer bias among clinicians or patients. Second, the sampled visits occurred early after EHR implementation and may not represent medication reconciliation after clinicians spent more time using the EHR, although other researchers have found that early EHR and communication behaviors may be similar to those measured later after implementation [12]. Third, this cross-sectional study was not designed to study process or clinical outcomes, including clinical slips or mistakes, and cannot be used to make causal inferences. Fourth, our study was not intended to identify the number and severity of discrepancies uncovered during medication reconciliation, which would be an important area for discovery in future research about EHR multitasking. Finally, as a qualitative study in a single safety net network, we are not able to develop theories based on particular provider or patient subgroup characteristics, and our findings may not be generalizable to other settings. The study strengths are its inclusion of primary care and specialty care providers, physicians and nurse practitioners, and a medically, socioeconomically, and linguistically diverse safety net population.

In summary, the multitasking clinician balances the cognitive and emotional demands posed by incoming information from multiple sources, attempts to synthesize and act on this information through EHR and communication tasks, and adopts strategies of silent EHR use and focused patient-clinician talk that may help mitigate the risks of multitasking. Future studies should explore diverse patient perspectives about clinician EHR multitasking, clinical outcomes related to EHR multitasking, and human factors and systems engineering interventions to improve the safety of EHR use during the complex process of medication reconciliation.

Acknowledgments

Research reported in this publication was supported by grants from the Agency for Healthcare Research and Quality (AHRQ; 1K08HS022561, P30HS023558, and R00HS022408). DS is supported by the Health Delivery Systems Center for Diabetes Translational Research funded through grant 2P30-DK092924 from the National Institute of Diabetes and Digestive and Kidney Diseases. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of AHRQ or National Institutes of Health.

Authors' Contributions

NR had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: GYM, DS, and NR; acquisition, analysis, or interpretation of data: GYM, ECK, CRL, DS, and NR; drafting of the manuscript: GYM and NR; critical revision of the manuscript for important intellectual content: GYM, ECK, CRL, DS, and NR; statistical analysis: GYM; obtained funding: NR; administrative, technical, or material support: GYM; study supervision: NR.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Codes associated with conceptual diagram: multitasking clinicians balancing the demands and risks of EHR and communication tasks during medication reconciliation.

[PNG File, 106KB - medinform_v6i2e10167_app1.png]

References

1. Ratanawongsa N, Barton JL, Schillinger D, Yelin EH, Hettema JE, Lum PJ. Ethnically diverse patients' perceptions of clinician computer use in a safety-net clinic. J Health Care Poor Underserved 2013 Nov;24(4):1542-1551 [FREE Full text] [doi: 10.1353/hpu.2013.0188] [Medline: 24185151]
2. The Joint Commission. National Patient Safety Goals Effective January 2018: Ambulatory Health Care Accreditation Program. 2018. URL: https://www.jointcommission.org/assets/1/6/NPSG_Chapter_AHC_Jan2018.pdf [accessed 2018-04-26] [WebCite Cache ID 6yyjnD0ba]
3. Rose AJ, Fischer SH, Paasche-Orlow MK. Beyond medication reconciliation: the correct medication list. JAMA 2017 May 23;317(20):2057-2058. [doi: 10.1001/jama.2017.4628] [Medline: 28426844]

4. Bayoumi I, Howard M, Holbrook AM, Schabort I. Interventions to improve medication reconciliation in primary care. Ann Pharmacother 2009 Oct;43(10):1667-1675. [doi: 10.1345/aph.1M059] [Medline: 19737997]
5. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
6. Persell SD, Osborn CY, Richard R, Skripkauskas S, Wolf MS. Limited health literacy is a barrier to medication reconciliation in ambulatory care. J Gen Intern Med 2007 Nov;22(11):1523-1526 [FREE Full text] [doi: 10.1007/s11606-007-0334-x] [Medline: 17786521]
7. Davis TC, Wolf MS, Bass PF, Thompson JA, Tilson HH, Neuberger M, et al. Literacy and misunderstanding prescription drug labels. Ann Intern Med 2006 Dec 19;145(12):887-894. [Medline: 17135578]
8. Wolf MS, Davis TC, Shrank W, Rapp DN, Bass PF, Connor UM, et al. To err is human: patient misinterpretations of prescription drug label instructions. Patient Educ Couns 2007 Aug;67(3):293-300. [doi: 10.1016/j.pec.2007.03.024] [Medline: 17587533]
9. Patel CH, Zimmerman KM, Fonda JR, Linsky A. Medication complexity, medication number, and their relationships to medication discrepancies. Ann Pharmacother 2016 Jul;50(7):534-540. [doi: 10.1177/1060028016647067] [Medline: 27147704]
10. Bedell SE, Jabbour S, Goldberg R, Glaser H, Gobble S, Young-Xu Y, et al. Discrepancies in the use of medications: their extent and predictors in an outpatient practice. Arch Intern Med 2000 Jul 24;160(14):2129-2134. [Medline: 10904455]
11. Centers for Medicare & Medicaid Services. Eligible Professional Meaningful Use Menu Set Measures: Measure 6 of 9. 2014 May. URL: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/7_Medication_Reconciliation.pdf [accessed 2018-04-26] [WebCite Cache ID 6yyjvQPb1]
12. Frankel R, Altschuler A, George S, Kinsman J, Jimison H, Robertson NR, et al. Effects of exam-room computing on clinician-patient communication: a longitudinal qualitative study. J Gen Intern Med 2005 Aug;20(8):677-682 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0163.x] [Medline: 16050873]
13. Alkureishi M, Lee W, Lyons M, Press VG, Imam S, Nkansah-Amankra A, et al. Impact of electronic medical record use on the patient-doctor relationship and communication: a systematic review. J Gen Intern Med 2016 May;31(5):548-560 [FREE Full text] [doi: 10.1007/s11606-015-3582-1] [Medline: 26786877]
14. Street RL, Liu L, Farber NJ, Chen Y, Calvitti A, Zuest D, et al. Provider interaction with the electronic health record: the effects on patient-centered communication in medical encounters. Patient Educ Couns 2014 Sep;96(3):315-319 [FREE Full text] [doi: 10.1016/j.pec.2014.05.004] [Medline: 24882086]
15. Ratanawongsa N, Barton JL, Lyles CR, Wu M, Yelin EH, Martinez D, et al. Computer use, language, and literacy in safety net clinic communication. J Am Med Inform Assoc 2017 Jan;24(1):106-112 [FREE Full text] [doi: 10.1093/jamia/ocw062] [Medline: 27274017]
16. Douglas HE, Raban MZ, Walter SR, Westbrook JI. Improving our understanding of multi-tasking in healthcare: drawing together the cognitive psychology and healthcare literature. Appl Ergon 2017 Mar;59(Pt A):45-55 [FREE Full text] [doi: 10.1016/j.apergo.2016.08.021] [Medline: 27890158]
17. Ratanawongsa N, Matta GY, Lyles CR, Koenig CJ, Barton JL, Yu K, et al. Multitasking and silent electronic health record use in ambulatory visits. JAMA Intern Med 2017 Sep 01;177(9):1382-1385. [doi: 10.1001/jamainternmed.2017.2668] [Medline: 28672379]
18. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. J Gen Intern Med 2009 Mar;24(3):341-348 [FREE Full text] [doi: 10.1007/s11606-008-0892-6] [Medline: 19130148]
19. Borycki E. Trends in health information technology safety: from technology-induced errors to current approaches for ensuring technology safety. Healthc Inform Res 2013 Jun;19(2):69-78 [FREE Full text] [doi: 10.4258/hir.2013.19.2.69] [Medline: 23882411]
20. Borycki EM, Kushniruk AW, Bellwood P, Brender J. Technology-induced errors. The current use of frameworks and models from the biomedical and life sciences literatures. Methods Inf Med 2012;51(2):95-103. [doi: 10.3414/ME11-02-0009] [Medline: 22101488]
21. Crampton NH, Reis S, Shachak A. Computers in the clinical encounter: a scoping review and thematic analysis. J Am Med Inform Assoc 2016 May;23(3):654-665. [doi: 10.1093/jamia/ocv178] [Medline: 26769911]
22. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Sep 6;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
23. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med 2017 Sep;15(5):419-426 [FREE Full text] [doi: 10.1370/afm.2121] [Medline: 28893811]
24. Hsiao C, Hing E, Socey T, Cai B. Electronic Medical Record / Electronic Health Record Systems of Office-Based Physicians: United States, 2009 and Preliminary 2010 State Estimates. Atlanta, GA: CDC National Center of Health Statistics; 2010.

25. Ratanawongsa N, Barton JL, Lyles CR, Wu M, Yelin EH, Martinez D, et al. Association between clinician computer use and communication with patients in safety-net clinics. JAMA Intern Med 2016 Jan;176(1):125-128 [FREE Full text] [doi: 10.1001/jamainternmed.2015.6186] [Medline: 26619393]
26. Karliner LS, Napoles-Springer AM, Schillinger D, Bibbins-Domingo K, Pérez-Stable EJ. Identification of limited English proficient patients in clinical care. J Gen Intern Med 2008 Oct;23(10):1555-1560 [FREE Full text] [doi: 10.1007/s11606-008-0693-y] [Medline: 18618200]
27. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008 May;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]
28. Sarkar U, Schillinger D, López A, Sudore R. Validation of self-reported health literacy questions among diverse English and Spanish-speaking populations. J Gen Intern Med 2011 Mar;26(3):265-271 [FREE Full text] [doi: 10.1007/s11606-010-1552-1] [Medline: 21057882]
29. San Francisco General Hospital. Administrative policy 16.35: medication reconciliation. In: SFGH Policies. San Francisco, CA: San Francisco Department of Public Health; Dec 2014.
30. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2nd Edition. Thousand Oaks, CA: Sage Publications; 1998.
31. Fernández A, Quan J, Moffet H, Parker MM, Schillinger D, Karter AJ. Adherence to newly prescribed diabetes medications among insured Latino and white patients with diabetes. JAMA Intern Med 2017 Mar 01;177(3):371-379 [FREE Full text] [doi: 10.1001/jamainternmed.2016.8653] [Medline: 28114642]
32. Harris LM, Dreyer BP, Mendelsohn AL, Bailey SC, Sanders LM, Wolf MS, et al. Liquid medication dosing errors by Hispanic parents: role of health literacy and English proficiency. Acad Pediatr 2017;17(4):403-410. [doi: 10.1016/j.acap.2016.10.001] [Medline: 28477800]
33. Barton JL, Schmajuk G, Trupin L, Graf J, Imboden J, Yelin EH, et al. Poor knowledge of methotrexate associated with older age and limited English-language proficiency in a diverse rheumatoid arthritis cohort. Arthritis Res Ther 2013;15(5):R157 [FREE Full text] [Medline: 24432366]
34. Schillinger D, Machtinger EL, Wang F, Palacios J, Rodriguez M, Bindman A. Language, literacy, and communication regarding medication in an anticoagulation clinic: a comparison of verbal vs. visual assessment. J Health Commun 2006 Nov;11(7):651-664. [doi: 10.1080/10810730600934500] [Medline: 17074733]
35. Karliner LS, Auerbach A, Nápoles A, Schillinger D, Nickleach D, Pérez-Stable EJ. Language barriers and understanding of hospital discharge instructions. Med Care 2012 Apr;50(4):283-289 [FREE Full text] [doi: 10.1097/MLR.0b013e318249c949] [Medline: 22411441]
36. Ratanawongsa N, Matta GY, Bohsali FB, Chisolm MS. Reducing misses and near misses related to multitasking on the electronic health record: observational study and qualitative analysis. JMIR Hum Factors 2018 Feb 06;5(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.9371] [Medline: 29410388]
37. Nuckols TK, Smith-Spangler C, Morton SC, Asch SM, Patel VM, Anderson LJ, et al. The effectiveness of computerized order entry at reducing preventable adverse drug events and medication errors in hospital settings: a systematic review and meta-analysis. Syst Rev 2014 Jun 04;3:56 [FREE Full text] [doi: 10.1186/2046-4053-3-56] [Medline: 24894078]
38. Prgomet M, Li L, Niazkhani Z, Georgiou A, Westbrook JI. Impact of commercial computerized provider order entry (CPOE) and clinical decision support systems (CDSSs) on medication errors, length of stay, and mortality in intensive care units: a systematic review and meta-analysis. J Am Med Inform Assoc 2017 Mar 01;24(2):413-422. [doi: 10.1093/jamia/ocw145] [Medline: 28395016]
39. Pohl JM, Tanner C, Hamilton A, Kaleba EO, Rachman FD, White J, et al. Medication safety after implementation of a commercial electronic health record system in five safety-net practices: a mixed methods approach. J Am Assoc Nurse Pract 2014 Aug;26(8):438-444. [doi: 10.1002/2327-6924.12089] [Medline: 24243832]
40. Frankel RM. Computers in the examination room. JAMA Intern Med 2016 Jan;176(1):128-129. [doi: 10.1001/jamainternmed.2015.6559] [Medline: 26619228]
41. Kumarapeli P, de Lusignan S. Using the computer in the clinical consultation; setting the stage, reviewing, recording, and taking actions: multi-channel video study. J Am Med Inform Assoc 2013 Jun;20(e1):e67-e75 [FREE Full text] [doi: 10.1136/amiajnl-2012-001081] [Medline: 23242763]
42. Lee WW, Alkureishi ML, Wroblewski KE, Farnan JM, Arora VM. Incorporating the human touch: piloting a curriculum for patient-centered electronic health record use. Med Educ Online 2017;22(1):1396171 [FREE Full text] [doi: 10.1080/10872981.2017.1396171] [Medline: 29103366]

Abbreviations

CMS: Centers for Medicare and Medicaid Services
EHR: electronic health record
IQR: interquartile range
PCP: primary care physician

Edited by G Eysenbach; submitted 18.02.18; peer-reviewed by A Georgiou, V Church, K Blondon; comments to author 15.03.18; revised version received 02.04.18; accepted 22.04.18; published 07.05.18.

Please cite as:
Matta GY, Khoong EC, Lyles CR, Schillinger D, Ratanawongsa N
Finding Meaning in Medication Reconciliation Using Electronic Health Records: Qualitative Analysis in Safety Net Primary and Specialty Care
JMIR Med Inform 2018;6(2):e10167
URL: http://medinform.jmir.org/2018/2/e10167/
doi:10.2196/10167
PMID:29735477

Original Paper

An Assistive Technology System that Provides Personalized Dressing Support for People Living with Dementia: Capability Study

Winslow Burleson[1], PhD; Cecil Lozano[2], PhD; Vijay Ravishankar[2], PhD; Jisoo Lee[2], PhD; Diane Mahoney[3], RN, PhD, FGSA, FAAN

[1]NYU-X, College of Nursing, New York University, New York, NY, United States

[2]Motivational Environments Research Group, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, United States

[3]School of Nursing, MGH Institute of Health Professions, Charlestown, MA, United States

Corresponding Author:
Winslow Burleson, PhD
NYU-X
College of Nursing
New York University
433 1st Ave
New York, NY, 10010
United States
Phone: 1 4802523929
Fax: 1 4802523929
Email: wb50@nyu.edu

Abstract

Background: Individuals living with advancing stages of dementia (persons with dementia, PWDs) or other cognitive disorders do not have the luxury of remembering how to perform basic day-to-day activities, which in turn makes them increasingly dependent on the assistance of caregivers. Dressing is one of the most common and stressful activities provided by caregivers because of its complexity and privacy challenges posed during the process.

Objective: In preparation for in-home trials with PWDs, the aim of this study was to develop and evaluate a prototype intelligent system, the DRESS prototype, to assess its ability to provide automated assistance with dressing that can afford independence and privacy to individual PWDs and potentially provide additional freedom to their caregivers (family members and professionals).

Methods: This laboratory study evaluated the DRESS prototype's capacity to detect dressing events. These events were engaged in by 11 healthy participants simulating common correct and incorrect dressing scenarios. The events ranged from donning a shirt and pants inside out or backwards to partial dressing—typical issues that challenge a PWD and their caregivers.

Results: A set of expected detections for correct dressing was prepared via video analysis of all participants' dressing behaviors. In the initial phases of donning either shirts or pants, the DRESS prototype missed only 4 out of 388 expected detections. The prototype's ability to recognize other missing detections varied across conditions. There were also some unexpected detections such as detection of the inside of a shirt as it was being put on. Throughout the study, detection of dressing events was adversely affected by the relatively smaller effective size of the markers at greater distances. Although the DRESS prototype incorrectly identified 10 of 22 cases for shirts, the prototype preformed significantly better for pants, incorrectly identifying only 5 of 22 cases. Further analyses identified opportunities to improve the DRESS prototype's reliability, including increasing the size of markers, minimizing garment folding or occlusions, and optimal positioning of participants with respect to the DRESS prototype.

Conclusions: This study demonstrates the ability to detect clothing orientation and position and infer current state of dressing using a combination of sensors, intelligent software, and barcode tracking. With improvements identified by this study, the DRESS prototype has the potential to provide a viable option to provide automated dressing support to assist PWDs in maintaining their independence and privacy, while potentially providing their caregivers with the much-needed respite.

(JMIR Med Inform 2018;6(2):e21) doi:10.2196/medinform.5587

KEYWORDS

Alzheimer disease; disorders, neurocognitive; image processing, computer-assisted

Introduction

Background

Dementia is a term that describes a broad category of symptoms related to declining memory and eroding thinking skills. It is a syndrome associated with a number of progressive illnesses affecting memory, thinking, behavior, and the ability to perform everyday activities [1]. An estimated two-thirds of dementia cases may be due to Alzheimer disease, the most common form of dementia [2]. The second most common cause is vascular dementia due to stroke. The World Health Organization Report estimates 7.7 million new cases of dementia are diagnosed every year [3] and the current estimate of cases to be 47.5 million. Over the next 40 years, an estimated 682 million people will live with dementia (persons with dementia, PWDs) [1]. It is important to note that the cognitive declines experienced because of dementia are not a normal part of the aging process. Beyond memory, dementia can affect communication and understanding, as well as the ability to focus and make decisions.

Dementia and cognitive decline make both basic and instrumental activities of daily living (ADL), such as bathing, dressing, eating, paying bills, cleaning, and doing laundry difficult and challenging for the individuals and their caregivers. Almost every person diagnosed with dementia eventually must either rely extensively on a caregiver (often a family member) at his or her home or relocate into a nursing home for professional supplemental care.

Data indicate that 86% of caregivers (either family members or professionals) help with ADL activities; the most common being dressing 61%, followed by feeding 52%, bathing 37%, toileting 34%, and incontinence care 26% [4]. Caregivers assisting individuals with dementia who perform ADL often feel stressed, frustrated at the amount of time required, and emotionally challenged. For individuals with early to middle stage dementia, dressing has been reported as the most pressing concern for both patients and caregivers [5]. Core challenges of dressing include the complexity of the activity and issues of privacy and independence of the PWD, particularly when caregivers are family members. Data indicate that adult children find it particularly challenging to help dress their parents, especially for opposite genders [6].

Our semiautonomous DRESS prototype was designed to address some of the challenges associated with dressing identified in literature and from interaction with focus groups composed of family and professional PWD caregivers. Integrating automated tracking and recognition with guided assistance for the dressing process, the DRESS prototype uses a combination of sensors and image recognition to detect dressing states and embedded intelligence to guide and prompt dressing tasks, assist in correction of dressing errors, and provide reinforcement for positive dressing performance. Initial input from caregiver focus groups provided a foundation for design and development of the prototype [6].

The goal was to provide assistance for the individual PWD to help them age in place more gracefully, while ideally allowing the caregiver to do other tasks as the PWD dresses, with the assurance that the prototype will monitor and alert when the dressing process is completed, or prompt if an intervention is needed. Although attempts have been made to automate real-time assistance for routine activities, we are not aware of other context-aware computing and human-computer-interaction projects that incorporate real-time prompting to assist with dressing processes.

The DRESS prototype leverages computer vision-based technologies to track and recognize progress during the dressing process. reacTIVision [7] is an image recognition system that the DRESS prototype uses to track fiducial markers [8] (a type of barcode, see Dressing Event Detection section) imprinted on clothing items (shirts and pants) to identify the type, location, and orientation of a garment. DRESS uses this data to recognize and track user progress while dressing and to determine whether the clothing is correctly positioned and oriented (ie, the front of the pants is facing forward). When a dressing error is detected, such as putting pants on backwards, the prototype generates an appropriate audio prompt recorded in the caregiver's voice to inform the PWD, noting the nature of the mistake and prompting recovery actions to correct the mistake. Once each step has been completed correctly, DRESS provides feedback and prompts the PWD to progress to the next step of the dressing activity. If the PWD continues to have problems—"freezing" or making little or no progress after multiple attempts or becoming frustrated—the caregiver is alerted so that they may provide personal assistance and support.

Before deploying the DRESS prototype in real-world environments with PWDs, we ran a capabilities study. Capabilities studies are employed in engineering to ensure that processes or products meet customer requirements, specifications, and functionality metrics. Our capabilities study involved 11 healthy participants emulating common dementia dressing scenarios, with a shirt and pants [9], to evaluate the DRESS prototype.

Prior Work

Many assistive technologies have been developed to help people perform daily activities. However, few systems specifically target the needs of PWDs. Mihailidis et al [10] developed a system for persons with moderate to severe dementia to assist with handwashing. Examples of other targeted behaviors include cooking [11] and taking medications [12]. Literature supports the use of cognitive interventions to assist and improve individual ability to perform ADL [9]. The current state-of-the-art in technology interventions involves attempts to mathematically predict patterns of behavior, but the results to date have been criticized as disappointing, as the predictive results are poor, and they are not sufficiently capable to be relevant to the design of systems that support the needs of PWDs [13-18].

Context-aware memory aids may have the potential to provide the support needed to assist with daily activities such as prompting to start dressing processes. However, memory aids alone are not sufficient—it is also important to determine context (current stage of dressing) to be able to create effective prompts. Efficient activity recognition systems are needed to acquire this information.

Usefulness of Vision Systems

Wu et al [19] presented an activity recognition system that combined radio-frequency identification (RFID) and video feedback in a kitchen setting. In testing for 16 activities with 33 subjects, they achieved a recognition rate of 80%. The system developed by Mihailidis et al [10] used video processing to recognize context and prompt actions performed in handwashing.

Behavioral Identification

An important contribution in the advancement of systems that identify ADL is Proact [20], a project designed to address recognition of 14 everyday activities. The system reports both the activity being performed and the extent to which it is performed. Related ADL research has been advanced by Dalton et al [21] and Fleury et al [22]. Dalton et al [21] developed a system that uses wireless kinematic sensors to identify accuracy of ADL identification based on position and on the manner of data processing. The authors reported dressing among the activities recognized. Fleury et al [22] developed a system using support vector machines and the data acquired from infrared sensors, microphones, door contact sensors, webcams, and accelerometers to recognize when a subject performs six types of daily activities, including dressing and undressing. Hayes et al [12] present another example of context-aware prompted feedback in an electronic pillbox that continuously monitors medication-taking over time.

Dementia Dressing Work

We found that little research emphasis has been placed on developing dressing support technologies to assist PWDs with this important ADL. An early effort in assisting PWDs in dressing was conducted by Namazi and Johnson [23]. They demonstrated how modifying closet arrangement, to organize the clothing in a visible and preplanned sequential order, can help independent activity.

Engelman et al [24] showed how human prompting using graduated procedures can be used to increase dressing independence for PWDs. Popleteev and Mayora [25] developed a smart assistive buttoning system for people with mild cognitive decline. Their system detects whether a button is “locked” with its correct counterpart and if incorrect (unlocked or locked with a wrong counterpart), triggers an event, and the system provides an alert (verbal feedback) to the user and records event details for further caregiver analysis.

Matic et al [26] developed an RFID and video system that detects dressing activity failures. Their system identifies the most common dressing failures, which are as follows: (1) putting clothes in an incorrect order, (2) putting on clothes partially, (3) incorrect orientation (such as putting on clothes backwards), (4) dressing incorrectly for temperature (too little or to many layers of clothing), and (5) putting clothes on the wrong part of the body. In addition to identifying errors, Matic’s system also recognizes when the correct dressing performance has occurred. However, this system does not use this information to provide feedback or to assist in rectifying mistakes identified during dressing.

Although each of these systems provides significant contributions, none are comprehensive in a manner that addresses the entire process from monitoring dressing activity, to identifying correct dressing and dressing failures, to providing feedback and guidance that rectify mistakes.

An important and challenging feature missing in existing systems is the ability to tailor or customize feedback and support interventions that address varying levels of cognitive function in an individual PWD. Cognitive function can progressively decline over time at various rates fluctuating throughout the day or over time as the system is used (eg, weeks to months or years) [9,13].

Literature and caregiver focus groups [27], consisting of 25 Latino (family member and professional) caregivers of PWDs, in three Arizona community service sites serving Latinos, clearly indicate an effective dressing assistance system for people with dementia should aim to be: (1) unobtrusive, (2) automated, (3) context-aware (ie, having the ability to recognize actions performed or missed), and (4) capable of providing personally tailored feedback and assistance as needed. The DRESS prototype was developed with attention to each of these criteria. The results, discussed below, indicate the potential of automated systems to assist in actual independent and assisted living settings.

The DRESS Prototype

The transdisciplinary team that developed the DRESS prototype included experts from engineering, gerontology, social science, psychology, nursing, speech, and occupational rehabilitation. The DRESS prototype was advanced through an iterative user-centered design approach, integrating participation and involvement of caregiver teams and family members of PWDs in the process of problem definition, design, and development [9] (Figures 1 and 2). Design was also informed by the Alzheimer’s Association’s dressing guidelines. The DRESS prototype uses our Game as Life-Life as Game (Figure 3) ubiquitous computing platform [9] that integrates a variety of sensors with digital systems, databases, and interactive software scripting to analyze context, then to create and provide prompts and interactions to support PWDs and caregiver’s dressing-related activities.

The intent of the DRESS prototype is to integrate typical routines and humanized interactions, promote normalcy and safety, and facilitate flexible customization to guide PWDs through dressing processes [9]. Continuous collection of data offers the DRESS prototype unique opportunities to not only provide real-time guidance and feedback but also to record, analyze, and understand patterns of usage to enhance development of appropriate interventions. We focused on tailoring and customizing feedback to suit each user’s need and

also recognized the progressive nature of dementia so that the prototype focuses on the individual, not on the disease. The DRESS prototype is further customizable to use a recording of the caregiver's voice to prompt the PWD.

The DRESS prototype also analyzes the marker data to determine which portion of the clothing is facing the dresser. Caregiver focus groups [9] indicated that this level of privacy was acceptable to (both professional and family) caregivers, but they recommended that future versions of the DRESS prototype use cameras embedded into the surface of the dresser, such that their visibility to PWDs would be minimized. Likewise, in future DRESS prototypes, we plan to make the coded markers (which are currently large, awkward, and would likely be stigmatizing, if used in public) "invisible," either through the use of infrared ink or through machine recognition and training of patterns in PWDs' existing clothes (see Discussion, for additional information).

Figure 1. Typical human interaction with the DRESS prototype.(9, reprinted with permission).

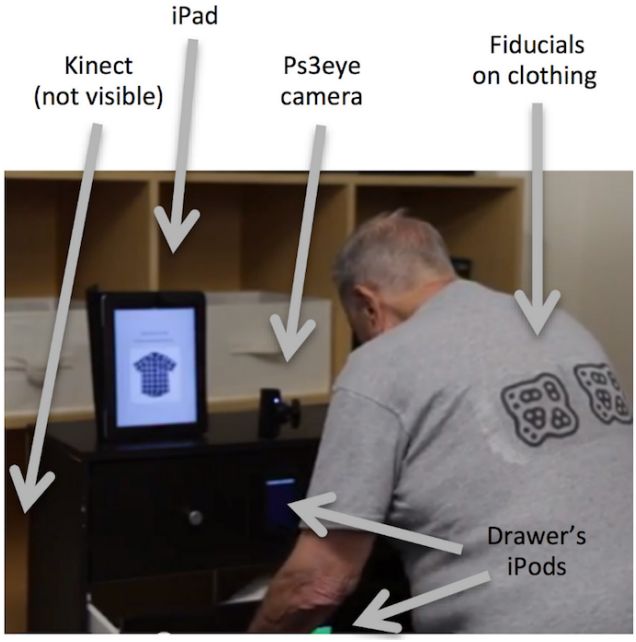


Figure 2. DRESS prototype initial architecture hardware and infrastructure; Kinect was subsequently removed.

Figure 3. Game as Life—Life as Game system architecture including physical components (lighting using X10 electrical protocol, wireless audio using AirTunes and an 802.11 wireless network, and door window sensor using X10 radio frequency) above the horizon and digital components (Java server, web, database logs, and scripts).

DRESS Functionality Aiding Dressing

The DRESS scenario begins with a caregiver assisting the PWD, who is wearing underwear and the wrist or leg skin conductance sensor, positioned in front of the dresser. Following the Alzheimer's Association's dressing guidelines for logical ordering and simplification of clothing choices, the five-drawer dresser is organized with one piece of clothing per drawer, with the first clothing item in the top drawer and the remaining clothing items sequenced in drawers below. The caregiver then initiates the assistive dressing procedure via their mobile device and leaves the room.

An X10 motion sensor on top of the dresser senses that the PWD is close to the dresser and transmits status to the Indigo server. Once the DRESS prototype confirms the presence of the PWD, the individual receives a verbal prompt to open the top drawer, and the iPod Touch on the front of the drawer displays a green light prompt. The other drawers display a red light. If the PWD opens the wrong drawer, the DRESS prototype prompts the PWD to close it and open the drawer with the green light (ie, the correct drawer).

Once the PWD opens the correct drawer and removes the clothing item, the RFID reader inside the drawer detects movement of the tag attached to the clothing. When this occurs, the DRESS prototype initiates a sequence, beginning with an Indigo action command (developed in the Processing language [7]) for the open source reacTIVision computer vision software to receive fiducial marker data from the cameras. reacTIVision provides information about the orientation and distance of the clothing-based fiducial markers. The Processing program initially used the fiducial marker data in combination with skeletal position data from Kinect for Windows [28] (responsible for skeletal tracking) to identify the current state of dressing, assess the need for intervention, and to provide audio prompts and assistance if needed. As the reacTIVision system captured the most important elements of clothing orientation, it was

determined that Kinect was not needed and that the DRESS prototype would be simpler without this component.

If caregiver selects "continuous" mode (see Figure 4, middle) or when the DRESS prototype determines the need for continuous intervention, chronological directions for each step of the dressing process are provided (eg, "put one arm through one hole of the t-shirt" followed by "Now, put the other arm through the other hole."). If the clothing is sensed as incorrectly worn (eg, inside out, back to front, and shirt Velcro misaligned) or the PWD is not taking any actions, the DRESS prototype identifies this state and guides the PWD through the process of correcting the error through a series of audio and visual prompts.

The DRESS prototype then continues to monitor, sense, and correct the PWD until the dressing process is completed. To enhance autonomy and independence of the PWD, the DRESS prototype does not provide audio prompts when an article of clothing is donned correctly in the "independent" mode. The goal is for the DRESS prototype to personally tailor support, at a level commensurate with the PWD's varying moment-to-moment and day-to-day needs.

Once the first clothing item is identified as being worn correctly, the DRESS prototype asks the PWD to close the drawer. Upon confirmation of drawer closure, the Indigo server initiates the next step in the dressing sequence, repeating a similar sequence of actions for each item of clothing in the remaining drawers until dressing is completed.

If the PWD becomes "stuck" (a nonoptimal experience state, eg, the user becomes confused or loses interest or focus) [29], as determined by a combination of the skin conductance level and context of recent sequence of problems in dressing activity, motivational prompting is provided to reengage the PWD.

The DRESS prototype continually monitors PWD stress levels via the skin conductance sensor, which is coupled with action monitoring to track progress and tailor interventions to mitigate frustration. If PWD stress levels continue to increase, the DRESS prototype initiates an activity previously identified by the caregiver as soothing, such as playing a favorite song or video clip. If this intervention is unsuccessful and stress levels continue to rise, the DRESS prototype notifies the caregiver via preferred communication (ie, cell phone or email). The goal is to avoid what caregivers term a "meltdown" situation.

The DRESS prototype continues to monitor progress, provide guided prompts and intervention until the dressing process is completed and then notifies the caregiver.

Dressing Event Detection

The fiducial tracking system uses a ps3eye camera to capture the current state of the shirt dressing process by detecting the ID and position of the specially designed 2D bar code fiducial markers attached to the clothing (Figure 5). The reacTIVision software extracts marker identification information from the cameras (identifiers indicate that the system detected the markers; see Figure 5, right), sensing tangible user interface object messages via UDP 3333 to the tangible user interface object–enabled client application [8]. This software determines (1) the dressing condition scenario based on the marker position with respect to the clothing, (2) its orientation and relation with other markers, (3) the time the marker is detected, and (4) the context of the dressing process. Table 1 shows the event detection and the rules used for the process of donning a shirt and a pair of pants.

The reacTIVision recognition process begins by searching for any marker within the camera field of view. reacTIVision uses marker data to determine orientation and placement of the garment. For example, sensing the back or incorrect positioning of the shirt prompts the PWD to reorient the shirt. Sensing the left or right side of the shirt marker can determine whether the user has put one side of the shirt on correctly (F and R/L in Table 1). Subsequent sensing of the markers on the opposite side (L/R) can indicate that both arms are worn (A). If the markers indicating reorientation of progress have not been sensed for more than 5 seconds, the DRESS prototype interprets this condition as partial dressing (p) and provides additional guidance and prompts.

Figure 4. Caregiver iOS mobile application user interface.

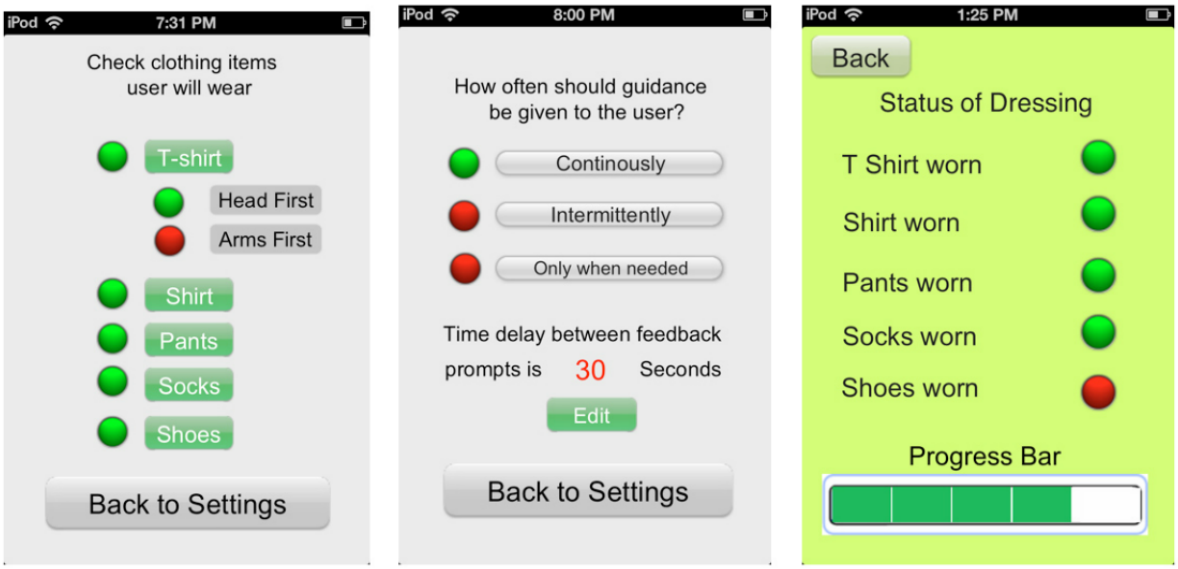


Figure 5. Fiducial markers (left) and example of fiducial markers (right) provided by a view of the ps3eye camera showing the markers from the shirt detected by the reacTIVision system.

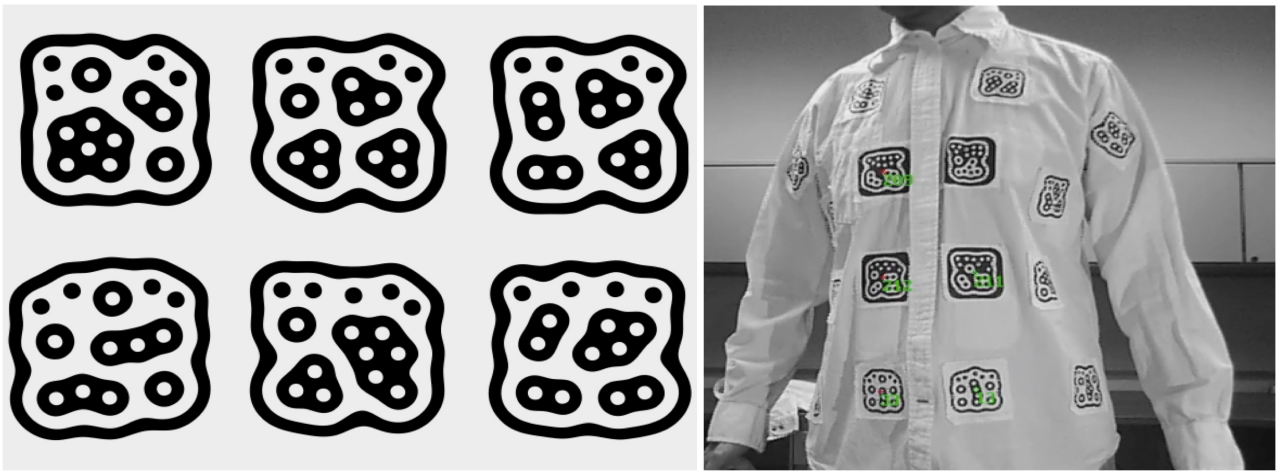


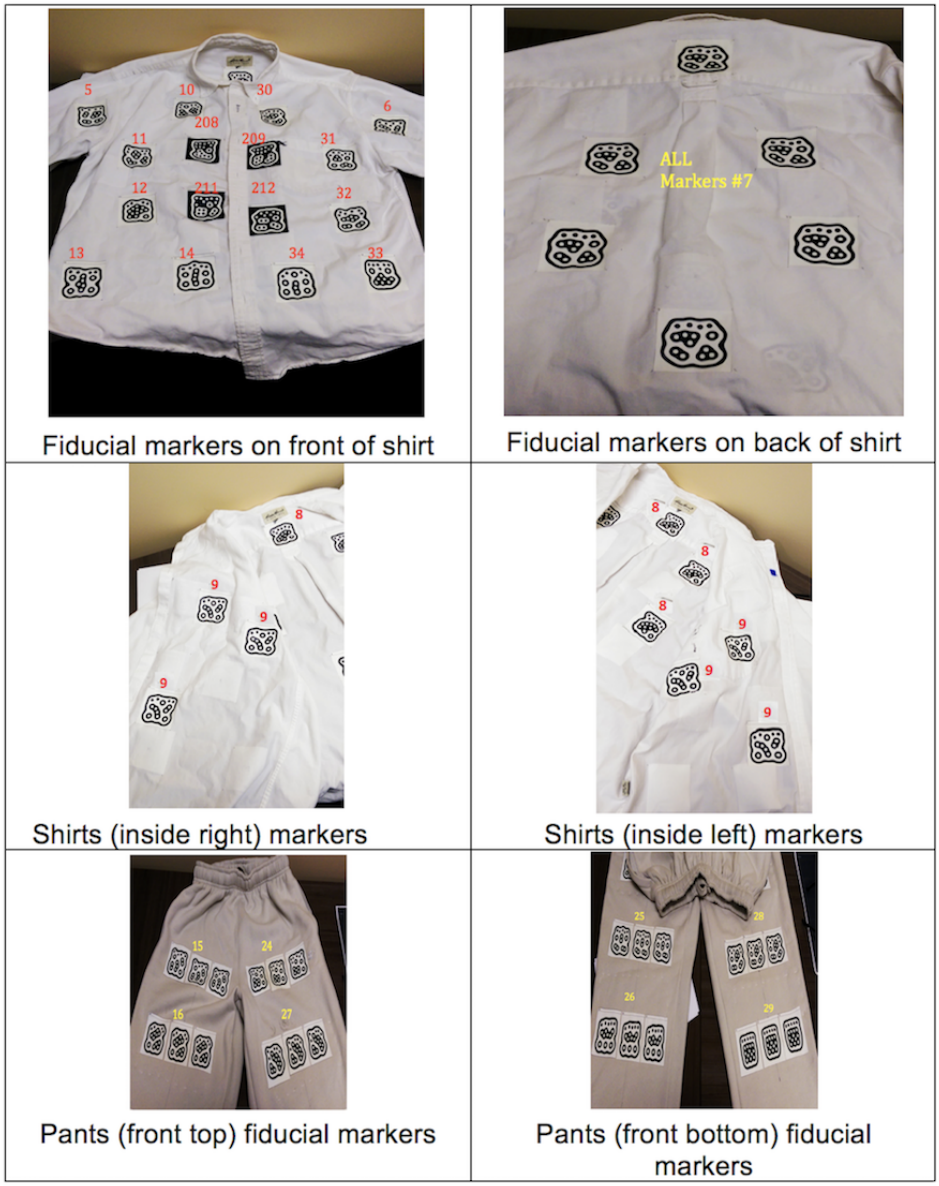
Table 1. Detection event descriptions and identification rules.

Clothing item[a]	Detection description	Identification rules
Shirt		
F	Front side of the shirt	Any marker from front of shirt (5,6,10-14,30-34,208,209,211,212) is visible for 2+ seconds.
B	Back side of the shirt	Any marker from back of shirt (7) is visible.
I	Inside part of the shirt	Any marker from inside of shirt (8,9) is visible for 2+ seconds. Marker 8 is for the center part of inside, whereas marker 9 is for the sides (left and right) of the inside part.
R	Right arm of the shirt worn	Any markers from front right of shirt (5,10,11,12,13,14,208,211) are visible for 2+ seconds.
L	Left arm of the shirt worn	Any markers from front left of shirt (6,30,31,32,33,34,209,212) are visible for 2+ seconds.
A	Both arms of the shirt worn	Any one marker from detecting R and one for detecting L is visible for 2+ seconds.
M	Velcro unevenly fastened	Any one of the following absolute differences of markers' distances are true: |Y208-Y209| >.05, |Y211-Y212| >.05, |X208-X209| >.18, or |X211-X212| >.18.
p	Partial dressing (incomplete)	Any one of the markers from either L or R is visible and the other is not visible for more than 5 seconds.
C	Shirt worn correctly	All the following absolute differences of markers' distances are true: |Y208-Y209| <.05, |Y211-Y212| <.05, |X208-X209| <.18, and |X211-X212| <.18.
Pants		
B	Back side of the pants	Any marker from back of pants (22) is visible for 2+ seconds.
I	Inside part of the pants	Any marker from inside of pants (17,19) is visible for 2+ seconds. Marker 17 is for inside out and front side, and marker 19 is for inside out and back side.
L	Left leg of the pants worn	Any markers from low part of the left side of the pants (28,29) are visible for 2+ seconds.
R	Right leg of the pants worn	Any markers from low part of the right side of the pants (25,26) are visible for 2+ seconds.
p	Partial dressing (incomplete)	Any one of the markers from either L or R is visible and the other is not visible for more than 5 seconds.
C	Pants worn correctly	All markers from upper part of the front of pants (15,16,24,27) are visible for 2+ seconds.

[a]The letters in this column indicate the ID.

To verify that the shirt is closed and worn correctly (C), the DRESS prototype searches for four markers (208/209 and 211/212) placed near the Velcro (Figure 6, top left). The close proximity between the right matching markers of both sides of the shirt and their orientation is used to identify any misalignment and generate corrective prompts. The proximity threshold between matching markers is fixed and previously determined by testing a correctly worn shirt. If the alignment conditions are not met, the DRESS prototype indicates a misalignment error (M) and prompts corrective action.

Figure 6. Fiducial markers provide DRESS with unique identification, orientation, and distance information on (a) shirt (upper row); (b) inside right and inside left markers (middle row); and (c) pants (lower row).



Dressing errors such as wearing the shirt back to front or inside out are identified if markers are sensed in the wrong sequence or orientation. For example, markers 7, 8, and 9 are attached to the back (B) and inside parts of the shirts (I) to identify the correspondent errors cases (see Figure 6, top row, right, and middle row). During laboratory testing, more than one marker of the same orientation ID was placed on each part of the clothing to assess validity and increase the robustness of the DRESS prototype's marker detection process.

Testing indicated that while donning the shirt, markers on the back or the inside can inadvertently become sensed. To minimize inaccurate detection errors, the DRESS prototype requires each marker to be continuously visible for a period of time before initiating action. On the basis of observations from our preliminary study [9], a 3-second duration appeared to be adequate to minimize these inaccurate detections.

Once the shirt donning has been successfully completed, the DRESS prototype prompts the PWD to close the shirt drawer, then provides audio and visual cues (green light on the associated iPod) to proceed to the pants drawer. For the pants dressing process, the DRESS prototype uses only the ps3eye camera located in the middle of the dresser. The DRESS prototype begins searching for the fiducial markers on the bottom half of the pants (25/26 and 28/29; see bottom right of Figure 6) to determine orientation and position. Similar to the process used to determine shirt orientation, the DRESS prototype then searches for markers indicating that one leg is worn (R/L) and then the other leg marker (L/R) is worn.

Partial or error detection follows the same protocol as the shirt. To identify whether the user has indeed stood, pulled the pants up, and worn them correctly (C), the DRESS prototype looks for markers in the upper half of the pants, specifically for marker pairs 15/16 and 24/27 (Figure 6, bottom left). This pant detection pattern was based on preliminary study observations of donning the pants while seated (PWDs are generally encouraged to don pants while seated, for safety, to minimize falls). As in the

example of the shirt, errors like reversal back to front (B) or inside out (I) are detected with markers 22 and 17/19 attached to these parts of the clothing, and appropriate corrective guidance and prompts are generated for the PWD.

Once the bottom markers of both legs have been detected, the DRESS prototype assumes that the user has donned both the legs of the pants correctly and is about to stand up. Supportive feedback is given and the caregiver is notified that the dressing process has been completed.

Methods

Aims

A laboratory study was conducted to evaluate the detection capabilities of the DRESS prototype. The study and recruitment of subjects was conducted according to the ASU IRB protocol STUDY1110006934. No conflicts of interest were noted. As the study was not a clinical trial, it was not registered with Clinicaltrials.gov.

Our overall aim for the capabilities study was to assess the DRESS prototype's ability to detect correct and incorrect dressing events. Accurate detection of these events is critical to developing effective sociotechnical systems that support a PWD's dressing activities. Although verification of our findings will be necessary with PWDs in home-based settings, home trials were not feasible during this phase of research. As the primary difference between a healthy adult and a PWD is a cognitive difference, the recruitment of healthy adults, of any age, was considered reasonable for this study.

To ensure the appropriateness of the prototype for PWDs, we built upon our prior work, supported by the Alzheimer's Association, involving caregiver focus groups [6], with the aim of advancing the prototype's ability to tailor customized support and feedback to the challenges that PWDs encounter as their conditions progress. We followed caregiver recommendations to engage adult participants in a laboratory setting to produce "acted errors" during DRESS prototype development and testing to assess the DRESS prototype's ability to accurately detect orientation, position, and errors.

To mitigate privacy and sensitivity issues related to dressing, study participant dressing actions were performed by donning fiducially-enhanced clothes on top of existing tee-shirt and athletic shorts. This study focused only on fiducial detection, and other sensor data were not included. We chose to test two clothing items: shirt and pants because of the requirement for different levels of motion and dexterity for both the top and bottom of the body. In addition, both clothing items are commonly worn by both women and men.

Study Design

The study was conducted with 11 healthy young participants, engaged for 1-hour sessions (7 female and 4 male, aged 19-41years—average age 25 years). Given the innovative nature of the prototype, we found that there was sufficient participant interest in the study to proceed effectively without the need for compensation. At the beginning of the session, participants gave informed consent, filled out a survey instrument, and were informed of the research goal. Pre- and postsurveys contained questions related to common dressing practices (eg, How often they use the specific clothing item? How often they put the clothing the wrong way? and if any, what have been the most common dressing mistakes they had made?) and reporting any discomfort about both the experimental setting and the tasks.

After the presurvey, participants were introduced to the DRESS prototype, highlighting the location of the fiducial tracking cameras and sensors and the preferred space between the dresser and the chair to use while getting dressed. Subjects were also asked to emulate common target population dressing practice and to use the chair while donning pants for the different dressing scenario study conditions. Finally, subjects were shown the special characteristics of the clothing items to be worn, especially the fact that they had to correctly align the shirt's Velcro closure without buttons or button holes to assist with alignment (Figure 7).

We were specifically interested in observing the ability of the prototype to use fiducial markers to accurately detect different stages of the process for nine dressing scenarios common to PWDs [22], clothing worn: correctly (shirt and pants); partially or on one limb, that is, one arm worn or one leg worn (shirt and pants); backwards with the back in front (shirt and pants); inside out (shirt and pants); and misaligned (for shirt only).

Before receiving specific instructions on how to perform each of the dressing conditions outlined in the study, participants were asked to retrieve two pieces of clothing from the drawer and to put them on in the manner they would normally perform at home. This step was included to identify whether any dramatic differences were identified because of the constraints of the dressing scenario instructions (not found to be the case) and for future analysis of the natural dressing process. As participants performed this "natural dressing" task, the chair was available to them, but none chose to use it.

Initially, the study incorporated a Kinect for event detection. Preliminary study determined that the Kinect could not reliably track participants' skeletal actions because of occlusion by the garment during dressing. Use of the fiducial marker tracking was found to offer more effective tracking of garment orientation and position during dressing and was selected to both simplify and increase the reliability of garment recognition to assist the DRESS prototype in developing prompting and interventions to guide dressing.

Participants were then given the dressing conditions to be performed. The experimenter explained each step in the process using pictures (Figure 7). Pictures of the target dressing conditions were used instead of video (or specific descriptions of intermediate procedures) to minimize introduction of experimenter bias.

Figure 7. Example pictures of dressing conditions shown to the participants, “Shirt misaligned” (left), “Pants partial” (right).

Participants were instructed that each trial would consist of the following steps: (1) wait for the experimenter to cue what dressing condition to perform and when to start; (2) pick the appropriate clothing item from the drawer; (3) don the garment in the way prescribed for the condition; (4) once completed, wait 3 seconds standing in front of the dresser saying slowly “DONE 1 2 3”; (5); if the condition involved an acted error to then engage in the DRESS prompted corrections; and (6) wait 3 seconds, again, saying slowly “DONE 1 2 3,” once the prompted dressing activity had been completed.

After completion, when prompted, the participant removed the clothing item off and gave it back to the experimenter. The experimenter would then orient the clothing (correctly, inside out, etc, as appropriate) and place the clothing back in the respective drawer to prepare for the next trial. Participants were reminded to complete each trial by wearing the clothing correctly in the manner they would normally prepare to leave their home and to use the chair when putting on or readjusting the pants. All participants performed each of the nine dressing conditions twice following a complete randomized block design. Finally, each participant session ended with the postsession survey.

Once we collected the data, we analyzed the DRESS prototype’s detection performance by comparing Indigo server recorded detection events (Table 1) with the expected detections for each of the six identified phases of the dressing process in each condition (first 2 title rows in Table 2). Phases of participants’ dressing actions included (1) Adjusting the clothing after being worn incorrectly spontaneously or by design (conditions with acted errors), (2) Putting the first limb on, (3) Transitioning between limbs, (4) Putting second limb on, (5) Transitioning to completion and adjustment, and (6) Completing the correct dressing process by standing in front of the camera.

Computer recordings of the trials were visually inspected to identify reasons for any missing or incorrect detections. Screen recordings included the video of the ps3view camera showing the participant’s action and the fiducial detection information (visible marker when detected; Figure 5, right), the experimenter’s actions (eg, setting, starting, and stopping the trial condition), and the computer’s time stamp.

Due to technical difficulties recovering data files for two trials, data were available for analysis for 108 of 110 experimental trials for the shirts and 86 of 88 trials for the pants.

Results

DRESS Prototype

Results indicated that the DRESS prototype was most reliable at reporting expected detections for acted errors of inside out pants and shirt in phase 1, followed by detection of each limb worn in phases 2 and 4 for both clothing items—missing only 4 out of 388 expected detections (see Table 2,Figures 8 and 9). The ability to recognize other missing detections varied across conditions. Furthermore, the DRESS prototype identified several initially unexpected detections, for example, recognizing the inside of the shirt or back of the pants during transition phases. On the basis of video analysis, the plausible detections that occurred most often among participants were identified as “expected detections” for each phase (presented in parenthesis, Table 2).

Detection of Shirt

When participants were asked to wear the shirt correctly, we expected that the DRESS prototype would sporadically detect and record the inside (I) or the back of the shirt (B) before putting it on as it was removed from the drawer. We expected the front of the shirt (F) to be detected in the first phase, then, when one arm is worn (R/L in phase 2), when the other arm is worn (L/R in phase 4), to detect wearing of both arms (A), and finally correctly completing the dressing process by aligning and attaching the Velcro (C).

In the transition phases, we expected to also see the back of the shirt (B) or the inside (I) when the shirt was folded or moved around or above the person while dressing. Between donning the first and second arm, we expected to detect partial dressing (p) even at times when this was not the condition. Finally, we

expected some misalignment (M) detections while adjusting preceding completion.

Table 2. Detections for each of the six identified phases for pants and shirt conditions. Letters indicate detection labels, and parentheses are used to indicate the expected detection events. Italicized fonts indicate unexpected detection results where the number shows the frequency (with positive and negative representing undesirable and missing expected detections, respectively).

Dressing Errors	Dressing phases					
	1	2	3	4	5	6
	Preliminary error or adjustment	1st limb worn	1st limb to 2nd limb transition	2nd limb worn	2nd limb to completion transition	Correct completion
Pants						
Correct	(B[a])	R[b]/L[c]	—	L/R	—	*−4C*[d]
Back to front	*−3B*	R/L	*4B*	L/*−1R*	—	*−6C*
Inside-out	I[e],(B)	R/L	*1B*	*−1L*/R	—	*−2C*
Partial	(B)	R/L	—	*−2L*/R	*1B*	*−7C*
Shirt						
Correct	F[f],(B),(I)	R/L	(p[g]),(I),(R/L)	L/R	A[h],(M[i]),(I),*(1p)*	*−10C*
Back to front	F,*−6B*,(I)	R/L	(p),(I),(R/L)	L/R	A,(M),(I),*(3p)*	*−10C*
Inside-out	F,(B),I	R/L	(p),(R/L)	L/R	A,(M),(I)	*−10C*
Misaligned	F,(B),(I)	R/L	(p),(I),(R/L)	L/R	A,*−7M*,(I)	*−5C*
Partial	F,(B),(I)	R/L	*−5p*,(I),(R/L)	L/R	A,(M),(I),*(4p)*	*−11C*

[a]B: detection label for back side.

[b]R: detection label for right arm/leg.

[c]L: detection label for left arm/leg.

[d]C: detection label for worn correctly.

[e]I: detection label for inside part.

[f]F: detection label for front side.

[g]p: detection label for partial dressing.

[h]A: detection label for both arms of the shirt.

[i]M: detection label for Velcro unevenly fastened or misaligned for shirt.

Figure 8. Shirt detection events indicate the number of trials (y-axis, maximum 108) for each recorded detection (x-axis), for each of the 5 conditions indicated by the filling pattern. Detection labels: I for Inside of clothing; B for back to front; BB+ for back to front detected more than twice in a row; IR+ or RI+ for sequence of inside then right or right and then inside detected more than once in a row; L for left side; R for right side; p for partial dressing; pI+ or pR+ for sequence of partial and inside or partial and right detected more than once in a row; C for correctly worn. Detections are presented for each of the 6 phases indicated on the top of the graph: (1) Preliminary error or adjustment; (2) 1st limb worn; (3) 1st limb to 2nd limb transition; (4) 2nd limb worn; (5) 2nd limb to completion transition; and (6) Correct completion.

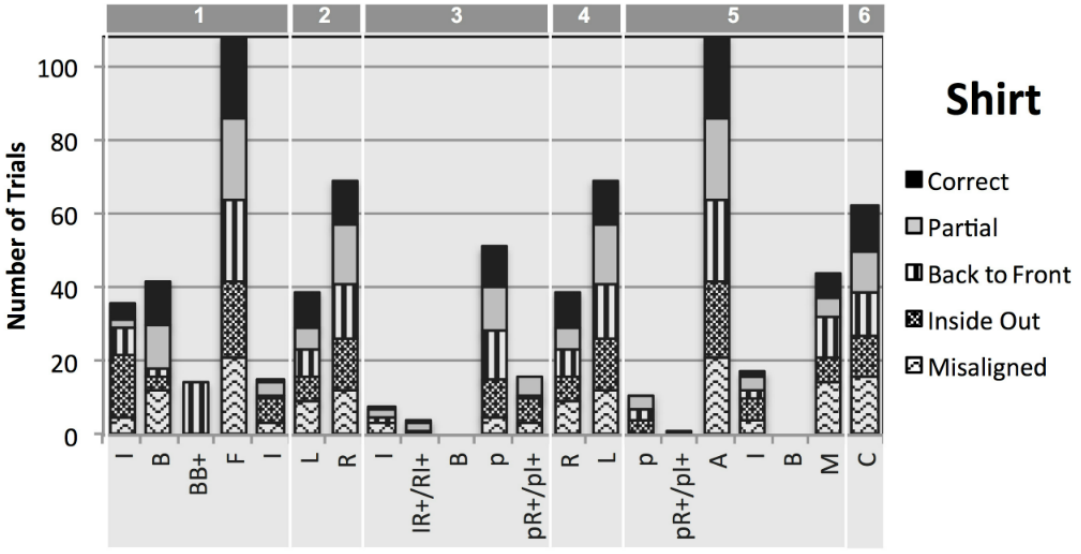
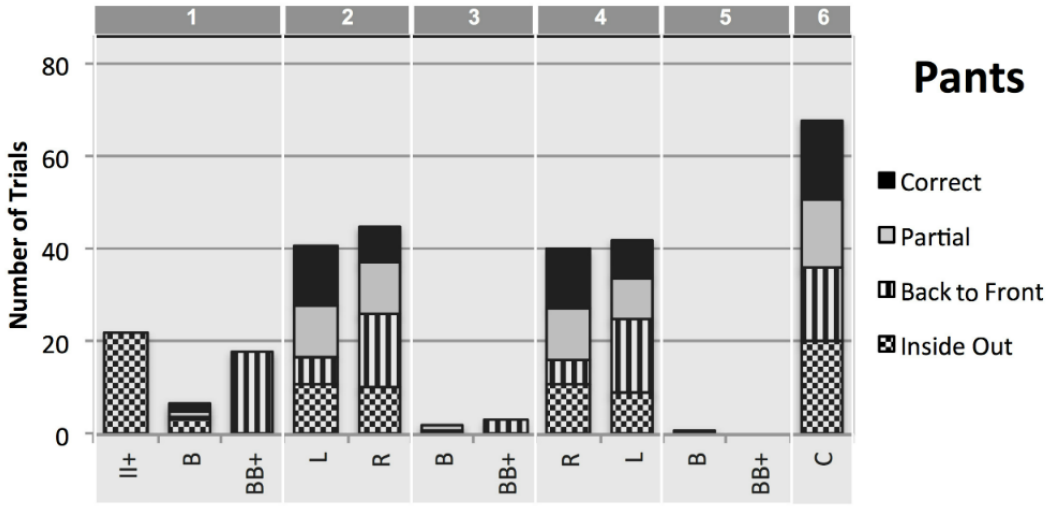


Figure 9. Pants detection events indicate the number of trials (y-axis, maximum 86) for each recorded detection (x-axis) for each of the 4 conditions indicated by the filling pattern. Detection labels: II+ for inside of clothing detected more than twice in a row; B for back to front; BB+ for back to front detected more than twice in a row; F for front side; L for left side; R for right side; C for correctly worn. Detections are presented for each of the 6 phases indicated on the top of the graph: (1) Preliminary error or adjustment; (2) 1st limb worn; (3) 1st limb to 2nd limb transition; (4) 2nd limb worn; (5) 2nd limb to completion transition; and (6) Correct completion.



We found that the DRESS prototype was completely reliable at detecting initial donning of the shirt (F) and was likewise completely reliable in phases 2 and 4 when the participant put on each of the two arms. Figure 9 shows the detections expected for each condition in each phase of the testing, and the black bars show the counts of detections in the correct condition. The DRESS prototype missed correct dressing completion detections in 10 of 22 cases. Reasons included limitations in the detection of markers, but also included the threshold of Velcro markers being too small (eg, if participants moved too far from the camera) and when participants did not correct the Velcro misalignment even when prompted to do so. Providing a mirror to check correctness and alignment might successfully alleviate this issue.

With respect to shirt error conditions, the DRESS prototype was most reliable in detecting inside out errors, missing only a single detection. In one occasion, the participant appeared confused about the orientation of the shirt and turned it inside out several times before completing donning, resulting in an unexpected but accurate condition recognition. Other conditions resulted in similar detection reliability, missing 5, 6, and 7 detections for partial, back to front, and misalignment conditions, respectively (see Table 2).

Unexpected detections included partial (P) detections after completing wearing of the second arm just before completion of donning (phase 5). Video inspection revealed that the unexpected detections were primarily because of lengthy adjustments by the subjects before closing the shirt. Adjustments included opening and closing the shirt several times to bring the two parts of the shirt together; holding the neck or Velcro occluding the markers; and for females, adjusting hair and occluding the markers while completing the dressing process. In one case, the shirt was too large for the participant, resulting in folds that impeded marker detection.

Detection of Pants

When participants were asked to don pants while seated, we expected the following detections through the phases of the process: sporadic back of pants detection when adjusting before putting the pants on (B in phase 1); right or left leg worn (R/L in phase 2); then other leg worn (L/R in phase 4); and finally that donning pants was correctly completed (C in phase 6). The DRESS prototype successfully recognized all the detections except phase 6, where completion detection was missed 5 out of 22 trials (see Table 2 and black bars in Figure 9).

In terms of detecting acted errors in phase 1, the DRESS prototype was most reliable in detecting when pants were donned inside out (100%) and only missed three back to front detections (see phase 1 of Table 2 and stripes and dotted bars in Figure 9). For the partial dressing condition, the DRESS prototype failed to correctly record partial dressing, Video examination indicated that partial dressing was detected in some occasions when the middle to upper fiducials on one pant leg were detected while the fiducials of the other pants leg were not. Data inspection indicated that no completed detections were recorded for only partial dressing. However, because of folding while pants are partially worn during donning, detection of partial dressing remains challenging.

Missing detections were found to occur as subjects put on the second leg (L/R phase 4) and upon correct completion of the dressing process (C phase 6). Visual inspection of the videos indicated that missing detections were because of (1) Inability of the camera to see markers, (2) Suboptimal position of the participant with respect to the camera (tilted, too close, too short), (3) Occlusion by folds in the cloth when the clothes were too large for the participant, (4) Failure of the Indigo server to recognize the visible markers or record the detection events on time, and (5) Participant donning the clothing too quickly for the marker to be captured and recognized.

Unexpectedly, the DRESS prototype detected back to front orientation in six trials during transition phases during donning

(3 and 5). Several events related to clothing adjustment were found to account for these detections, including (1) Turning the pants around while wearing one leg; (2) Holding the pants in front of the camera long enough in an orientation that the DRESS prototype incorrectly detected wearing one leg, affecting the interpretation of the following detections; (3) Taking the pants off after completing the partial scenario; (4) Readjusting the pants in a manner in which the back markers were visible momentarily; and (5) While adjusting the pants after turning them inside out in the inside-out dressing condition.

Discussion

The main findings of this capabilities study are that the DRESS prototype incorrectly identified 10 of 22 cases for shirts and only 5 of 22 cases for pants. Through this process, we identified several significant opportunities to improve the DRESS prototype's reliability, such that it can provide substantive support for PWDs engaged in dressing activities in subsequent in-home trials.

Intelligent dressing systems that support PWD need to understand and adapt to the complex and dynamic processes involved in donning each clothing item. These systems must also be able to adapt to challenges that may occur in these processes that can lead to incorrect dressing.

The DRESS prototype identifies states within the dressing process through its ability to detect fiducial markers on clothing items. Marker detection is used to infer context and improve the accuracy of the DRESS prototype. Additional trials are needed to further determine optimal positions for marker placements to minimize potential interference and to ensure that the DRESS prototype can accurately detect and respond to a PWD's dressing activities.

Improving the accuracy of marker detection is only a part of the solution. To provide efficient feedback and guidance to the user, DRESS enables caregivers to customize its support so as to foster a PWD's personalized dressing sequence. Awareness of this sequence of actions may help DRESS in determining dressing progress and generating prompts and guidance. As dementia is often a progressive condition, the data generated in the tracking of progress and provision of these prompts may be useful in the assessment of cognitive decline overtime. At times when PWDs' dressing process has stalled, caregivers will receive an alert prompting them to intervene; as an alternative to responding to these alerts, in person, we plan on providing caregivers with tools to provide remote intervention via the DRESS prototype. Using a mobile phone, caregivers will be able to control the iPad and iPods on the drawers to provide voice and visual prompts, a technique known in the field of human-computer interaction as the "Wizard of Oz" method.

We have discussed laboratory testing of our DRESS prototype, integrating sensors and fiducial tracking to identify and guide the dressing process. However, successful deployment in the homes of PWD is more complex. For example, we must address the physical and cognitive differences between the population used in our study and PWDs. We plan to assess potential participants for our initial home trials to ensure that their physical capabilities will not encumber either their dressing activities or use of DRESS. In our initial home trials with PWDs, we will exclude individuals with limited physical capabilities (an issue that might encumber either their dressing activities or use of DRESS) to focus on the systems' ability to attend to the cognitive differences between the population used in this study and PWDs. This will enable our investigative efforts to focus on fine-tuning the DRESS prototype in ways that will appropriately attend to the cognitive needs of PWDs. We will pay particular attention to determining how the DRESS prototype prompts and the diverse nature of individuals' home environments might alter the detecting process and results.

Additional efforts will involve use of more acceptable markers, for example, using infrared inks to provide "invisible" markings on PWD's existing clothes. Although these "invisible markers" cannot be seen by humans, they can still be detected by the reacTIVision system. Members of our team are also exploring the potential for using machine-vision to train and recognize patterns inherent to individual PWD's existing clothing.

As part of the home deployment phase, we will also conduct cost-benefit analyses to determine the economic value of DRESS. Although it is too early to predict final product expense or initial markets segments (domestic, group homes, etc) a rough cost or benefit estimate can be considered. Assume a conservative rate of US $20/day to represent the combined value of increased PWD independence and reduced caregiver stress and effort, approximately US $600/month. In 2 to 4 months, a system cost of US $1200 to US $2000 would be recouped. Subscription plans rather than purchase might also make such systems broadly affordable.

Our study has shown that the DRESS prototype can detect clothing orientation and position and infer current state of dressing using a combination of sensors, intelligent software, and fiducial tracking. The DRESS prototype demonstrates a promising step toward automated dressing support to assist PWDs in maintaining their independence and privacy, while potentially providing their caregivers with much needed respite. We have identified several opportunities for improvement of the DRESS prototype. We plan to improve the markers, making them less obtrusive by making them "invisible" and/or training recognition systems to detect the natural patterns present in PWDs' existing clothes. These endeavors aim to optimize the overall detection of dressing status, as we embark on the next stage of our research agenda—deployment of the next iteration of the DRESS prototype in the homes of PWDs.

Acknowledgments

Research reported in this publication was supported by the National Institute of Nursing Research of the National Institutes of Health under award number R21NR013471.

Conflicts of Interest

None declared.

References

1. Batsch NL, Mittelman MS. ALZ.: Alzheimer's Disease International; 2012. World Alzheimer Report 2012: Overcoming the stigma of dementia URL: https://www.alz.co.uk/research/WorldAlzheimerReport2012.pdf [accessed 2018-03-23] [WebCite Cache ID 6y8YsMYYT]
2. Alzheimer's Association. ALZ. 2017. 2017 Alzheimer's Disease Facts and Figures URL: https://www.alz.org/documents_custom/2017-facts-and-figures.pdf [accessed 2018-03-28] [WebCite Cache ID 6yG45mISU]
3. World Health Organization. 2015. 10 Facts on Dementia URL: http://www.who.int/features/factfiles/dementia/dementia_facts/en/ [accessed 2018-03-23] [WebCite Cache ID 6y8ZvCI3K]
4. National Alliance for Caregiving, UnitedHealthcare. Caregiving. 2011 Jan. e-Connected Family Caregiver: Bringing Caregiving into the 21st Century URL: http://www.caregiving.org/data/FINAL_eConnected_Family_Caregiver_Study_Jan%202011.pdf [accessed 2018-03-24] [WebCite Cache ID 6yAgyGoWR]
5. Nichols LO, Martindale-Adams J, Greene WA, Burns R, Graney MJ, Lummus A. Dementia caregivers' most pressing concerns. Clin Gerontol 2008;32(1):1-14. [doi: 10.1080/07317110802468546]
6. Mahoney DF, LaRose S, Mahoney E. Family caregivers' perspectives on dementia related dressing difficulties at home: the preservation of self model. Dementia (London) 2015;14(4):494-512. [doi: 10.1177/1471301213501821] [Medline: 24339112]
7. reacTIVision. URL: http://reactivision.sourceforge.net/ [accessed 2017-04-03] [WebCite Cache ID 6pRwpOMdP]
8. Bencina R, Kaltenbrunner M. The design and evolution of fiducials for the reacTIVision system. 2005 Presented at: International Conference on Generative Systems in the Electronic Arts; 2005; Melbourne, Australia URL: https://modin.yuri.at/publications/reactivision_3rditeration2005.pdf
9. Mahoney DF, Burleson W, Lozano C, Ravishankar V, Mahoney EL. Prototype Development of a Responsive Emotive Sensing System (DRESS) to aid older persons with dementia to dress independently. Gerontechnology 2015;13(3):345-358 [FREE Full text] [doi: 10.4017/gt.2015.13.3.005.00] [Medline: 26321895]
10. Mihailidis A, Barbenel J, Fernie G. The efficacy of an intelligent cognitive orthosis to facilitate handwashing by persons with moderate to severe dementia. Neuropsychol Rehabil 2004;14(1-2):135-171. [doi: 10.1080/09602010343000156]
11. Mynatt E, Melenhorst A, Fisk A, Rogers W. Aware technologies for aging in place: understanding user needs and attitudes. IEEE Pervasive Computing 2004;3(2):36-41. [doi: 10.1109/MPRV.2004.1316816]
12. Hayes TL, Hunt JM, Adami A, Kaye JA. An electronic pillbox for continuous monitoring of medication adherence. 2006 Presented at: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 30-September 3, 2006; New York, NY. [doi: 10.1109/IEMBS.2006.260367]
13. Alm N, Astell A, Gowans G, Dye R, Ellis M, Vaughan P, et al. Lessons learned from developing cognitive support for communication, entertainment, and creativity for older people with dementia. In: Stephanidis C, editor. Universal Access in Human-Computer Interaction: Addressing Diversity. Berlin, Heidelberg: Springer; 2009:195-201.
14. Bharucha AJ, Anand V, Forlizzi J, Dew MA, Reynolds 3rd CF, Stevens S, et al. Intelligent assistive technology applications to dementia care: current capabilities, limitations, and future challenges. Am J Geriatr Psychiatry 2009 Feb;17(2):88-104 [FREE Full text] [doi: 10.1097/JGP.0b013e318187dde5] [Medline: 18849532]
15. Hamada T, Kuwahara N, Morimoto K, Yasuda K, Akira U, Abe S. Preliminary study on remote assistance for people with dementia at home by using multi-media contents. In: Stephanidis C, editor. Universal Access in Human-Computer Interaction: Addressing Diversity. Berlin, Heidelberg: Springer; 2009:236-244.
16. Kang HG, Mahoney DF, Hoenig H, Hirth VA, Bonato P, Hajjar I, Center for Integration of Medicine and Innovative Technology Working Group on Advanced Approaches to Physiologic Monitoring for the Aged. In situ monitoring of health in older adults: technologies and issues. J Am Geriatr Soc 2010 Aug;58(8):1579-1586. [doi: 10.1111/j.1532-5415.2010.02959.x] [Medline: 20646105]
17. Nygård L. The meaning of everyday technology as experienced by people with dementia who live alone. Dementia 2008 Nov;7(4):481-502. [doi: 10.1177/1471301208096631]
18. Tak SH, Benefield LE, Mahoney DF. Technology for long-term care. Res Gerontol Nurs 2010 Jan;3(1):61-72 [FREE Full text] [doi: 10.3928/19404921-20091103-01] [Medline: 20128544]
19. Wu J, Osuntogun A, Choudhury T, Philipose M, Rehg JM. A scalable approach to activity recognition based on object use. 2007 Presented at: IEEE International Conference on Computer Vision; October 14-21, 2007; Rio de Janeiro, Brazil. [doi: 10.1109/ICCV.2007.4408865]

20. Philipose M, Fishkin KP, Perkowitz M, Patterson DJ, Fox D, Kautz H, et al. Inferring activities from interactions with objects. IEEE Pervasive Comput 2004;3(4):50-57. [doi: 10.1109/MPRV.2004.7]
21. Dalton AF, Morgan F, OLaighin G. A preliminary study of using wireless kinematic sensors to identify basic Activities of Daily Living. Conf Proc IEEE Eng Med Biol Soc 2008;2008:2079-2082. [doi: 10.1109/IEMBS.2008.4649602] [Medline: 19163105]
22. Fleury A, Noury N, Vacher M. Supervised classification of activities of daily living in smart homes using SVM. 2009 Presented at: IEEE Annual International Conference of the Engineering in Medicine and Biology Society; November 13, 2009; Minneapolis, MN, USA. [doi: 10.1109/IEMBS.2009.5334931]
23. Namazi K, Johnson B. Dressing independently: a closet modification model for Alzheimer's disease patients. Am J Alzheimers Dis Other Demen 1992 Jan;7(1):22-28. [doi: 10.1177/153331759200700106]
24. Engelman K, Mathews R, Altus D. Restoring dressing independence in persons with Alzheimer's disease: a pilot study. Am J Alzheimers Dis Other Demen 2002 Jan;17(1):37-43. [doi: 10.1177/153331750201700102]
25. Popleteev A, Mayora O. Smart Buttoning: Assisting the dressing activities of Alzheimer patients. 2011 Presented at: Workshop on Ambient Intelligence on Healthcare and Social Wellness at the International Symposium on Ubiquitous Computing and Ambient Intelligence; 2011; Riviera Maya, Mexico.
26. Matic A, Mehta P, Rehg JM, Osmani V, Mayora O. Monitoring dressing activity failures through RFID and video. Methods Inf Med 2012;51(1):45-54. [doi: 10.3414/ME10-02-0026] [Medline: 21533305]
27. Mahoney DF, Coon DW, Lozano C. Latino/Hispanic Alzheimer's caregivers experiencing dementia-related dressing issues: corroboration of the Preservation of Self model and reactions to a “smart dresser” computer-based dressing aid. Digit Health 2016 Nov;2(1-12). [doi: 10.1177/2055207616677129]
28. Microsoft Corporation. 2017. Kinect for Windows URL: https://developer.microsoft.com/en-us/windows/kinect [accessed 2017-04-03] [WebCite Cache ID 6pRwVikLw]
29. Burleson W, Picard R. Evidence for gender specific approaches to the development of emotionally intelligent learning companions. IEEE Intelligent Systems 2007;22(4):62-69. [doi: 10.1109/MIS.2007.69]

Abbreviations

ADL: activities of daily living
PWD: person with dementia
RFID: radio-frequency identification

Edited by CL Parra-Calderón; submitted 01.02.16; peer-reviewed by B Hattink, AM Almeida, C Bailey; comments to author 30.03.16; revised version received 15.11.16; accepted 28.04.17; published 01.05.18.

Please cite as:
Burleson W, Lozano C, Ravishankar V, Lee J, Mahoney D
An Assistive Technology System that Provides Personalized Dressing Support for People Living with Dementia: Capability Study
JMIR Med Inform 2018;6(2):e21
URL: http://medinform.jmir.org/2018/2/e21/
doi:10.2196/medinform.5587
PMID:29716885

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org