JMIR Medical Informatics

Contents

Review

Original Papers

Review

Patient Similarity in Prediction Models Based on Health Data: A Scoping Review

Anis Sharafoddini[1], MSc; Joel A Dubin[1,2], PhD; Joon Lee[1], PhD

[1]Health Data Science Lab, School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada
[2]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Joon Lee, PhD
Health Data Science Lab
School of Public Health and Health Systems
University of Waterloo
200 University Avenue West
Lyle Hallman North
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567 ext 31567
Fax: 1 519 746 6776
Email: joon.lee@uwaterloo.ca

Abstract

Background: Physicians and health policy makers are required to make predictions during their decision making in various medical problems. Many advances have been made in predictive modeling toward outcome prediction, but these innovations target an average patient and are insufficiently adjustable for individual patients. One developing idea in this field is individualized predictive analytics based on patient similarity. The goal of this approach is to identify patients who are similar to an index patient and derive insights from the records of similar patients to provide personalized predictions..

Objective: The aim is to summarize and review published studies describing computer-based approaches for predicting patients' future health status based on health data and patient similarity, identify gaps, and provide a starting point for related future research.

Methods: The method involved (1) conducting the review by performing automated searches in Scopus, PubMed, and ISI Web of Science, selecting relevant studies by first screening titles and abstracts then analyzing full-texts, and (2) documenting by extracting publication details and information on context, predictors, missing data, modeling algorithm, outcome, and evaluation methods into a matrix table, synthesizing data, and reporting results.

Results: After duplicate removal, 1339 articles were screened in abstracts and titles and 67 were selected for full-text review. In total, 22 articles met the inclusion criteria. Within included articles, hospitals were the main source of data (n=10). Cardiovascular disease (n=7) and diabetes (n=4) were the dominant patient diseases. Most studies (n=18) used neighborhood-based approaches in devising prediction models. Two studies showed that patient similarity-based modeling outperformed population-based predictive methods.

Conclusions: Interest in patient similarity-based predictive modeling for diagnosis and prognosis has been growing. In addition to raw/coded health data, wavelet transform and term frequency-inverse document frequency methods were employed to extract predictors. Selecting predictors with potential to highlight special cases and defining new patient similarity metrics were among the gaps identified in the existing literature that provide starting points for future work. Patient status prediction models based on patient similarity and health data offer exciting potential for personalizing and ultimately improving health care, leading to better patient outcomes.

XSL•FO
RenderX

Introduction

Medicine is largely reactive—a disease is treated only after it is observed [1]. However, a move toward proactive medicine has been initiated by advances in technologies for analyzing the nature of a disease or estimating individual susceptibility to disease [2]. Moreover, a sharp increase in electronic health record (EHR) adoption has facilitated the move toward proactive medicine, which will hopefully lead to improved care and better patient outcomes. However, it is challenging for clinicians to examine and derive insights from multidimensional, large-scale EHR data. One pathway to proactive medicine employs predictive analytics to accurately derive insights from EHR data to predict disease progression. Predictive analytics, by employing EHRs, can also lead to personalized decision making based on the unique characteristics of a given patient [3].

Many studies have analyzed large populations to answer a wide range of health-related questions, including the study that developed Acute Physiology and Chronic Health Evaluation II (APACHE-II) [4]. These studies often provide statistically rigorous results for an average patient but are also expensive, time-consuming, and prone to selection bias [1]. Moreover, one of the major challenges for population-based studies is comorbidity, which limits generalizing a study to many patients [5,6]. Typically, these studies provide "the average best choice" [3]. Therefore, physicians cannot solely rely on the evidence from such population-based studies when facing a patient with conditions that deviate from the average.

One developing idea in this field is personalized predictive modeling based on patient similarity. The goal of this approach is to identify patients who are similar to an index patient and derive insights from the records of similar patients to provide personalized predictions. Employing patient similarity helps identify a precision cohort for an index patient, which will then be used to train a personalized model. Compared to conventional models trained on all patients, this approach has the potential to provide customized prediction. This approach has been widely used for personalized predictions in other fields, including music [7], movies [8], and sales pricing [9], and is referred to as *collaborative filtering* [10]. It can potentially be employed to manage a real-world patient with a complex health status and comorbidity profile. Patient similarity analytics also has the potential to assess the similarity between an index patient and trial population in conventional studies and help clinicians choose the most appropriate clinical trial [11].

Although the concept of patient similarity is not new—blood typing has been used for blood transfusion for more than a century [12]—advanced application of patient similarity is missing in the new era of data-driven medicine. The online PatientsLikeMe website provides a patient-reported database, where an index patient can find a cohort of similar patients and explore their data including symptoms, treatments, and tests [13]. Although PatientsLikeMe received the Drug Information Association 2014 President's Award for Outstanding Achievements in World Health, the full potential of patient similarity, especially in predictive modeling, has not been uncovered. Although there have been some attempts to embed patient similarity in health predictive modeling, a comprehensive picture of patient health predictive analytics based on health data (including EHRs) and patient similarity is lacking in the literature. The objectives of this paper are to provide an analysis and summary of the studies on patient health prediction models based on health data and patient similarity, identify any gaps in this area, and suggest ideas for future work. Overall, this review aims to address the following three research questions:

In which context (applications) have patient health prediction models based on health data and patient similarity been used?

Which modeling techniques have been considered in the literature?

How do patient similarity-based models affect health predictions in comparison to conventional models?

We hope the results could also contribute to the broad field of case-based reasoning (CBR)—with the core component of similarity assessment—to meet the challenges in medical applications [14].

Methods

A systematic search approach in line with guidelines of Kitchenham et al [15] was taken to review and analyze the literature on patient similarity in health prediction models based on health data. However, this paper does not aim to report the performance of particular models and identify the best model because various health data types and performance measures are possible.

Inclusion and Exclusion Criteria

Studies included in this review had to be journal articles or conference proceedings written in English. They had to focus on prediction in the health domain, devise a model for prediction, embed explicit patient similarity analytics, and utilize health data for training their model. Studies were excluded if (1) they entirely relied on human input for predictions or similarity assessment, (2) the model was tested on seen data—the part of the data used for training the algorithm, and (3) the algorithm was trained using only genomic data. If the same study appeared in multiple publications, only the most comprehensive and latest version was included.

Paper Selection

The literature search was finalized in December 2015. Scopus, PubMed, and ISI Web of Science, all databases covering health-related publications, were searched for peer-reviewed studies with keywords related to "prediction," "health data," and "patient similarity." The search strings used in each of these search engines are given in Multimedia Appendix 1. After removal of duplicates, the title and abstract of each identified article were screened. The remaining articles were further examined in full text to finalize the set of included articles.

Data Extraction and Analysis

Data from included articles were extracted into a matrix table and analyzed with respect to the following criterion: publication information, context, predictors (or features), missing data, modeling algorithms, performance measures, and outcomes.

The context was further examined from two points of view: data source and application area. The employed patient similarity-based modeling algorithms were also synthesized in three categories: neighborhood-based, clustering-based, and other algorithms, with the majority falling in the first category. Because measuring predictive performance is essential to model development (model selection/model tuning)—and can also be used to compare a given model with other methods (performance estimation)—evaluation metrics along with validation techniques used in the reviewed studies were also extracted.

Results

A total of 22 articles were included in the review (Figure 1). Tables 1 and 2 summarize the data extracted from input data/predictors and outcome perspectives, respectively.

Table 1. Summary of the reviewed articles in terms of data type, data origin, number of predictors, and number of instances (N=22).

Authors	Data type	Data origin[a]	Predictors, n[b]	Instances, n[c]
	Cross-sectional			
Jurisica et al [16]	Cross-sectional	NR	55	788
Bobrowski [17]	Cross-sectional	The Gastroenterological Clinic of the Institute of Food and Feeding in Warsaw [18]—a database consisting of hepatological patient data	40	511
Park et al [19]	Cross-sectional	UCI repository [20]-Dermatology	35	350
	Cross-sectional	UCI repository [20]-Heart Disease: Cleveland Clinic;	13	270
	Cross-sectional	UCI repository[20]-Breast Cancer Wisconsin: University of Wisconsin Hospital and Clinics	31	560
	Cross-sectional	UCI repository[20]-Pima Indians Diabetes: NR	8	760
	Cross-sectional	UCI repository[20]-Liver Disorders: BUPA Medical Research	7	340
Saeed et al [21]	Longitudinal	MIMIC-II [22]-The Multiparameter Intelligent Monitoring in Intensive Care at Boston's Beth Israel Deaconess Medical Center	50	377
Chattopadhyay et al [23]	Cross-sectional	Hospital-history of suicidal attempts and committed suicides collected from hospital records	15	50
Sun et al [24]	Longitudinal	MIMIC-II [22]-The Multiparameter Intelligent Monitoring in Intensive Care at Boston's Beth Israel Deaconess Medical Center	50	74
Sun et al [25]	Longitudinal	MIMIC-II [22]-The Multiparameter Intelligent Monitoring in Intensive Care at Boston's Beth Israel Deaconess Medical Center	10	1500
David et al [26]	Cross-sectional	Laboratory results generated by two Beckman-Coulter Gen-S analyzers at an acute care facility in Brooklyn	NR	4900
Houeland [27]	Cross-sectional	A dataset focused on palliative care for cancer patients	55	1486
Wang et al [28]	Cross-sectional	UCI repository [20]-Breast Cancer Wisconsin: University of Wisconsin Hospital and Clinics	31	560
	Cross-sectional	UCI repository [20]-Pima Indians diabetes	8	760
	Cross-sectional	A real-world EHR data warehouse of a health network consisting of data from 135K patients over a year	NR	135K
Wang et al [29]	Cross-sectional	A real-world EHR data warehouse of a health network consisting of data from 135K patients over a year	2388	3946
Campillo-Gimenez et al [30]	Cross-sectional	French Renal Epidemiology and Information Network (REIN) registry [31]	19	1137
Gottlieb et al [32]	Cross-sectional and longitudinal	Hospital dataset-Stanford Medical Center, USA	16	9974
	Cross-sectional and longitudinal	Hospital dataset-Rabin Medical Center, Israel	16	5513
Lowsky et al [33]	Cross-sectional	A dataset by the United States Renal Data System (USRDS) consisting of all kidney transplant procedures from 1969 to 1999	13	51,088
Hielscher et al [34]	Cross-sectional	The Study of Health in Pomerania (SHIP) [35]-a dataset consisting of a comprehensive examination program including but not limited to ultrasound tests and laboratory analysis	65/57	578
Zhang et al [36]	Longitudinal	A 3-year longitudinal EHR data of 110,157 patients	NR	1219
Henriques et al [37]	Longitudinal	myHeart home telemonitoring study [38]-daily physiological records including blood pressure, respiration rate, heart rate, and body weight	NR	41
Lee et al [39]	Cross-sectional and longitudinal	MIMIC-II [22]-The Multiparameter Intelligent Monitoring in Intensive Care at Boston's Beth Israel Deaconess Medical Center	76	17,152

Authors	Data type	Data origin[a]	Predictors, n[b]	Instances, n[c]
	Cross-sectional			
Ng et al [40]	Cross-sectional and longitudinal	A longitudinal medical claims database consisting of data from over 300,000 patients during four years	8500	15038
Panahiazar et al [41]	Cross-sectional	The Mayo Clinic	33	1386
Wang [42]	Cross-sectional	UCI repository [20]-Breast Cancer Wisconsin: University of Wisconsin Hospital and Clinics	31	560
	Cross-sectional	UCI repository[20]-Pima Indians diabetes	8	760
	Cross-sectional	A real-world EHR data warehouse	NR	135K
Wang et al [43]	Cross-sectional	A real-world EHR data warehouse	127	3946

[a] NR: not reported.

[b] Predictors: the total number of predictors.

[c] Instances: the total number of data points used in each study including the training and test.

**Table 2.** Summary of reviewed articles in terms of outcome, evaluation metrics, and comparing methods (N=22).

Authors	Outcome[a]	Evaluation metrics[b]	Compared against[c]
Jurisica et al [16]	Suggesting hormonal therapy (day of human chorionic gonadotrophin administration and the number of ampoules of human menopausal gonadotrophin) after in vitro fertilization and predicting pregnancy outcome (pregnancy, abortion, ectopic pregnancy, and ovarian hyperstimulation syndrome)	Accuracy	NR
Bobrowski [17]	Four types of liver disease (cirrhosis hepatis biliaris primaria, cirrhosis hepatis decompensata, hepatitis chronica activa, and hepatitis chronica steatosis)	Accuracy	Classic k-NN (k=10)
Park et al [19]	(1) Six types of dermatology diseases (psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, pityriasis rubra pilaris); (2) diagnosis of heart disease (angiographic disease status); (3) diagnosis of a breast tumor as malignant or benign; (4) diagnosis of diabetes; (5) diagnosis of liver disorder	Accuracy; sensitivity; specificity	LR; C5.0; CART; neural network; conventional CBR (k=5) with five neighbors
Saeed et al [21]	Hemodynamic stability or instability of an episode	Sensitivity; positive predictive value	NR
Chattopadhyay et al [23]	Suicidal risk levels (level 1: suicidal plans or thoughts; level 2: single suicidal attempt; level 3: multiple suicidal attempts)	NR	NR
Sun et al [24]	Occurrence of acute hypotensive episode within the forecast window of an hour	Accuracy	Human expert's idea based on the Euclidean [44]; k-NN over low-dimensional space after applying PCA
Sun et al [25]	Occurrence of acute hypotensive episode within the forecast window of an hour	Accuracy	Human expert's idea based on the Euclidean [44]; k-NN over low-dimensional space after applying PCA
David et al [26]	Seven disease diagnoses (microcytic anemia, normocytic anemia, mild SIRS, thrombocytopenia, leukocytopenia, moderate/severe SIRS, normal)	Accuracy	Human expert's idea
Houeland [27]	Pain levels	Error rate (1-accuracy).	Random retrieval; k-NN (k=1) with the Euclidian distance; random forest
Wang et al [28]	(1) Diagnosis of a breast tumor as malignant or benign; (2) diagnosis of diabetes; (3) diagnosis of dementia without complications (HCC352) or diabetes with no or unspecified complications (HCC019)	Accuracy; sensitivity; precision; F-measure	PCA; LDA [45]; LSDA [45]; LSML [24]
Wang et al [29]	Diagnosis of CHF 6 months later	Accuracy; sensitivity; precision; F-measure	LLE; LE; PCA; Euclidean distance.
Campillo-Gimenez et al [30]	Registration on the renal transplant waiting list: yes/no	ROC curve	k-NN; LR; k-NN with weighted predictors; k-NN with weighted patients
Gottlieb et al [32]	Patient discharge diagnosis *ICD* codes	ROC curve; F-measure	NR
Lowsky et al [33]	Graft survival probability	IPEC	Cox model; RSF [46]
Hielscher et al [34]	Three levels of liver fat concentration measured by magnetic resonance tomography: (1) fat concentration <10%; (2) fat concentration of 10%-25%; (3) fat concentration ≥25%	Accuracy; sensitivity; specificity	Multiple variants of the k-NN: majority voting; weighted voting; with/without predictor selection
Zhang et al [36]	Four effective drugs for hypercholesterolemia treatment: atorvastatin, lovastatin, pravastatin, and simvastatin	ROC curve	Patient similarity; patient similarity with drug structure similarity; patient similarity with drug target similarity
Henriques et al [37]	Early detection of heart failure: decompensation or normal condition	Sensitivity; specificity; F-measure; G-measure	Coefficients' distance; linear correlation of signals; Euclidean distance
Lee et al [39]	30-day in-hospital mortality	Area under ROC curve; area under precision-recall curve	Population-based and personalized versions of: majority vote; LR; DT

XSL•FO
RenderX

Authors	Outcome[a]	Evaluation metrics[b]	Compared against[c]
Ng et al [40]	The risk of diabetes disease onset	ROC curve	Global LR; *k*-NN; patient similarity-based LR with Euclidean distance
Panahiazar et al [41]	Medication plans for heart-failure patients (angiotensin-converting enzyme, angiotensin receptor blockers, β-adrenoceptor antagonists, statins, and calcium channel blocker)	Sensitivity; specificity; F-measure; accuracy	K-means; hierarchical clustering
Wang [42]	(1) Diagnosis of a breast tumor as malignant or benign; (2) diagnosis of diabetes; (3) occurrence of CHF within 6 months	Precision; F-measure; sensitivity; accuracy	*kd*-tree; PCA- *kd*-tree; ball-tree; spectral-tree.
Wang et al [43]	Occurrence of CHF within 6 months	Precision; F-measure; sensitivity; accuracy	PCA; Laplacian regularized metric learning [47]; LLE [48]; LSR; LSML [24]

[a] CHF: congestive heart failure; *ICD*: *International Classification of Diseases*.

[b] IPEC: integrated prediction error curve ; NR: not reported; ROC: receiver operating characteristic: SIRS: systemic inflammatory response syndrome.

[c] CART: classification and regression tree; CBR: case-based reasoning; DT: decision tree; *k*-NN: *k*- nearest neighbor; *kd*-tree: *k* dimensional tree; LDA: linear discriminant analysis; LE: Laplacian embedding; LLE: locally linear embedding; LR: logistic regression; LSDA: locality sensitive discriminant analysis; LSML: locally supervised metric learning; LSR: local spline regression; NR: not reported; PCA: principal component analysis; RSF: random survival forest.

XSL•FO
RenderX

Figure 1. Flow diagram of article selection procedure.
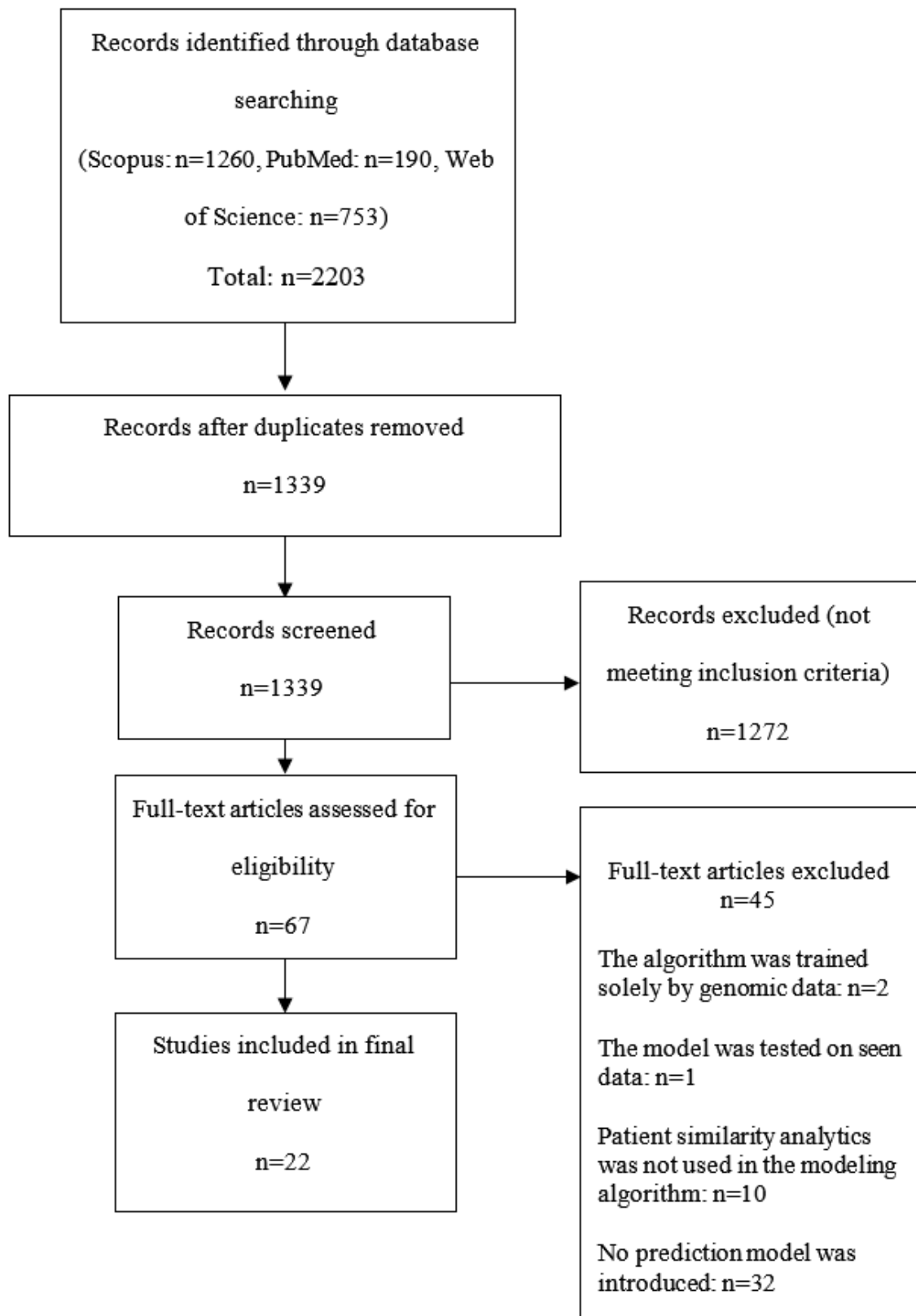
Publication Information

The level of interest could be gauged by the increase in publication on this topic in recent years (Figure 2). Fifteen studies of 22 were journal publications [16,17,19,21,23,25,26,30,32,33,36,37,39,40,42,43] and seven were conference articles [24,27-29,34,41].

Figure 2. Distribution of publications by year.
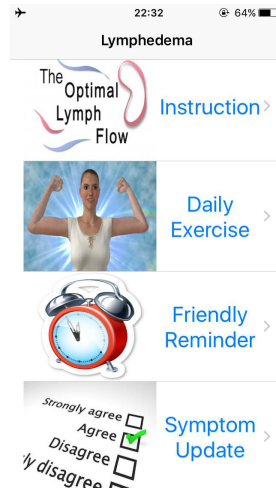
The Context

Although a considerable number of articles did not clearly state the source of data—some articles used more than one dataset—hospitals were named (10/22); within hospitals, intensive care units (ICUs) were the main sources of data (5/10).

In addition, one study [19] used data from a research center and another study [37] utilized telemonitoring data, also known as wearable-based remote patient monitoring data. From the application area perspective, chronic diseases were the most prevalent context. For detailed distributions, refer to Figure 3.

Figure 3. Focused application areas of studies. Some studies featured more than a single application area and were counted more than once.
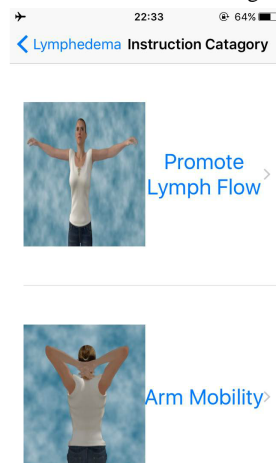
The Predictors

Raw health data can be in various formats, including narrative/textual data (eg, history of a present illness), numerical measurements (eg, laboratory results, vital signs, and measurements), recorded signals (eg, electrocardiograms), and pictures (eg, radiologic images). Numerical measurements and recorded signals were the format used most in the reviewed articles. Three main approaches were used for extracting predictors from raw health data. First, for some variables, including age and gender, the exact/coded value was used as a predictor. Second, in articles employing recorded signals and/or longitudinal numerical measurements, numeric variables, including wavelet coefficients, minimums, maximums, means, and variances, were extracted from within particular time windows [24,25,37,39,41]. Third, three studies [21,29,43] employed the term frequency-inverse document frequency (TF-IDF) technique from text mining to produce predictors for their model.

Although predictor extraction affects the performance of the model [24], one of the challenging tasks in patient similarity-based predictive modeling is identifying the most relevant and important patient characteristics for patient similarity assessment. Patient similarity assessment is generally defined as investigating the similarity of patients' data in terms of their symptoms, comorbidities, demographics, and treatments, but there is no predefined list of predictors to be considered. Most of the studies proposed an arbitrary list of predictors or limited their work to the available predictors, but selected predictors must be representative of patient's condition in each particular application. Two studies [26,30] employed weighting schemes to adjust the importance of the predictors based on the outcome. One study [34] showed that predictors selected by a correlation-based feature algorithm could vary according to gender. Although feature selection methods can help with predictor selection, selected predictors may not be the most appropriate ones for each individual patient because they are derived from general analysis of the population. One study [40]

showed that a group of similar patients has a similar set of predictors, but the predictors' importance was different between individuals. Two studies [29,43] suggested utilizing expert knowledge on the similarity of cases to implicitly consider case-specific predictors. One study [16] proposed a context-based similarity metric in which an expert determined a set of predictors and their allowable values for patient similarity assessment.

Missing Data

One common challenge in using health data in predictive analytics is missing data. Most of the modeling techniques cannot handle an incomplete data matrix. Nevertheless, all studies that mentioned this challenge [19,23,25,30,32,39,41], except for two [23,25], simply excluded patients with incomplete data. In Chattopadhyay et al's study [23], missing values were replaced with the most common value of the corresponding predictor. Sun et al [25] evaluated two methods in overcoming missingness: replacing the missing value with the mean of the sensor measurements within a time window or imputing based on the correlations among multiple sensors using linear regression models. The latter method consistently performed better than the former. Three studies [21,29,43] that mapped EHRs to TF-IDF space handled the missing data challenge indirectly. Other studies did not discuss missing data.

The Modeling Algorithms

Neighborhood-Based Algorithms

The neighborhood-based algorithms indicate studies in which a group of patients similar to an index patient is retrieved and a prediction is produced by a model trained on similar patients' data. This category is comparable to memory-based techniques in collaborative filtering [10]. Various types of similarity metrics can describe the similarity between patients. Studies in this category [16,17,19,21,23-26,28,30,33,34,37,39,40,43,47] were organized based on the type of similarity metric they employed for calculating patient similarity.

Distance-Based Similarity Metrics

Twelve studies (of 18) used various types of distance-based similarity. One study [23] utilized the sum of absolute distances for each predictor to retrieve a cohort of similar patients and find the closest class to a new patient.

Five studies [17,19,26,34,43] utilized the Euclidean distance. Bobrowski [17] designed a linear transformation by solving a convex optimization problem to maximize between-class distances and minimize in-class distances. In this study, the k-nearest neighbor (k-NN) method on the transformed data outperformed the classical k-NN algorithm. Park et al [19] investigated the optimum number of neighbors for each patient. In this study, a grid search found a cut-off probability based on the distribution of pairwise distances to define a distance threshold. This method outperformed several conventional machine learning algorithms, including logistic regression (LR), C5.0 decision tree (DT), classification and regression tree, neural network, and conventional CBR.

David et al [26] employed the Euclidean distance on weighted predictors to select neighbors for an index patient. Although

their method strongly agreed with a human reviewer, no comparison with other methods was reported. Hielscher et al [34] suggested the idea of subgrouping the training set based on gender and then applying a k-NN method. This study showed that using a predictor selection algorithm can reduce the dimension of the predictor space and improve performance. Furthermore, the results demonstrated that only a few of the predictors with highest predictive power within each subgroup are common, thus highlighting the efficiency of subgrouping a population, then considering customized predictors for each subgroup.

Six studies utilized the Mahalanobis distance [24,25,28,29,33,40]. Sun et al [24] defined a Mahalanobis distance by solving an optimization problem aimed at minimizing the within-class squared distances and maximizing between-class squared distances. A sensitivity analysis of the parameter k— number of neighbors — revealed that small k resulted in lower classification error, confirming the idea of using local information. The proposed metric outperformed the Euclidean distance. As an extension of a previous study [12], Sun et al [25] trained a linear regression model based on a least squared error fitting technique on the retrieved data. The proposed method outperformed the previous method and k-NN with Euclidean distance on a lower dimensional space mapped by linear discriminant analysis.

Wang et al [28] focused on integrating multiple patient similarity metrics learned independently without sharing the training datasets. In combining the metrics, various degrees of importance were considered for each individual Mahalanobis metric. The proposed method outperformed all compared methods and improved accuracy even when some individual metrics were biased. Building on that study, Wang et al [29] proposed a new algorithm in which human experts' ideas could be embedded. To incorporate expert knowledge, two matrices were defined by an expert: a similarity matrix and dissimilarity matrix. The proposed method outperformed k-NN with the Euclidean distance in the original feature space and low-dimensional spaces derived by PCA, locally linear embedding and Laplacian embedding.

Wang [43] then proposed a two-term objective function for Mahalanobis distance learning: a part based on human experts' knowledge (following the same procedure as in the previous study) and a part based on available historical data. The proposed online distance metric learning method outperformed locally supervised metric learning [24]. In addition, the results showed that the performance increased from 20 to 200 neighbors, but decreased after 200. This result supports the advantage of using local neighborhood data.

Lowsky et al [33] proposed a neighborhood-based survival probability prediction model based on a Mahalanobis distance and constructed a weighted Kaplan-Meier survival curve on the basis of retrieved similar cases. Although their method did not show consistent advantage over the Cox model on the original dataset, its performance improved as the proportional hazards violation was highlighted on the simulated datasets.

Ng et al [40] compared personalized predictive modeling and population-based predictive models. The proposed algorithm

made predictions using Mahalanobis and an LR model. Clustering analysis of risk factors revealed that patients with similar risk factors were grouped together, whereas patients with different risk factors were distributed in groups far apart in the cluster tree. Furthermore, a large number of risk factors were not captured by the population-based model, whereas personalized models highlighted them.

Correlation-Based Similarity Metrics

Saeed et al [21] utilized a correlation coefficient to retrieve the *k* most similar patients. This correlation coefficient measured the extent of the linear correlation between two data points. The proposed algorithm was not benchmarked against other methods.

Cosine-Similarity Metrics

Lee et al [39] examined the hypothesis that predictive modeling based on patient similarity analytics can outperform conventional predictive modeling in which all available patient data are analyzed. Their study employed cosine patient similarity and focused on characterizing neighborhood size and model performance. Results confirmed that patient similarity analytics can outperform not only population-based models but also well-known clinical scoring systems. Moreover, a reasonably small and homogenous neighborhood improved predictive performance; however, a very small neighborhood compromised performance due to small sample size effects.

Other Similarity Metrics

Four studies used other similarity metrics. One of the earliest proposed methods [16] retrieved patients based on the context defined by a user. A context was defined as a set of predictors and had allowable values for these predictors in the retrieval task. Houeland [27] proposed a combination of a Euclidean distance and a tree-based distance. Each case in the training set was stored with its associated terminal node for every tree in a forest of randomly grown trees. For a new patient, half of the most similar patients in the training set were retrieved based on Euclidian distance. Then, two patients were considered to be more similar if they shared the same terminal node assignments for a higher number of trees. The proposed method outperformed conventional random forest and *k*-NN with the Euclidean distance.

Campillo-Gimenez et al [30] employed an exclusive OR-based patient similarity metric with an LR model. This method outperformed compared methods, including population-based LR, and performed well, after randomly generated predictors were added to the relevant predictors. Henriques et al [37] utilized a similarity metric based on the signs of Haar wavelet coefficients derived from telemonitoring data. A metric based on the coefficients' signs outperformed similarity metrics based on the coefficients' distances, Euclidian distance, and linear correlation of the actual data points.

Cluster-Based Algorithms

Cluster-based algorithms group patients in a training set based on their profiles and relationships. Therefore, a new patient is assigned to a predefined cluster based on his/her similarity to each cluster. These methods have a trade-off between prediction performance and scalability for large datasets. Only one study

[41] employed supervised and unsupervised clustering approaches with a Mahalanobis distance in recommending a medication to a heart-failure patient. Then, the most frequently prescribed medication in the most similar cluster was selected for the index patient. The proposed supervised clustering outperformed hierarchical clustering and *k*-means.

Other Algorithms

Gottlieb et al [32] focused on associations between hospitalization data and discharge diagnoses, considering eight similarity metrics between hospitalization data and two similarity measures for *International Classification of Diseases* codes. Then, they combined these measures into 16 hospitalization-discharge code associations. For a new patient's hospitalization data, the score of a potential discharge code was calculated by considering the similarity to the known discharge code-hospitalizations' associations, and then an LR classifier was trained to distinguish true associations (of medical history with diagnosis) from false ones. Using various similarity metrics helped overcome the limitations of using only one particular similarity metric—using just one similarity metric for all predictors may miss information relevant to prediction [49].

Zhang et al [36] augmented patient similarity analytics with drug similarity analytics and proposed an algorithm for personalized drug recommendations in hypercholesterolemia treatment. Based on the Jaccard similarity metric in their label propagation algorithm, they defined three sets of similarities: (1) patient-patient, (2) drug-drug, and (3) patient-drug. This study suggested that combining patient similarity with drug similarity can help achieve personalized medicine.

Wang [42] proposed an adaptive semisupervised recursive tree partitioning (ART) approach to reduce the computational burden of pairwise patient similarity calculations. This algorithm can also leverage expert knowledge. The algorithm constructs a tree used to index patient profiles and then rapidly retrieve the nearest neighbors to a new patient. The ART series methods generally performed better than compared methods.

Outcomes

The outcomes of prediction models normally take six forms: continuous, binary, categorical (but not ordered), ordinal, count, and survival. The studies reviewed targeted continuous outcomes [16], such as hormonal therapy dosage; binary outcomes [19,21,24,25,28-30,37,39,42,43], such as disease diagnosis or patient death; categorical outcomes [17,19,26,32,36,41], such as multiple-disease diagnosis; and ordinal outcomes [23,27,34,40], such as the grade of an illness. One study also aimed to predict a survival outcome [34] (ie, the prediction of the time to an event of interest) [50]. No study had a count outcome, which is a nonnegative integer value derived from counting rather than grading.

Evaluation Metrics and Validation Techniques

Evaluation Metrics

Evaluation metrics are widely used to tune the parameters of a model and compare the model with other methods.

Evaluation Metrics Based on a Confusion Matrix

A confusion matrix is a cross-tabulation representation of observed and predicted classes. Various evaluation metrics extracted from a confusion matrix—including accuracy, sensitivity, specificity, F-measure, G-measure, precision, and positive predictive value—were used in the included articles.

Receiver Operating Characteristic Curve

Five articles [30,32,36,39,40] used the receiver operating characteristic (ROC) curve, and one [39] used the precision-recall curve in combination with the ROC curve to overcome the optimistic estimate of ROC curves in the presence of imbalanced data—where class distribution is not approximately uniform among the classes.

Measures Based on Model Residuals

When a model generates a continuous outcome, a common performance measure is the mean squared error. This metric is based on model residuals, which are the difference between the observed and predicted responses, and can be calculated by taking the average of squared model residuals. One study [25] used a relative error rate and another [33] used integrated prediction error curve, a time-invariant measure that calculates the weighted quadratic difference of prediction and observed survival outcome.

Validation Techniques

Validation techniques can generally be grouped into two categories: internal and external [50].

Internal Validation Techniques

Internal validation techniques randomly split the available dataset into two parts using various approaches: a training set and a test set. Seven studies [21,23,25,26,28,30,51] used various ratios for this splitting. Eight studies [19,32,34,36,39-42] employed a *k*-fold cross-validation technique and five studies [16,17,24,27,37] used leave-one-out.

External Validation Techniques

External validation means assessing the performance of the prediction model in other scenarios or settings (eg, assessing the geographic or temporal transportability of the model). Only one study [33] used temporal validation for assessing their model's performance. External validation better evaluates generalization of a model to new patients.

Discussion

Over the period of 1989 to 2015, we found 22 articles that focused on patient similarity in predictive modeling using EHR data, with an increase in the number of these studies over time. Overall, three main approaches were employed in these studies to leverage patient similarity: neighborhood-based modeling, clustering modeling, and other algorithms. This section discusses the results from this review study to address the research questions then identifies gaps and future research directions.

Predictive Modeling Based on Patient Similarity and Health Data Context

This study showed that patient similarity-based predictive modeling has been widely used on hospital data, which sheds light on the need for patient similarity-based predictive modeling in tackling big data. In addition, further analysis revealed that ICUs are the central focus in hospitals. ICUs treat patients with severe and life-threatening illnesses that require continuous monitoring. Thus, ICU patients are surrounded by equipment that constantly generates a large amount of data. However, this large volume usually overwhelms clinicians and highlights the need for a computerized system. In addition, the critical health status of the patients in ICUs requires more proactive (rather than reactive), precise, and personalized care. Therefore, ICUs are a suitable environment for personalized prediction models.

Furthermore, chronic disease prognosis was one of the common application areas for personalized predictive modeling. Such analytics can help in improving patient health status if used in planning new therapies or interventions to prevent further complications. Patient similarity analytics can also be used for predicting a patient's risk of developing further complications or disease. In particular, patient similarity analytics can overcome the challenge of comorbidities in chronic disease risk stratification and provide customized plans for a given patient. It is worth mentioning that cardiovascular diseases and diabetes were common application domains among the reviewed studies.

Modeling Techniques

Most of the studies focused on neighborhood-based modeling. These models are easy to implement and they typically perform well. However, their performance depends greatly on the chosen patient similarity metric. Although there are a variety of similarity metrics in data mining [52], distance-based similarity metrics were the most popular in the reviewed studies. These methods are also constrained by their limited scalability for big data. Although Lee et al [39] suggested that computational load can be parallelized, the high computational load of neighborhood-based methods in comparison to other models is not trivial.

Cluster-based methods exhibit better scalability than neighborhood-based modeling, but there is a trade-off between prediction accuracy and scalability. These methods may not satisfactorily address the prediction for patients with rare conditions because they work based on predefined clusters. Especially in hierarchical clustering methods, in which final clusters are derived based on merging smaller clusters [53], the algorithm may fail to provide personalized predictions for patients with a rare condition.

Four studies embedded patient similarity analytics in their modeling approach even though they did not explicitly compute a patient similarity metric [27,32,36,42]. These studies reported improved prediction performance and overcame the limitations of neighborhood-based algorithms. However, these methods tend to be associated with increased computational and mathematical complexity. Mathematical complexity can lead to decreased interpretability in the context of how the model has learned to solve a problem. Nevertheless,

neighborhood-based methods and cluster-based methods maintain a fair level of interpretability (a summary of the reviewed articles in terms of methodology is provided in Multimedia Appendix 2).

Patient Similarity-Based Models Versus Conventional Models

Only two studies [39,40] directly compared the performances of patient similarity-based models and population-based models. Both demonstrated that patient similarity-based models resulted in better predictive performance. Lee et al [39] also compared the performance of patient similarity-based models to the Sequential Organ Failure Assessment [54] and the Simplified Acute Physiology Score [55], two widely used scoring systems in ICUs, and the patient similarity-based models showed a significant improvement.

Gaps and Future Work

One of the factors that strongly affects predictive performance is the choice of predictors. Results show that researchers are searching for reliable predictors to enhance the performance of patient similarity-based models. In the context of personalized prediction models, the best possible predictors should have at least two characteristics: (1) be capable of capturing the progression of a patient's health status, and (2) be as discriminative as possible. Applying TF-IDF technique could help boost the accuracy of similarity assessment for patients with rare conditions [21] in the predictor extraction phase because the IDF value is low for common clinical observations and high for rare observations. Although identifying the relevant predictors for patient similarity assessment is of special importance for precise prediction, only a few studies have considered this component in their proposed framework. Although feature selection techniques, predictor weighting schemes, and experts' opinions were used in the reviewed articles to address this question, further studies are needed to identify appropriate predictors. However, as the number of predictors increases, the performance of many types of prediction models may decline and this can lead to a generalizability concern; hence, the need for external validation of prediction models. This challenge may be encountered in patient similarity predictive modeling, particularly with neighborhood-based methods where a model is developed from a small cohort of similar patients and the number of predictors may exceed the number of training instances. Therefore, creating a balance between the number of training instances—training sample size—and the number of predictors is important.

As observed in several studies, some values for a given patient may be missing. Although imputation methods can help deal with missing data, it is important to determine why the values are missing. Sometimes, associations exist between patterns of missing data and the outcomes. This type of information gap is referred to as informative missingness [56]. Further studies that account for this type of missingness are needed.

As mentioned previously, a wide variety of techniques have been employed in efforts to achieve personalized prediction. Neighborhood-based methods are among the most popular techniques. However, abundant room remains for progress in defining new patient similarity metrics. In addition, as suggested by Gottlieb et al [32], various similarity metrics based on different predictors can be combined to devise better similarity metrics.

There are some limitations to this review. First, although the article selection protocol was devised by all reviewers, there could have been a bias in selecting articles because title and abstract screening was done by only one reviewer. Second, the search process focused on the more generic terms covering the concept of EHR, and it might have excluded articles in which domain-specific words (eg, "diabetes data") were used to describe the data source. Finally, due to inaccessibility to some EHR data in the included studies, data quality assessment was infeasible and all the studies received equal importance in the interpretation of the findings, which might have caused a bias in the results.

Conclusion

Personalized medicine has the potential to facilitate predictive medicine, provide tailored prognoses/diagnoses, and prescribe more effective treatments. Interest is increasing in the use of personalized predictive modeling and various patient similarity-based models using EHRs have been described in the literature. This review has demonstrated the value of patient similarity-based models in critical health problems and noted the results of two studies [39,40] on the superiority of patient similarity-based models over population-based ones. The suggested future work could improve the capabilities of these models.

Authors' Contributions

Study conception and review design were conducted by AS, JAD, and JL. AS screened titles and abstracts of references identified in the databases. AS, JAD, and JL designed the data extraction instrument. AS extracted data from the original studies and prepared the initial results. Interpretation of the results was provided by all authors and study supervision was provided by JAD and JL. All authors contributed in writing the manuscript and approved the final version of the review.

XSL•FO

RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strings used to search databases.

[PDF File (Adobe PDF File), 60KB - medinform_v5i1e7_app1.pdf]

Multimedia Appendix 2

Summary of the reviewed articles in terms of methodology (N=22).

[PDF File (Adobe PDF File), 70KB - medinform_v5i1e7_app2.pdf]

References

1. Miner L, Bolding P, Hilbe J, Goldstein M, Hill T, Nisbet R, et al. Biomedical informatics. In: Practical Predictive Analytics and Decisioning Systems for Medicine. Cambridge, MA: Academic Press; 2015:42-59.
2. Snyderman R. Personalized health care: from theory to practice. Biotechnol J 2012 Aug;7(8):973-979. [doi: 10.1002/biot.201100297] [Medline: 22180345]
3. Bellazzi R, Ferrazzi F, Sacchi L. Predictive data mining in clinical medicine: a focus on selected methods and applications. WIREs Data Mining Knowl Discov 2011 Feb 11;1(5):416-430. [doi: 10.1002/widm.23]
4. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985 Oct;13(10):818-829. [Medline: 3928249]
5. Campbell-Scherer D. Multimorbidity: a challenge for evidence-based medicine. Evid Based Med 2010 Dec;15(6):165-166. [doi: 10.1136/ebm1154] [Medline: 21106673]
6. Gotz D, Stavropoulos H, Sun J, Wang F. ICDA: a platform for Intelligent Care Delivery Analytics. AMIA Annu Symp Proc 2012;2012:264-273 [FREE Full text] [Medline: 23304296]
7. Shardanand U, Maes P. Social information filtering: algorithms for automating "word of mouth". In: Proceedings of the Conference on Human Factors in Computing Systems. 1995 Presented at: Conference on Human Factors in Computing Systems; May 7-11, 1995; Denver, CO p. 210-217. [doi: 10.1145/223904.223931]
8. Hill W, Stead L, Rosenstein M, Furnas G. Recommending and Evaluating Choices in a Virtual Community of Use. In: Proceedings of the Conference on Human Factors in Computing Systems. 1995 Presented at: Conference on Human Factors in Computing Systems; May 7-11, 1995; Denver, CO p. 194-201. [doi: 10.1145/223904.223929]
9. Xu R, Nettleton D, Nordman DJ. Case-specific random forests. J Comput Graph Stat 2016 Mar 09;25(1):49-65. [doi: 10.1080/10618600.2014.983641]
10. Su X, Khoshgoftaar TM. A survey of collaborative filtering techniques. Adv Artif Int 2009;2009:1-19. [doi: 10.1155/2009/421425]
11. Cahan A, Cimino JJ. Visual assessment of the similarity between a patient and trial population: Is This Clinical Trial Applicable to My Patient? Appl Clin Inform 2016;7(2):477-488. [doi: 10.4338/ACI-2015-12-RA-0178] [Medline: 27437055]
12. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795. [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
13. PatientsLikeMe. URL: https://www.patientslikeme.com/ [accessed 2016-09-30] [WebCite Cache ID 6ktqwoOLi]
14. Begum S, Ahmed MU, Funk P, Xiong N, Folke M. Case-based reasoning systems in the health sciences: a survey of recent trends and developments. IEEE Trans Syst Man Cybern C 2011 Jul;41(4):421-434. [doi: 10.1109/Tsmcc.2010.2071862]
15. Kitchenham B, Charters S, Budgen D, Brereton P. Guidelines for Performing Systematic Literature Reviews in Software Engineering: Technical report, Ver 23 EBSE Technical Report EBSE. 2007 Jul 09. URL: https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf [accessed 2016-10-02] [WebCite Cache ID 6kxyjH02W]
16. Jurisica I, Mylopoulos J, Glasgow J, Shapiro H, Casper RF. Case-based reasoning in IVF: prediction and knowledge mining. Artif Intell Med 1998 Jan;12(1):1-24. [Medline: 9475949]
17. Bobrowski L. Induction of similarity measures and medical diagnosis support rules through separable, linear data transformations. Methods Inf Med 2006;45(2):200-203. [Medline: 16538289]
18. Bobrowski L, Wasyluk H. Diagnosis supporting rules of the Hepar system. Stud Health Technol Inform 2001;84(Pt 2):1309-1313. [Medline: 11604939]
19. Park Y, Kim B, Chun S. New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis. Expert Systems 2006 Feb;23(1):2-20. [doi: 10.1111/j.1468-0394.2006.00321.x]
20. UCI Machine Learning Repository: Center for Machine Learning and Intelligent Systems. 2013. URL: http://archive.ics.uci.edu/ml/ [accessed 2016-09-30] [WebCite Cache ID 6ktr5KfZ4]
21. Saeed M, Mark R. A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations. AMIA Annu Symp Proc 2006:679-683 [FREE Full text] [Medline: 17238427]

22. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L, Moody G, et al. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. Crit Care Med 2011 May;39(5):952-960 [FREE Full text] [doi: 10.1097/CCM.0b013e31820a92c6] [Medline: 21283005]

23. Chattopadhyay S, Ray P, Chen H, Lee M, Chiang H. Suicidal risk evaluation using a similarity-based classifier. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). LNAI. 2008 Presented at: 4th International Conference on Advanced Data Mining and Applications; Oct 8-10, 2008; Chengdu, China p. 51-61. [doi: 10.1007/978-3-540-88192-6-7]

24. Sun J, Sow D, Hu J, Ebadollahi S. Localized supervised metric learning on temporal physiological data. 2010 Presented at: 20th International Conference on Pattern Recognition (ICPR); Aug 23-26, 2010; Istanbul, Turkey p. 23-26. [doi: 10.1109/ICPR.2010.1009]

25. Sun J, Sow D, Hu J, Ebadollahi S. A system for mining temporal physiological data streams for advanced prognostic decision support. 2010 Presented at: 2010 IEEE International Conference on Data Mining; Dec 13-17, 2010; Sydney, Australia. [doi: 10.1109/ICDM.2010.102]

26. David G. Generating evidence based interpretation of hematology screens via anomaly characterization. TOCCHEMJ 2011 Mar 1;4(1):10-16. [doi: 10.2174/1874241601104010010]

27. Houeland T. An efficient random decision tree algorithm for case-based reasoning systems. In: Proceedings of the 24th International Florida Artificial Intelligence Research Society. 2011 Presented at: 24th International Florida Artificial Intelligence Research Society; May 18-20, 2011; Palm Beach, FL p. 401-406.

28. Wang F, Sun J, Ebadollahi S. Integrating distance metrics learned from multiple experts and its application in patient similarity assessment. In: Proceedings of the 11th SIAM International Conference on Data Mining. 2011 Presented at: 11th SIAM International Conference on Data Mining; Apr 28-30, 2011; Mesa, AZ p. 59-70.

29. Wang F, Hu J, Sun J. Medical prognosis based on patient similarity and expert feedback. 2012 Presented at: 21st International Conference on Pattern Recognition (ICPR); Nov 11-15, 2012; Tsukuba Science City, Japan URL: http://europepmc.org/abstract/MED/25717413

30. Campillo-Gimenez B, Jouini W, Bayat S, Cuggia M. Improving case-based reasoning systems by combining k-nearest neighbour algorithm with logistic regression in the prediction of patients' registration on the renal transplant waiting list. PLoS One 2013 Sep;8(9):e71991 [FREE Full text] [doi: 10.1371/journal.pone.0071991] [Medline: 24039730]

31. Couchoud C, Stengel B, Landais P, Aldigier J, de Cornelissen F, Dabot C, et al. The renal epidemiology and information network (REIN): a new registry for end-stage renal disease in France. Nephrol Dial Transplant 2006 Feb;21(2):411-418 [FREE Full text] [doi: 10.1093/ndt/gfi198] [Medline: 16234286]

32. Gottlieb A, Stein GY, Ruppin E, Altman RB, Sharan R. A method for inferring medical diagnoses from patient similarities. BMC Med 2013 Sep;11:194 [FREE Full text] [doi: 10.1186/1741-7015-11-194] [Medline: 24004670]

33. Lowsky DJ, Ding Y, Lee DK, McCulloch CE, Ross LF, Thistlethwaite JR, et al. A K-nearest neighbors survival probability prediction method. Stat Med 2013 May 30;32(12):2062-2069. [doi: 10.1002/sim.5673] [Medline: 23653217]

34. Hielscher T, Spiliopoulou M, Volzke H, Kuhn J. Using participant similarity for the classification of epidemiological data on hepatic steatosis. In: Proceedings of the 27th IEEE International Symposium on Computer-Based Medical Systems. 2014 Presented at: 27th IEEE International Symposium on Computer-Based Medical Systems; May 27-29, 2014; New York p. 1-7. [doi: 10.1109/CBMS.2014.28]

35. Völzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011 Apr;40(2):294-307 [FREE Full text] [doi: 10.1093/ije/dyp394] [Medline: 20167617]

36. Zhang P, Wang F, Hu J, Sorrentino R. Towards personalized medicine: leveraging patient similarity and drug similarity analytics. AMIA Jt Summits Transl Sci Proc 2014;2014:132-136 [FREE Full text] [Medline: 25717413]

37. Henriques J, Carvalho P, Paredes S, Rocha T, Habetha J, Antunes M, et al. Prediction of heart failure decompensation events by trend analysis of telemonitoring data. IEEE J Biomed Health Inform 2015 Sep;19(5):1757-1769. [doi: 10.1109/JBHI.2014.2358715] [Medline: 25248206]

38. Harris M, Habetha J. The MyHeart project: a framework for personal health care applications. 2007 Presented at: Computers in Cardiology; Sep 30-Oct 3, 2007; Durham, NC. [doi: 10.1109/CIC.2007.4745440]

39. Lee J, Maslove DM, Dubin JA. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. PLoS One 2015 May;10(5):e0127428 [FREE Full text] [doi: 10.1371/journal.pone.0127428] [Medline: 25978419]

40. Ng K, Sun J, Hu J, Wang F. Personalized predictive modeling and risk factor identification using patient similarity. AMIA Jt Summits Transl Sci Proc 2015;2015:132-136 [FREE Full text] [Medline: 26306255]

41. Panahiazar M, Taslimitehrani V, Pereira NL, Pathak J. Using EHRs for heart failure therapy recommendation using multidimensional patient similarity analytics. Stud Health Technol Inform 2015;210:369-373 [FREE Full text] [Medline: 25991168]

42. Wang F. Adaptive semi-supervised recursive tree partitioning: the ART towards large scale patient indexing in personalized healthcare. J Biomed Inform 2015 Jun;55:41-54 [FREE Full text] [doi: 10.1016/j.jbi.2015.01.009] [Medline: 25656756]

43. Wang F, Sun J. PSF: a unified patient similarity evaluation framework through metric learning with weak supervision. IEEE J Biomed Health Inform 2015 May;19(3):1053-1060. [doi: 10.1109/JBHI.2015.2425365] [Medline: 25910264]

44. Chen X, Xu D, Zhang G, Mukkamala R. Forecasting acute hypotensive episodes in intensive care patients based on a peripheral arterial blood pressure waveform. In: Computers in Cardiology. 2009 Presented at: Computers in Cardiology; Sep 13-16, 2009; Park City, UT p. 13-16 URL: http://cinc.org/archives/2009/pdf/0545.pdf

45. Cai D, He X, Zhou K, Han J, Bao H. Locality sensitive discriminant analysis. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence. 2007 Presented at: 20th International Joint Conference on Artifical intelligence; Jan 6-12, 2007; Hyderabad, India p. 708-713.

46. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat 2008 Sep;2(3):841-860. [doi: 10.1214/08-AOAS169]

47. Wang F. Semisupervised metric learning by maximizing constraint margin. IEEE Trans Syst Man Cybern B 2011 Aug;41(4):931-939. [doi: 10.1109/Tsmcb.2010.2101593]

48. Hoi SC, Liu W, Chang S. Semi-supervised distance metric learning for collaborative image retrieval and clustering. ACM Trans Multimedia Comput Commun Appl 2010 Aug 01;6(3):1-26. [doi: 10.1145/1823746.1823752]

49. Perlman L, Gottlieb A, Atias N, Ruppin E, Sharan R. Combining drug and gene similarity measures for drug-target elucidation. J Comput Biol 2011 Feb;18(2):133-145. [doi: 10.1089/cmb.2010.0213] [Medline: 21314453]

50. Reddy CK, Li Y. A review of clinical prediction models. In: Reddy CK, Aggarwal CC, editors. Healthcare Data Analytics. Boca Raton, FL: CRC Press; 2015:343-378.

51. Wang X, Wang F, Wang J, Qian B, Hu J. Exploring patient risk groups with incomplete knowledge. 2013 Presented at: 13th IEEE International Conference on Data Mining; Dec 7-10, 2013; Dallas, TX. [doi: 10.1109/ICDM.2013.129]

52. Deza M, Deza E. Dictionary of Distances. Amsterdam: Elsevier; 2006.

53. Milligan GW, Cooper MC. Methodology review: clustering methods. Appl Psych Meas 1987 Dec 01;11(4):329-354. [doi: 10.1177/014662168701100401]

54. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996 Jul;22(7):707-710. [Medline: 8844239]

55. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. Crit Care Med 1984 Nov;12(11):975-977. [Medline: 6499483]

56. Kuhn M, Johnson K. Data pre-processing. In: Applied Predictive Modeling. New York: Springer-Verlag; 2013:27-59.

Abbreviations

APACHE-II: Acute Physiology and Chronic Health Evaluation II
ART: adaptive semisupervised recursive tree partitioning
CBR: case-based reasoning
DT: decision tree
EHR: electronic health data
ICU: intensive care unit
kd-tree: k dimensional tree
k-NN: k-nearest neighbor
LR: logistic regression
LSDA: locality sensitive discriminant analysis
LSR: local spline regression
PCA: principal component analysis
ROC: receiver operating characteristic
RSF: random survival forest
TF-IDF: term frequency–inverse document frequency

XSL•FO
RenderX

XSL•FO
RenderX

Original Paper

A Software Framework for Remote Patient Monitoring by Using Multi-Agent Systems Support

Chrystinne Oliveira Fernandes[1*], MSc; Carlos José Pereira De Lucena[1*], PhD

Department of Informatics, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil
[*]all authors contributed equally

Corresponding Author:
Chrystinne Oliveira Fernandes, MSc
Department of Informatics
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)
RDC Bldg, 4th Fl
225 Marquês de São Vicente St
Rio de Janeiro, 22451-900
Brazil
Phone: 55 521 3527 1510
Fax: 55 21 3527 1530
Email: chrystinne@gmail.com

Abstract

Background: Although there have been significant advances in network, hardware, and software technologies, the health care environment has not taken advantage of these developments to solve many of its inherent problems. Research activities in these 3 areas make it possible to apply advanced technologies to address many of these issues such as real-time monitoring of a large number of patients, particularly where a timely response is critical.

Objective: The objective of this research was to design and develop innovative technological solutions to offer a more proactive and reliable medical care environment. The short-term and primary goal was to construct IoT4Health, a flexible software framework to generate a range of Internet of things (IoT) applications, containing components such as multi-agent systems that are designed to perform Remote Patient Monitoring (RPM) activities autonomously. An investigation into its full potential to conduct such patient monitoring activities in a more proactive way is an expected future step.

Methods: A framework methodology was selected to evaluate whether the RPM domain had the potential to generate customized applications that could achieve the stated goal of being responsive and flexible within the RPM domain. As a proof of concept of the software framework's flexibility, 3 applications were developed with different implementations for each framework hot spot to demonstrate potential. Agents4Health was selected to illustrate the instantiation process and IoT4Health's operation. To develop more concrete indicators of the responsiveness of the simulated care environment, an experiment was conducted while Agents4Health was operating, to measure the number of delays incurred in monitoring the tasks performed by agents.

Results: IoT4Health's construction can be highlighted as our contribution to the development of eHealth solutions. As a software framework, IoT4Health offers extensibility points for the generation of applications. Applications can extend the framework in the following ways: identification, collection, storage, recovery, visualization, monitoring, anomalies detection, resource notification, and dynamic reconfiguration. Based on other outcomes involving observation of the resulting applications, it was noted that its design contributed toward more proactive patient monitoring. Through these experimental systems, anomalies were detected in real time, with agents sending notifications instantly to the health providers.

Conclusions: We conclude that the cost-benefit of the construction of a more generic and complex system instead of a custom-made software system demonstrated the worth of the approach, making it possible to generate applications in this domain in a more timely fashion.

XSL•FO

RenderX

Introduction

Innovative Technological-Based Solutions

Technological solutions can be applied to deal better with current operational problems involving the delivery of health care. The application of computational tools to hospital activities has the capacity to transform the present operational environment through activities such as improvement of work processes. Examples of improvements brought by the use of innovative technological solutions are as follows:

1. Change in the way the physician-patient–relationship occurs, because of Remote Patient Monitoring (RPM) possibilities [1]

2. Ease of information access and sharing among the medical team and the patients' relatives [2]

3. More mobility for patients, whose health status can be monitored from home or work, without being restricted to hospital facilities

4. Possibility of collaborative work between the local team and external professionals; it allows a second opinion about patients' diagnoses and treatments, as patient information is already in a distributed database

5. Possibility of automatic processes such as vital patient data collection by using sensors

6. Remote and real-time monitoring of patient health conditions

7. Alerts to health care professionals in emergency situations

8. Decrease in elapsed time for detection of anomalies in the vital signs of monitored patients, by using software agents; in this context, software agents consist of computational entities that perform activities in response to emergency situations

Investments in RPM technology can provide better support for patients from their health care team and perhaps make resources available for other health-related activities.

Theoretical Background

Internet of Things (IoT)

IoT is a field within Computer Science that has grown quickly in recent years. Kevin Ashton introduced the term "Internet of Things" in 1999 [3]. One can define IoT as a global network of smart devices that can sense and interact with their environment for communication with users and other things (smart devices)

and systems. In this context, things could be identified solely by using radio-frequency identification (RFID) [4] tags in order to be connected to the Internet and publish their information. Things are physical objects such as refrigerators, cars, walking sticks, dog collars, and whatever object comes to mind.

Thus, using sensors, actuators, and RFID-like technology, objects in the environment could be viewed, identified, and controlled more autonomously. In this case, things themselves could specify when they needed to be replaced, fixed, or report if they could provide data [3].

IoT Technologies: RFID, Microcontrollers, and Sensors

To develop the IoT patient-monitoring application described in this paper, 3 main IoT technologies have been used: RFID, micro-controllers, and sensors.

RFID is an automatic identification method that utilizes radio signals, recovering and storing data remotely through devices called RFID tags. These devices are used for identification, sensing, and communication [5].

Arduino [6] microcontrollers, which are open source platforms for electronic prototyping, are also used: Uno R3 [7] and Yún [8] models (Figure 1). Microcontrollers can be programmed to process inputs and outputs of connected external components (Figure 1). One can use embedded computing to allow the construction of systems that interact with the environment using hardware and software [9].

A variety of sensors can be used to collect data for IoT applications such as temperature, humidity, light level, oxygen level, and sensor presence, among others.

In eHealth, it is common for some devices to contain a number of sensors linked together, such as in the HealthPatch MD [10] Vital Connect health-monitoring sensor (Figure 2). The sensor is a small adhesive patch with a module that measures heart rate, breathing frequency, body temperature, posture, detection of falls, and also has Internet connectivity. Another example is the eHealth Sensor Platform Complete Kit [11]. It contains an eHealth Sensor Shield compatible with Arduino and Raspberry Pi [12] microcontrollers (Figure 2), plus 10 sensors to collect biometric data (Figure 2): pulse, oxygen levels in blood, airflow (breathing), body temperature, electrocardiogram (ECG), glucometer, galvanic skin response, blood pressure, patient position (accelerometer), and muscle or electromyography sensor (EMG).

Figure 1. Arduino micro-controllers, Uno R3 (on the left) and Yún (on the right) models.
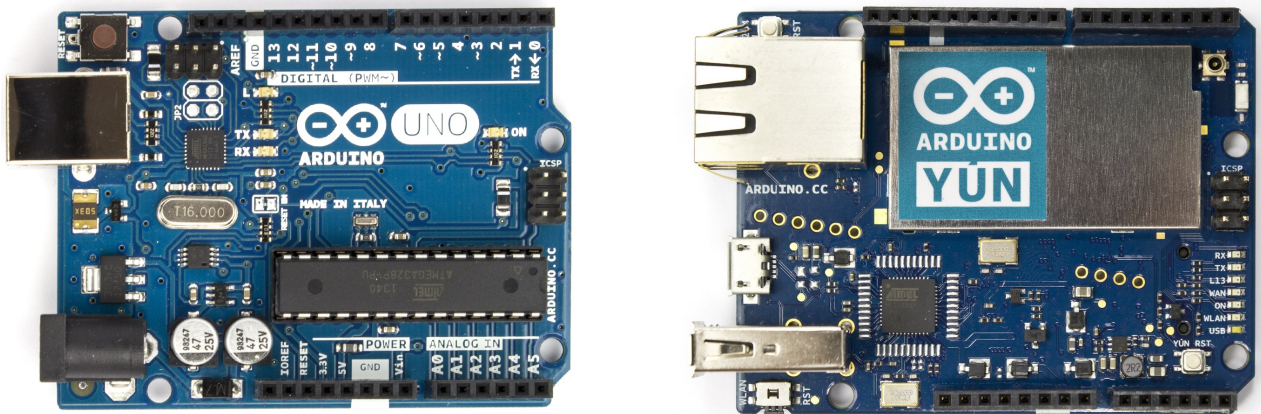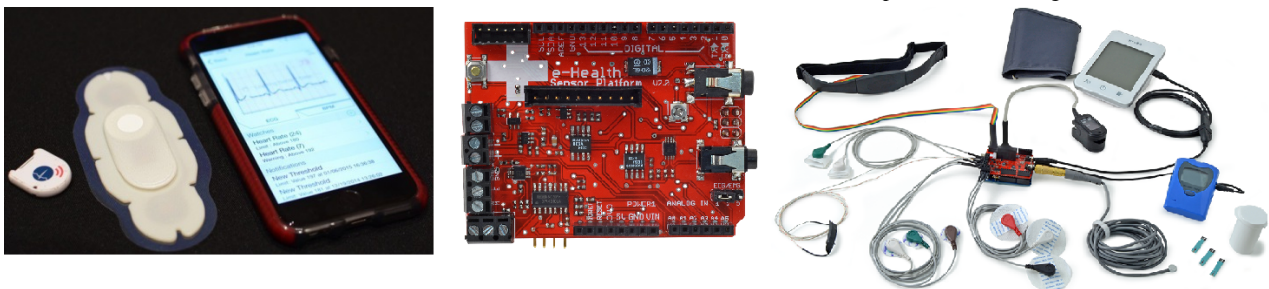

Figure 2. HealthPatch MD (on the left), e-Health Sensors Shield and e-Health Sensor Platform Complete Kit (on the right).

Software Agents

A software agent [13] is an element of a computational system that is situated in an environment where it can perform autonomous actions in order to reach its assigned goals. An agent is both autonomous and capable of learning from its experience. Autonomy has been acknowledged as a key characteristic of an agent in satisfying its goals [14]. In this context, autonomy means operating without the intervention of humans or other systems, although the set of possible actions should be previously defined.

Although agents control their behavior and internal states, they do not have full control of the environment in which they operate. Agents contain a set of actions that can carry out tasks, the execution of which can result in changes in the environments. For this reason, one can consider that an agent can have partial control and influence over its environment depending on the action performed [15].

In general, the use of software agents is justified by the fact that asynchronous software systems require autonomous operation, a general argument that can be applied to our solution.

General Concepts About Software Frameworks

Frameworks are tools used to generate applications related to a specific domain; that is, to cope with a family of related problems [16]. The choice of using existing frameworks or developing new application generators is based on whether the framework can offer design and code reuse. Thus, frameworks can usually increase software development productivity and shorter time-to-market, compared with traditional approaches.

Frameworks contain fixed and flexible points known as frozen spots and hot spots, respectively. Hot spots are extension points that allow developers to create a new application from the framework instantiation process. In this case, developers should create specific application code for each hot spot, through the implementation of abstract classes and methods defined in the framework. Frozen spots consist of the framework's kernel, corresponding to its fixed parts, previously implemented and hard to change. A frozen spot calls one or more of the application's hot spots and is present in each framework's instance [16].

Creating a new instance of a framework consists of 3 main steps: (1) Domain analysis, (2) Design, and (3) Instantiation. The domain analysis step includes requirements elicitation including definitions of hot and frozen spots. The design step is responsible for specifying the hot and frozen spots through a modeling language such as UML [17] diagrams. Design patterns [18] are also used in this phase. The instantiation phase corresponds to the application generation phase through hot spot implementation [16].

Related Work

Our proposal takes a similar approach to that in [19]. This paper shows the implementation of a distributed information infrastructure that uses the intelligent agent paradigm for: (1) automatically notifying the patient's medical team regarding the abnormalities in his or her health status; (2) offering medical advice from a distance; and (3) enabling continuous monitoring of a patient's health status. In addition, the authors have promoted the adoption of ubiquitous computing systems [20] and apps that allow immediate analysis of a patient's physiological data such as a personalized feedback of their

condition in real time, by using an alarm-and-remember mechanism. In this solution, patients can be evaluated, diagnosed, and cared for through a mode that is both remote and ubiquitous. In the case of rapid deterioration of a patient's condition, the system automatically notifies the medical team through voice calls or SMS messages, providing a first-level medical response. This proposal differs from ours, in that the resulting application is closed, as opposed to our broader eHealth application generator.

The approach in [21] focuses on design and development of a distributed information system based on mobile agents to allow automatic and real-time fetal monitoring. Devices such as a PDA, mobile phone, laptop, and personal computer are used to capture and display the monitored data.

In [22], mobile health apps are proposed as solutions for (1) overcoming personalized health service barriers; (2) providing opportune access to critical information on a patient's health status; (3) avoiding duplication of exams, delays and errors in patient treatment.

Methods

Main Research Goals

Our main research goal is to demonstrate that the formulation of a software framework to generate IoT applications in the eHealth domain does effectively support RPM. The aim is to analyze the tradeoffs involved in the challenge of building a flexible and powerful tool to help deal with the constraints found in a medical care environment. This initial version is totally experimental; it has not been tested in real medical care environments. Regarding the long-term goals of the research, the aim is to apply this software framework in a real medical care environment to assess its effective use as well as adequacy in terms of regulatory approval.

Methodology

We decided to build an IoT framework to allow the characterization of the RPM domain by using framework design techniques that encompass software agents. Framework methodology was chosen to assess its suitability for the RPM domain and its potential to generate customized applications that achieve the stated goals of more closely connecting patients to their health care team. As a proof of the concept, 3 applications were developed with different implementations for each hot spot of the framework.

An application named, Agents4Health was selected to illustrate the instantiation process and the IoT4Health framework operation. Furthermore, IoT devices were built from scratch to collect patient data for the Agents4Health application by using hardware prototypes comprised of biometric sensors and Internet-enabled microcontrollers to send the sensed data to the cloud automatically.

To measure the ability of the tool to respond proactively to adverse conditions such as anomalies in patients' vital signs, and its capacity to notify health providers in real time, the following step-by-step experiment was conducted:

1. Five measurement points were identified in the Agents4Health's workflow related to the tasks performed by agents and were labeled as Timestamps (T1 through T5) as follows:

T1. The Agents4Health application retrieves the patient data from the cloud and the monitoring agent analyses them, searching for anomalies. If no anomaly is detected, the system remains in a loop collecting more data until an anomaly is found. Once an anomaly is detected the application continues to T2

T2. This second step is reached when the monitoring agent detects an anomaly and then calls the notification agent.

T3. The notification agent initiates the routine to notify the health care providers;

T4. The notification agent sends information about the detected anomaly to the patient's health care providers;

T5. The health care providers receive the notification message on their mobile phones.

2. Agents4Health is executed and the timestamps are measured and registered.

3. Four delays defined as follows are captured for the different agent's execution tasks:

Detection anomaly interval (DAI)=T2−T1. The anomaly's detection delay in the monitoring routine.

Notification start interval (NSI)=T3−T2. The delay between the anomaly detection and the initiation of the notification routine.

Notification period (NP)=T4−T3. Duration of the notification routine by agents.

Notification routine interval (NRI)=T5−T4. Time elapsed between the sending of the notification and its receipt by the health provider.

These delays were calculated to serve as a concrete measure of how quickly and proactively the solution can respond to the environment, as well as to support the assertion that this system performs anomaly detection in real time.

To confirm the fulfillment of the main research goal, the experiment described above was conducted and the relevant results have been tabulated in the Discussion section.

Results

IoT4Health Framework

Domain Analysis

In this step, problems that health professionals currently deal with in their patient monitoring routines are considered. As mentioned earlier in this paper, the decision was made to build a software framework instead of one or more apps. The choice to use framework design techniques was motivated by the fact that the construction of a more generic and complex system would provide a cost-benefit, in that frameworks can usually increase software development productivity and shorter time-to-market.

IoT4Health Design

Regarding the IoT4Health's design, besides the hot and frozen spots that were modeled as UML diagrams, the following architecture was defined.

IoT4Health Architecture

The IoT4Health's architecture is structured in 3 layers, each with well-defined functionality (Figure 3).

Figure 3. The IoT4Health's architecture with its three layers (L1-L3). Data Management Layer (L3) and Distribution Data Layer (L1) interact through the Data Communication Layer (L2). The IoT4Health's frozen spots can be extended by hot spots.
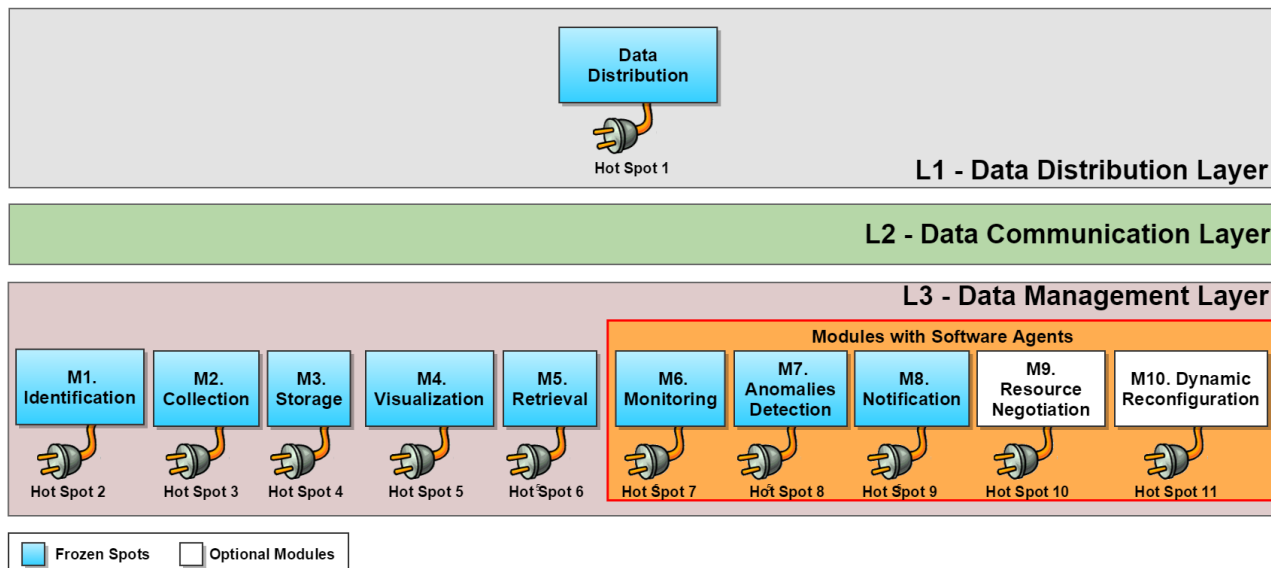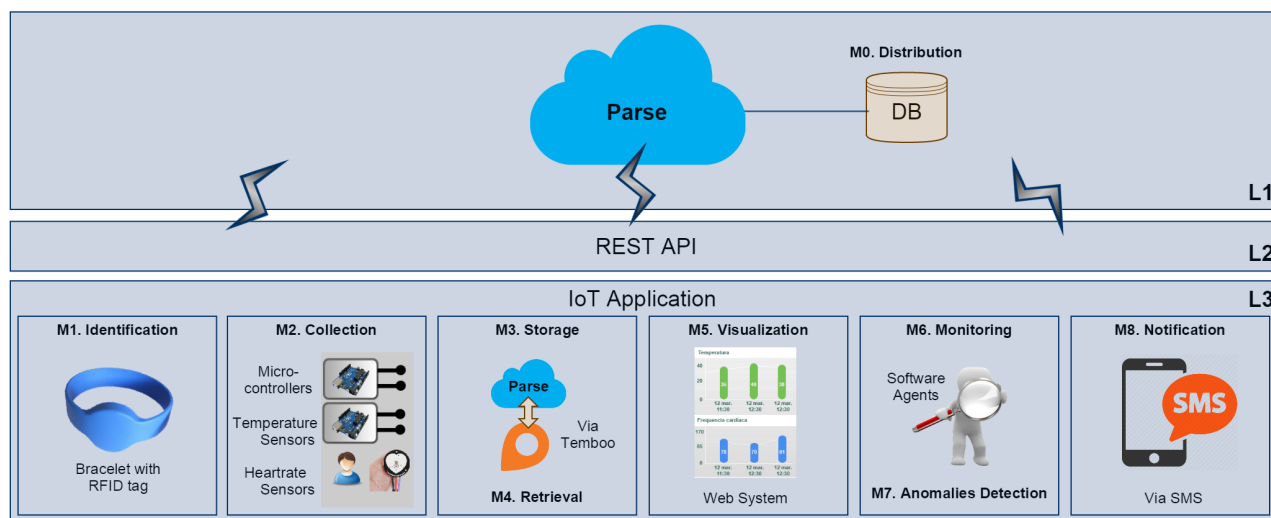

Figure 4. The Agents4Health's architecture with its three layers (L1-L3). The Agents4Health application (L3) interacts with Parse (L1) through REST API (L2).

L1: Data Distribution Layer

This layer works as a remote database of a patient's vital signs.

L2: Data Communication Layer

Through this layer, the L3 can communicate with the L1 layer's remote database.

L3: Data Management Layer

The L3 Layer is responsible for the entire information management of the instantiated applications. It is composed of 10 modules with well-defined responsibilities. These modules are shown next, along with a description of their purpose and

examples of how our framework can be extended by specific application code implemented in IoT4Health's instances.

Identification Module (M1)

It should be implemented by IoT4Health's instances to support the patient identification process. The IoT4Health's architecture offers the possibility of customizing this process, allowing developers to use different strategies, including: (1) The use of a unique identification code, such as the patient's ID Card; (2) utilization of RFID tags that can be inserted into objects like bracelets, cards, or other elements with radio-frequency identification capability; and (3) Biometry.

Collection Module (M2)

It provides the collection of both patient and his or her environment data. The collection process can be realized manually or automatically.

Storage Module (M3)

Its implementation lets the application store the vital collected data. Examples of storage strategies that can be developed as an extension of IoT4Health's architecture are (1) Local storage, (2) Cached storage, and (3) Remote storage.

Visualization Module (M4)

It was designed to provide users with ways to visualize storage data. Developers can implement some visualization strategies, utilizing the Web or a mobile application.

Recovery Module (M5)

It is responsible for recovering patient data stored on the cloud-based platform.

Monitoring Module (M6)

It was designed to continuously monitor the sensed data through software agents. Agents evaluate if the sensed data are within normal ranges, thus monitoring them to find anomalous values (AV). These normal ranges are defined for each patient, accounting for age, gender, other individual patient conditions, and each sensor in use. The system has a mandatory configuration step for each patient that can easily be completed by an administrator filling out a form through a system interface. In this step, the following parameters are defined:

1. Desired value range (DVR): They are the normal values collected from sensors; that is, values within an acceptable limit. They can correspond to an interval such as 36.0-36.6 for body temperature, for example. They should be defined for each sensor in use.

2. AVs: They are values outside the DVR, which are associated with anomalies. Regarding the DVRs from the previous example, one could have 37.8 as an example for a temperature AV.

3. Label of Anomaly: They are associated with the AVs and must also be defined for each sensor. Regarding data like temperature, it can be associated with the following anomalies: (1) Hyperthermia, for example, can be the label specified for anomalies associated with AVs higher than 36.6; (2) Hypothermia, to AVs lower than 36.0.

As one can observe, each such anomaly will receive a meaningful label regarding the health care context, so that it makes sense to a domain specialist. The goal is to enable a health care provider to identify quickly what problem is occurring when the system has detected an anomaly.

Anomalies Detection Module (M7)

It is supported by the use of reactive agents. This entity triggers alerts to health providers when case anomalies are detected.

Notification Module (M8)

It offers the possibility of using different strategies to send alerts to the medical team, such as by short message service (SMS),

email message, voice call, or by Bluetooth. This module also requires the configuration of some parameters, as follows:

1. Health provider responsible for an anomaly: a health professional should be selected to deal with each anomaly described.

2. Notification details: The type of message for each health care provider indicated previously should be specified (ie, SMS, email, voice call, or Bluetooth), along with the details such as email address or phone number.

This module's result is the communication process between agents. Agents that monitor patient data send a message to agents that send notifications when they detect an abnormality in the patient's condition, based on the predefined anomaly settings already mentioned.

Resource Negotiation Module (M9)

It utilizes the concept of cognitive agents that, in this context, would be responsible for the use of argumentation techniques [23] to achieve resource sharing in a collaborative way, by making its management more effective. An application could implement cognitive agents, responsible for adopting negotiation strategies, to obtain hospital resources for a particular patient.

Dynamic Reconfiguration Module (M10)

Its goal is to provide applications with context-sensitive capability so that these systems could be capable of responding to changes in the environment. A change of a patient's room could affect the defined parameters for monitoring, anomaly detection, and notification modules, becoming inappropriate in the new context. In this case, the applications' values must be reconfigured. This reconfiguration can be carried out manually by an administrator user or autonomously by cognitive agents.

Frozen Spots and Hot Spots

The IoT4Health contains 11 hot spots, offering developers the opportunity to create customized applications. Each one of these modules has extension points that broaden our framework's architecture, as shown above in Figure 3.

The Application Agents4Health as an Illustrative Instantiation of IoT4Health

The Agents4Health application [1] is an example of the IoT4Health's instantiation process, which was developed to illustrate the generative power of our framework Figure 4. It consists of a multi-agent system that autonomously conducts monitoring and notifying tasks. To access the patient data sensed by real biometric sensors and remotely stored through Arduino, the Agents4Health communicates with the cloud via REST application programming interface (API).

The Arduino integrated development environment (IDE) was used to implement the M2 and M3 modules of the Agents4Health in the C++ language. The other modules were created with the Java language. The software agents were programmed with the version 4.3.0 of the JADE tool [24]. JADE is a free software distributed by Telecom Italia (the copyright holder), in open source under the terms and conditions of the

XSL•FO
RenderX

second version of the Lesser General Public License (LGPL) license. It is a framework to develop agent systems in Java. It simplifies multi-agent systems' implementation through Foundation for Intelligent, Physical Agents (FIPA)-compliant middleware [25]. The JADE API offers 2 types of behavior classes that can be extended by agents: Primitive and Composite. The Agents4Health agents' behavior was implemented using the Primitive behavior class. Each application agent is an extension of the *Agent* Class and has a corresponding behavior to *Behavior's* extension class. The behavior of each of the system's agents is defined by its setup method, where behavior was configured through the *addBehavior* method.

The Agents4Health application's reactive agents present 2 types of behavior: *TickerBehavior* and *OneShotBehavior*. *TickerBehavior* type behavior is executed cyclically. That is, agents in our scenario that must carry out continuous monitoring activities implement this behavior, as is the case of *MonitoringSensorTemperatureDataAgent* class. Other agents are responsible for executing tasks that are not realized in a predefined interval of time, occurring only on demand in response to a specific event. This is the case of *NotificationBySMSAgent*, which sends messages to the medical team.

The Agents4Health Instance

L1: Data Distribution Layer

To provide the data distribution service to the application, a remote data storage service called Parse [26] was utilized.

However, because the Parse hosted service will be retired in early 2017, we are moving our database to another platform called MongoDB [27].

L2: Data Communication Layer

The Agents4Health communicates with Parse (L1) through the REST API [28]. The application sends and retrieves data to and from the cloud through HTTP requests.

L3: Data Management Layer

The data management layer comprises the IoT application, with its 8 modules (M1-M8) as follows:

M1

To identify a patient in the application, an RFID strategy has been chosen. We have used an RFID system that includes a tag and a reader. This process is performed through an RFID interface, where each patient receives a bracelet containing an RFID tag that will be used as a unique ID code in the system.

M2

In the Agents4Health application, both pulse and body temperatures are collected. This module may be extended to collect other patient data such as electroencephalography (EEG), EMG, as well as environment data such as light, noise levels, and data about the device such as battery status. This process, which is also called sensing, is performed automatically in Agents4Health. Arduino is used, together with sensors for heartbeat and temperature that form an IoT device capable of collecting patient data without human intervention (Figure 5).

Figure 5. Our IoT device for patient monitoring that contains an Arduino microcontroller and biometric sensors (on the left), Pulse and Oxygen in Blood Sensor (SPO2) sensors and Body Temperature sensor (on the right).
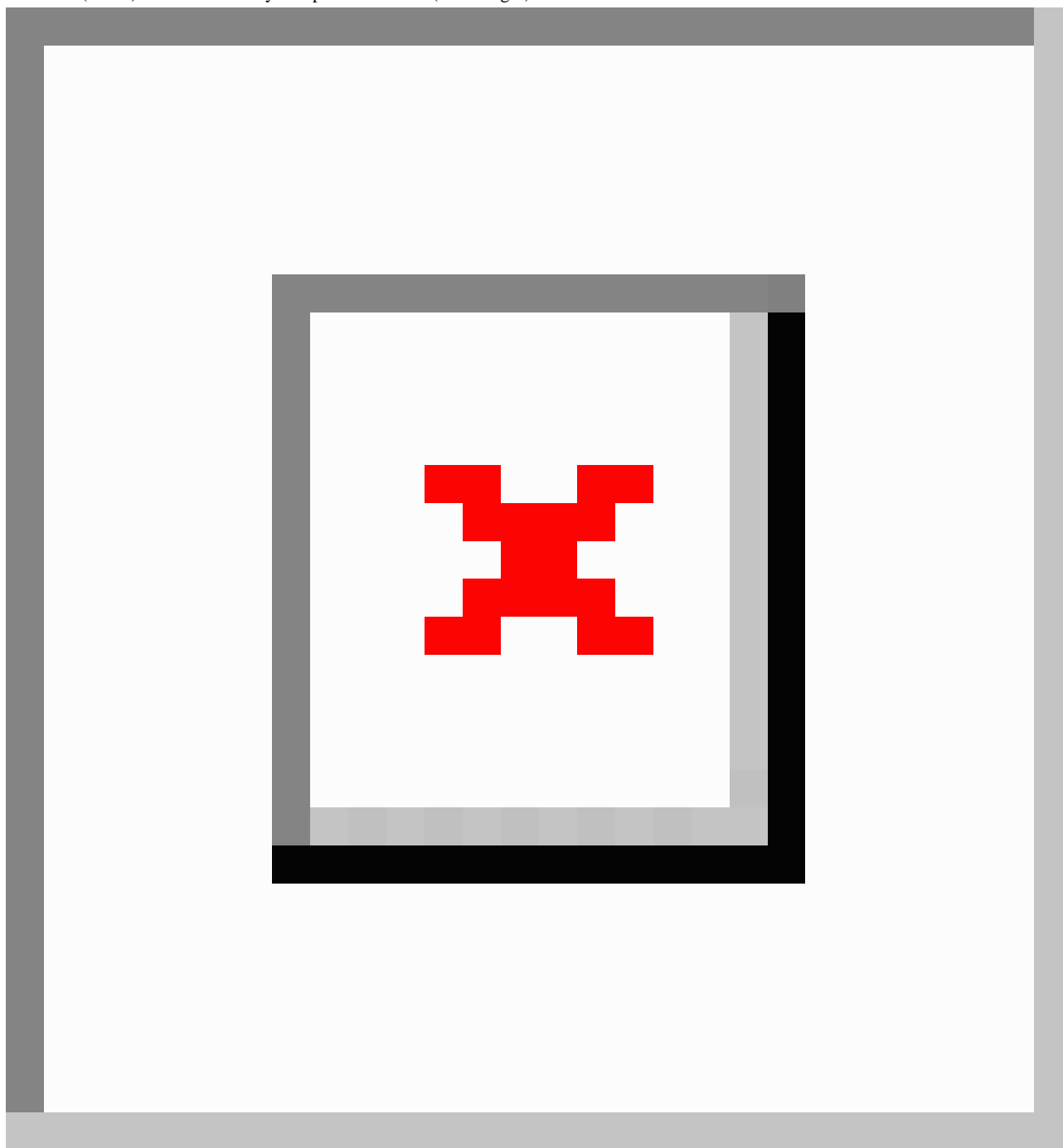
M3

Once collected, the application transfers the patient's data over the Internet to the Parse.

M4

The remote storage allows any authorized user to access the data by means of a user-friendly interface [1], through any device (computer, mobile phone, or tablet).

M5

A Web application is provided to support the visualization of the patient data. In the current implementation, there is a line chart for each one of the sensors used and they are updated in real time (Figure 6).

Figure 6. Example of the visualization module (on the left). Example of an action taken by the NotificationBySMSAgent agent (on the right).
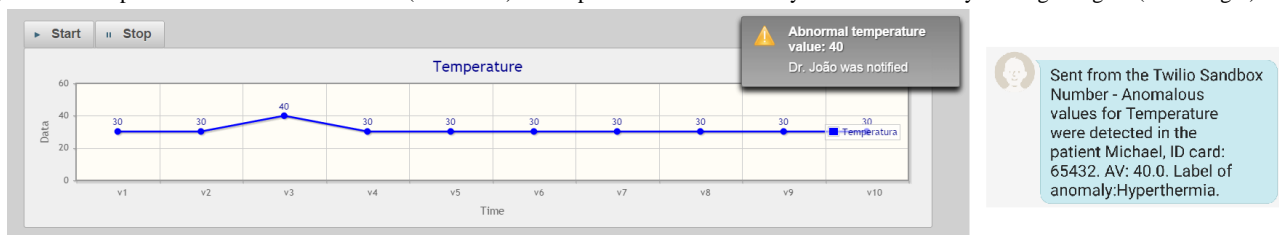
M6

Following the IoT4Health's protocol, in this step, an administrative user defines the DVR and the AV for each sensor.

M7

During this phase, a specific label for each anomaly is defined for each of the sensors: heartrate sensor (Table 1) and temperature sensor (Table 2).

Table 1. Configuring an example for the anomaly detection module, considering cardiac heartbeat.

AV[a] for cardiac heartbeat	Kinds of associated types of anomalies
Heartbeat values <60	Bradyarrhythmias such as sinus bradycardia or atrioventricular block
Heartbeat values >110	Tachyarrhythmias such as atrial fibrillation, supraventricular tachycardia, and ventricular tachycardia

[a]AV: anomalous value.

Table 2. Configuring an example for the anomaly detection module considering temperature.

AV[a] for temperature	Kinds of associated types of anomalies
Temperature values <36	Hypothermia
Temperature values >36.6	Hyperthermia

[a]AV: anomalous value.

In Agents4Health, the criteria used by the reactive agents to detect anomalies are defined by the domain specialists and coded in the XML language. They will form the agents' knowledge bases (Figure 7).

To prevent the system from detecting false abnormalities and triggering false alarms caused by simple patient movements or exercising, 2 strategies are being developed: (1) filtering the sensed data by using information provided by its own sensors related to the signal quality; (2) adding the environment's sensors to collect information about the context of the measurement. The former is performed when the sensors in use provide information about signal quality. Sensors such as the Mindwave Mobile Headset (NeuroSky) [29] are used to collect EEG data to provide this type of information. In this particular case, if the sensor is not in contact with the skin or if there is some interference such as a strand of hair between the sensor and the skin, the signal quality will indicate this situation. In that case, the application can be configured to ignore the sensed data until the signal quality provides a reliable value. The latter is useful to make the AVs flexible, taking into consideration the context of the patient being monitored. To avoid mistakenly detecting a heartbeat anomaly, for example, when a patient is engaged in physical activity, we can use sensors such as an accelerometer to collect context information.

M8: For Agents4Health, the choice was to send SMSs as a notification strategy, using the Twilio [30] library. Twilio is a platform using API communication that offers Web-service APIs, allowing users to construct their own SMS communication applications.

XSL•FO
RenderX

Figure 7. XML document that constitutes the agents' knowledge base.


```xml
<anomaly>
    <type>Hypothermia</type>
    <criteria type="comparison">
        <function symbol="<"/>
        <data>Temperature</data>
        <value>35</value>
    </criteria>
</anomaly>


<anomaly>
    <type>Bradyarrhythmia</type>
    <criteria type="comparison">
        <function symbol="<"/>
        <data>Heartrate</data>
        <value>50</value>
    </criteria>
</anomaly>
```

```xml
<anomaly>
    <type>Hyperthermia</type>
    <criteria type="comparison">
        <function symbol=">"/>
        <data>Temperature</data>
        <value>38</value>
    </criteria>
</anomaly>


<anomaly>
    <type>Tachyarrhythmias</type>
    <criteria type="comparison">
        <function symbol=">"/>
        <data>Heartrate</data>
        <value>110</value>
    </criteria>
</anomaly>
```

Discussion

Conclusions and Future Work

The main objective of this paper was to report on the construction of the IoT4Health framework based on a requirements analysis of the RPM domain. This was accomplished by using framework design techniques and the introduction of software agents in the framework design to allow autonomic behavior [31]. To deal with the aforementioned goal, experiments for the IoT4Health framework were created and tried.

In a main application experiment, generating a complete instance of the framework validated IoT4Health. This experiment, called Agents4Health, made it possible to observe that its design contributed toward making the patient's environment more proactive. Through this experimental system, it has also been possible to detect anomalies in real time and to send alerts instantly and autonomously to health providers. Thereby, professionals responsible for taking action in the case of abnormalities in a patient's condition can immediately react to these events.

As mentioned in the Methods section, Table 3 offers the results of the experiment conducted to offer a more concrete way to indicate the performance of the simulated environment:

Table 3. Examples of timestamps for agents' behavior and task delays.

Timestamp T1	Timestamp T2	Timestamp T3	Timestamp T4	Timestamp T5	DAI[a] (s)	NSI[b] (s)	NP[c] (s)
2016-11-04-173424	2016-11-04-173427	2016-11-04-173427	2016-11-04-173428	2016-11-04-1734	3	0	1
2016-11-04-173430	2016-11-04-173434	2016-11-04-173434	2016-11-04-173436	2016-11-04-1735	4	0	2
2016-11-04-173442	2016-11-04-173445	2016-11-04-173445	2016-11-04-173448	2016-11-04-1735	3	0	3
2016-11-04-173450	2016-11-04-173454	2016-11-04-173454	2016-11-04-173455	2016-11-04-1735	4	0	1
2016-11-04-173458	2016-11-04-173501	2016-11-04-173501	2016-11-04-173502	2016-11-04-1735	3	0	1
2016-11-04-173504	2016-11-04-173507	2016-11-04-173507	2016-11-04-173509	2016-11-04-1735	3	0	2

[a]DAI: detection anomaly interval.

[b]NSI: notification start interval.

[c]NP: notification period.

On average, the DAI for the Agents4Health experiment results is 3.5 s. The NSI presented zero delays for all results in this experiment. The NP averaged 1.75 s. And, finally, as mobile phones do not provide the SMS reception time with millisecond precision, there is only an approximate measurement for NRI, which was less than 1 min on average.

We have been involved in a number of practical developments based on our framework. One consists of the use of Bluetooth for communication with the medical team in the absence of

Internet access. The use of machine learning is also examined in the patient monitoring domain. This approach is performed by creating melanoma and mammography classifications as a black box accessible to the system agents. IoT4Health is also being used as the basis of a complex patient monitoring system under development in our laboratory [32], with our participation and which has been named portable care.

As future work, we are planning a rigorous formal characterization of the patient monitoring domain as well as the formal characterization of the family of applications reachable through the framework flexible points. The application of cognitive agents as elements of the software framework is also being considered.

Conflicts of Interest

None declared.

References

1. Fernandes CO, Lucena CJP. An Internet of Things Application with an Accessible Interface for Remote Monitoring Patients. In: Design, User Experience, and Usability: Interactive Experience Design. Springer International Publishing Switzerland: Springer International Publishing; 2015 Jul 06 Presented at: HCI International 2015; August 2-7, 2015; Los Angeles, CA, USA p. 651-661 URL: http://doi.org/10.1007/978-3-319-20889-3_60 [doi: 10.1007/978-3-319-20889-3_60]

2. Fernandes CO. Enabling a Smart and Distributed Communication Infrastructure in Healthcare. In: Lucena CJP, Lucena CAP, Azevedo BA, editors. Innovation in Medicine and Healthcare 2015. Springer International Publishing Switzerland: Springer International Publishing; Aug 12, 2015:435-446.

3. RFIDjournal. 2009. That 'Internet of Things' Thing - In the real world, things matter more than ideas URL: http://www.rfidjournal.com/articles/view?4986 [accessed 2016-08-22] [WebCite Cache ID 6jxDkcPuS]

4. Finkenzeller D, Cox MK. RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication. Third edition. Hoboken, US: Wiley; Jun 14, 2010.

5. Atzori L, Iera A, Morabito G. The Internet of Things: a survey. Computer Networks 2010 Oct;54(15):2787-2805. [doi: 10.1016/j.comnet.2010.05.010]

6. Arduino. URL: https://www.arduino.cc/ [accessed 2016-08-21] [WebCite Cache ID 6jwDCGjRi]

7. Arduino. Arduino UNO & Genuino UNO URL: https://www.arduino.cc/en/Main/ArduinoBoardUno [accessed 2016-10-22] [WebCite Cache ID 6lSEHmYSS]

8. Arduino. Arduino Yún Shield (USA Only) & Genuino Yún Shield (Outside USA) URL: https://www.arduino.cc/en/Main/ArduinoYunShield [accessed 2016-10-22] [WebCite Cache ID 6lSEaH2wB]

9. Doukas C. Building Internet of Things with the Arduino (Volume 1). USA: CreateSpace Independent Publishing Platform; 2012.

10. Engadget. 2015. HealthPatch MD alerts your doctor about heart problems in real time URL: https://www.engadget.com/2015/01/06/healthpatch-md-vital-connect/ [accessed 2016-08-22] [WebCite Cache ID 6jxH7bcek]

11. Cooking-hacks. e-Health Sensor Platform V2.0 for Arduino and Raspberry Pi [Biometric / Medical Applications] URL: https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical [accessed 2016-08-22] [WebCite Cache ID 6jxCIvFSO]

12. Raspberrypi. URL: https://www.raspberrypi.org/ [accessed 2016-08-22] [WebCite Cache ID 6jxI7NEyb]

13. Wooldridge M. Intelligent Agents. In: Multi-agent Systems. Cambridge, MA: The MIT Press; 1999.

14. Wooldridge MJ. An introduction to multiagent systems. Chichester, UK: John Wiley & Sons; 2009.

15. Russell SJ, Norvig P. Artificial intelligence: a modern approach. Upper Saddle River, NJ: Prentice Hall; 2010.

16. Markiewicz M, de Lucena CJ. Object oriented framework development. Crossroads 2001 Jul 01;7(4):3-9. [doi: 10.1145/372765.372771]

17. UML. What is UML URL: http://www.uml.org/ [accessed 2016-08-22] [WebCite Cache ID 6jxJ32ugU]

18. Gamma E, Helm R, Johnson R, Vlissides J. Design patterns: elements of reusable object-oriented software. Boston, MA: Addison-Wesley; 1995.

19. Su CJ, Wu C. JADE implemented mobile multi-agent based, distributed information platform for pervasive health care monitoring. Appl Soft Comput 2011 Jan;11(1):315-325. [doi: 10.1016/j.asoc.2009.11.022]

20. Weiser M. Some computer science issues in ubiquitous computing. Commun ACM 1993;36(7):75-84. [doi: 10.1145/159544.159617]

21. Su CJ, Chu TW. A mobile multi-agent information system for ubiquitous fetal monitoring. Int J Environ Res Public Health 2014 Jan 02;11(1):600-625 [FREE Full text] [doi: 10.3390/ijerph110100600]

22. Mohammadzadeh N, Safdari R. Patient monitoring in mobile health: opportunities and challenges. Med Arch 2014;68(1):57-60 [FREE Full text] [Medline: 24783916]

XSL•FO

RenderX

23.   Laera L, Blacoe I, Tamma V, Payne T, Euzenat J, Bench-Capon T. Argumentation over ontology correspondences in MAS.
      2007 May 14 Presented at: Aamas Conference 2007; May 14 - 18, 2007; Honolulu, Hawaii p. 228. [doi:
      10.1145/1329125.1329400]

24.   Jade.tilab. JAVA Agent DEvelopment Framework: an open source platform for peer-to-peer agent based applications URL:
      http://jade.tilab.com/ [accessed 2016-08-21] [WebCite Cache ID 6jwCwVQvX]

25.   FIPA. FOUNDATION FOR INTELLIGENT PHYSICAL AGENTS URL: http://www.fipa.org/docs/wps/f-wp-00018/
      f-wp-00018.html [accessed 2016-08-22] [WebCite Cache ID 6jxCpiSP5]

26.   Parse. Open Source Hub URL: http://parse.com/ [accessed 2016-08-21] [WebCite Cache ID 6jwBcbFYW]

27.   Mongodb. MongoDB Atlas: Hosted MongoDB-as-a-Service URL: https://www.mongodb.com/ [accessed 2016-11-04]
      [WebCite Cache ID 6ll1mFYUV]

28.   Parseplatform.github. Parse URL: http://parseplatform.github.io/docs/rest/guide/ [accessed 2016-08-22] [WebCite Cache
      ID 6jxLgHHwW]

29.   Neurosky. URL: http://neurosky.com/biosensors/eeg-sensor/biosensors/ [accessed 2016-11-04] [WebCite Cache ID
      6lm37mphP]

30.   Twilio. URL: https://www.twilio.com/ [accessed 2016-08-21] [WebCite Cache ID 6jwCh91Vd]

31.   Nunes IO, Kulesza U, Nunes C, Cirilo E, Lucena C. Extending web-based applications to incorporate autonomous behavior.
      New York, NY, USA: ACM; 2008 Presented at: 14th Brazilian Symposium on Multimedia and the Web (WebMedia '08);
      October 26 - 29, 2008; Vila Velha, Brazil p. 115-122 URL: http://dl.acm.org/citation.cfm?doid=1666091.1666112 [doi:
      10.1145/1666091.1666112]

32.   LES. 2016. URL: http://www.les.inf.puc-rio.br/wiki/index.php/P%C3%A1gina_principal [accessed 2016-08-22] [WebCite
      Cache ID 6jxZarAkW]

Abbreviations

API: Application Programming Interface

AV: anomalous values

DAI: detection anomaly interval

DVR: desired value range

ECG: electrocardiography

EEG: electroencephalography

EMG: electromyography

FIPA: Foundation for Intelligent Physical Agents

HTTP: Hypertext Transfer Protocol

IDE: integrated development environment

IoT: Internet of Things

JADE: Java Agent Development Framework

NP: notification period

NRI: notification routine interval

NSI: notification start interval

RFID: radio-frequency identification

SMS: short message service

UML: Unified Modeling Language

XML: eXtensible Markup Language

Original Paper

The Value of Electronic Medical Record Implementation in Mental Health Care: A Case Study

Sanaz Riahi[1,2], RN, MSN; Ilan Fischler[1,2], MD, FRCPC; Melanie I Stuckey[1], PhD; Philip E Klassen[1,2], MD, FRCPC; John Chen[1], Hon BA, MAcc

[1]Ontario Shores Centre for Mental Health Sciences, Whitby, ON, Canada

[2]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Sanaz Riahi, RN, MSN
Ontario Shores Centre for Mental Health Sciences
700 Gordon Street
Whitby, ON,
Canada
Phone: 1 905 430 4055
Fax: 1 905 430 4020
Email: riahis@ontarioshores.ca

Abstract

Background: Electronic medical records (EMR) have been implemented in many organizations to improve the quality of care. Evidence supporting the value added to a recovery-oriented mental health facility is lacking.

Objective: The goal of this project was to implement and customize a fully integrated EMR system in a specialized, recovery-oriented mental health care facility. This evaluation examined the outcomes of quality improvement initiatives driven by the EMR to determine the value that the EMR brought to the organization.

Methods: The setting was a tertiary-level mental health facility in Ontario, Canada. Clinical informatics and decision support worked closely with point-of-care staff to develop workflows and documentation tools in the EMR. The primary initiatives were implementation of modules for closed loop medication administration, collaborative plan of care, clinical practice guidelines for schizophrenia, restraint minimization, the infection prevention and control surveillance status board, drug of abuse screening, and business intelligence.

Results: Medication and patient scan rates have been greater than 95% since April 2014, mitigating the adverse effects of medication errors. Specifically, between April 2014 and March 2015, only 1 moderately severe and 0 severe adverse drug events occurred. The number of restraint incidents decreased 19.7%, which resulted in cost savings of more than Can $1.4 million (US $1.0 million) over 2 years. Implementation of clinical practice guidelines for schizophrenia increased adherence to evidence-based practices, standardizing care across the facility. Improved infection prevention and control surveillance reduced the number of outbreak days from 47 in the year preceding implementation of the status board to 7 days in the year following. Decision support to encourage preferential use of the cost-effective drug of abuse screen when clinically indicated resulted in organizational cost savings.

Conclusions: EMR implementation allowed Ontario Shores Centre for Mental Health Sciences to use data analytics to identify and select appropriate quality improvement initiatives, supporting patient-centered, recovery-oriented practices and providing value at the clinical, organizational, and societal levels.

XSL•FO
RenderX

Introduction

Electronic medical records (EMRs) have been adopted in health care facilities around the world with the purpose of providing clinical, organizational, and societal value [1]. According to the Health Information Management Systems Society (HIMSS) EMR adoption model, a hospital can range from stage 0, which denotes zero to very minimal electronic record function, to stage 7, signifying a complete EMR with data analytic function that is meaningfully used to improve care [2]. In Canada, EMR adoption is low with 62.7% of Canadian health care facilities at stages 2 to 3 compared with 63.0% of American facilities at stages 5 to 6 in 2015 [2]. Although beyond the scope of this paper, there are many differences between the American and Canadian health care systems, including funding and incentive structures, that explain these differences. There are a number of benefits to EMR systems that are not being realized in Canada due to lack of adoption. For example, EMRs provide organizational and societal value by increasing research ability, averting costs, improving legal and regulatory compliance, increasing career satisfaction, and monitoring public health [1]. Clinically, EMRs have the potential to improve health care delivery across the 6 dimensions of quality care (safety, effectiveness, efficiency, patient-centered, timely, equitable [3]), although most literature has reported outcomes related to safety, effectiveness, and efficiency [1,4-9]. There are also a number of disadvantages of EMRs, which are likely significant factors explaining the low adoption rates in Canada. These include the large upfront investment to cover costs, potential interruptions in workflow with system initiation, privacy concerns, limitations to physicians' autonomy, and unintended adverse outcomes [1,10]. Proper planning and support, however, can mitigate these challenges.

Ontario Shores Centre for Mental Health Sciences (Ontario Shores) implemented an EMR in 2009 and has since been committed to its meaningful use. In October 2014, Ontario Shores became the first Canadian hospital to achieve HIMSS Stage 7 designation and in October 2015, the first to achieve the HIMSS Nicholas E Davies Enterprise Award of Excellence (for the outstanding achievement of organizations who have used health information technology to substantially improve patient outcomes while achieving return on investment). Ontario Shores' EMR adoption journey is unique in that the EMR needed not only to provide the basic medical and business intelligence functions but also to be customized to suit the specific needs of mental health care. Most of the benefits of EMRs reported are from general health care settings. Mental health care facilities have unique needs for monitoring quality care. Perhaps most notably, rather than following a medical model of care, many mental health care facilities have adopted a recovery model of care [11-13], which is strongly focused on patient-centered care and involves collaborative treatment planning between patients and clinicians to support the patients' personal journeys toward achieving their individual goals. These goals may be related to symptom control or to promoting a meaningful, satisfying, hopeful, and contributing life, thereby minimizing limitations caused by illness. This case study reports on the implementation and customization of an EMR for a specialized, recovery-oriented mental health care facility. The objective of this evaluation was to determine the value that the EMR brought to the organization by examining the outcomes of quality improvement initiatives driven by the EMR.

Methods

Setting

Ontario Shores (Whitby, Ontario, Canada) is a public teaching hospital specializing in comprehensive mental health and addiction services for those with complex, serious, and persistent mental illness. The facility has 16 specialized inpatient units and extensive outpatient and community services serving a total regional population of approximately 2.8 million. The organization is staffed by approximately 1300 employees with 326 inpatients beds and approximately 60,000 annual outpatient visits. Ethical approval for this corporate evaluation study was not required by the institution as per Tri-Policy Statement guidelines.

Implementation

In 2007, Ontario Shores began its journey to implement a fully integrated EMR system. The project consisted of two main steps: first, the organization-wide implementation of the EMR and second, the EMR modules customized to support recovery-oriented mental health care practices and data used to identify and evaluate quality improvement initiatives. A number of evidence-informed strategies were incorporated in the implementation plan [14]. Additionally, it was recognized that staff behavior change would be important to the successful adoption of new practices. Hence, change models were incorporated in the implementation plan. At the onset, Kotter's change model [15] was followed to engage staff throughout the process. The Canada Health Infoway change management framework [16] was adopted when it was released.

Step 1: Electronic Medical Record Implementation

Overview

Readiness work prior to building and launching the EMR system was emphasized. The clinical informatics portfolio was strategically aligned under the professional practice umbrella to ensure that clinical practice was the focus of all functionality, design, and development of the EMR, such that the technology would enable but not drive practice. A governance model was created, which necessitated engagement of key clinical stakeholders including a physician champion [17]. A key success in the readiness phase and overall implementation was early and continuous engagement of clinical staff, who were involved in vendor and system selection (Meditech 6.0, Westwood, MA, USA) and working groups along with information technology, clinical informatics, and professional practice specialists. Working groups designed new documentation forms that conformed to evidence-based guidelines and future state processes. New workflows were initiated with paper-based forms prior to EMR implementation to allow staff to become comfortable with new processes using a familiar charting medium.

Staff Training

Basic computer skill training was offered to all clinical staff. Intensive EMR-specific training sessions, led by professional practice, were required for all nursing staff, allied staff, and physicians. Content was based on an "a day in the life of" concept and walked the clinicians through EMR documentation from the beginning to the end of shift. Real-time practice with a test patient was incorporated to increase understanding and retention. Super users for training and on-unit support were important facilitators for successful implementation.

Go Live

Go live occurred in "Big Bang" fashion for all departments as applicable modules were ready (Table 1). Go live for financial, admissions, pharmacy, material management, human resources, and staffing and scheduling modules occurred from October to December 2009. Go live for all inpatient services occurred in October 2010 and for all outpatient services in August 2011. All advanced clinical applications including computerized physician order entry, electronic medication administration record and bedside medication verification, plan of care, patient care system, imaging and therapeutic services, laboratory, and physician care manager were included in the initial launch.

Table 1. Timeline for electronic medical record implementation, quality improvement initiatives, and significant milestones.

Phase	Period	Milestone
Readiness phase		
	2007/08	Request for information vendor shortlist and request for proposal. Business case approved. Contract signed. Project resource plan developed and core team assembled.
	2007/08	Device selection. Vendor demonstrations and evaluation.
	September 2008	Readiness work.
	March 2009	Design of new clinical forms.
Implementation		
	August 2009	Completion of Advanced Clinical System Readiness Assessment.
	October 2009	Go live: financial, admissions, pharmacy, material management modules.
	December 2009	Go live: human resources, staffing and scheduling.
	March 2010	Implementation of paper forms with new workflows.
	Summer 2010	Basic computer skills training offered.
	September 2010	Inpatient clinical staff EMR[a]-specific training.
	October 2010	Inpatient go live: Go live with EMR advanced clinical applications, patient care system, imaging and therapeutic services, laboratory, and physician care manager.
	December 2010	Data repository functional.
	June 2011	Outpatient clinical staff EMR-specific training.
	August 2011	Outpatient go live: Go live with all applicable modules for outpatient services.
Customization and quality improvement		
	2011/12	Full system upgrade plus integration of the Resident Assessment Instrument—Mental Health and optimization of restraint minimization practices and documentation.
	2012/13	Achievement of HIMSS[b] stage 6 (June 2012). Plan of care optimization.
	2013/14	Full system upgrade plus implementation of integrated assessment record, business intelligence, and smoking cessation module. Optimization of outpatient and laboratory modules.
	2014/15	Achievement of HIMSS stage 7 (October 2014). Implementation of CPGs[c] for schizophrenia, level of care utilization system, front-end speech recognition, Ontario Common Assessment of Need, infection prevention and control surveillance status board; optimization of quality risk management. Launch of patient portal for patient engagement and self-management.
	2015/16 (to date)	Achievement of HIMSS Nicholas E Davies Enterprise Award of Excellence (October 2015). Technology-sharing partnership with other mental health facilities. EMR optimization through evaluation and enhancement of existing modules. Participation in health information exchange initiatives.

[a]EMR: electronic medical record.

[b]HIMSS: Health Information Management Systems Society.

[c]CPG: clinical practice guideline.

XSL•FO

RenderX

Step 2: Customization, Sustainability, Optimization, and Quality Improvement

In addition to the standard modules, a number of modules were created to support recovery-oriented care. Once modules were implemented, data mining was used to identify the need for quality improvement initiatives as described below.

Standard Modules

Closed Loop Medication Administration

The computerized physician order entry, electronic medical administration record, and bedside medication verification modules were standard and included in the initial launch. An adherence rate of 95% was needed in order to qualify for HIMSS stage 7 designation. An audit of bedside medication verification data showed that medication and patient scan rates were below this target. Letters of expectation communicating the standards of practice and importance of meeting target adherence levels in regard to safety were issued to all nursing staff whose adherence was lower than 90% over a 6-month period. A working group with point-of-care nurses was deployed to identify and remedy practice and process issues that were contributing to the low adherence rates.

Business Intelligence

Data mining has resulted in the evolution of a business analytics culture. With the implementation of a business intelligence tool, opportunities were created to build data models and engage in data analytics. Summary data for key performance indicators were provided on a dashboard with drill-through functionality. Detailed information related to program needs was accessible to leadership staff to help guide decision-making processes related to operations.

Mental Health Specific Modules

Resident Assessment Instrument—Mental Health

The Resident Assessment Instrument—Mental Health (RAI-MH), which is necessary for mandatory reporting to the ministry in the province of Ontario, Canada, was historically completed in another solution. With the implementation of the EMR, this mandatory assessment has been integrated in the EMR, and clinician daily documentation assists in completing this assessment. The RAI-MH is an interdisciplinary tool in the EMR with alerts to notify the clinical team of any missing measures and timely completion.

Plan of Care

The plan of care is a patient-centered treatment and discharge plan that informs care with the intention of ensuring a seamless transition and continuity of care between service providers. The patient, substitute decision maker, and family caregivers (if applicable) work collaboratively with the interprofessional team to develop the plan of care. The patient's values, strengths, goals, and vision are provided by the patient and family while the assessment, medical history, and behavioral profiles are added by the interprofessional team. The plan of care contains a number of themes that need to be addressed in mental health care (such as harm to self, illness management, exercise and nutrition, and leisure and education). On admission, relevant themes are selected and goals are developed in collaboration between the patient, relevant family, and clinicians.

Important to recovery-oriented treatment, each plan of care includes a recovery plan with 3 sections: patient story (subjective information gathered from the patient upon admission and on an ongoing basis to explore personal values, interests, cultural and religious practices, support systems, and personal views on hospitalization and treatment), crisis prevention plan (information regarding a patient's unique behavioral pattern, antecedents to maladaptive behaviors, de-escalation preferences, strategies, and intervention techniques), and sensory diet (sensory modalities that provide patients with information for recognizing and reducing their level of self-perceived distress and for modulating and learning to self-regulate their mood and behaviors).

Restraint and Seclusion

To align practices with recovery values, a policy supporting least restraint and seclusion was implemented [18]. Decision support was embedded directly within the physician orders for restraint and seclusion, and appropriate documentation of best practices were reflexed for clinical staff. Specifically, clinicians were alerted to complete a reassessment hourly for the duration of the event. Prompts were included to encourage clinicians to engage the patient with alternative interventions with the goal to reduce their duration of time in restraint or seclusion. Additionally, prompts were built in to remind clinicians to update patient de-escalation preferences in the plan of care following the event with the intention of preventing future incidents. A section was added to the module to document the patient and clinician debrief following the incident. Daily, monthly, and quarterly restraint data were reviewed by the senior management team, clinical leadership, and staff. Data were used to guide and facilitate meaningful, nonpunitive discussions with unit staff and to inform future practices aimed at minimizing restraint and seclusion use.

Clinical Practice Guidelines for Schizophrenia

Reflex orders and decision support were embedded in electronic documentation templates to prompt clinicians to follow evidence-based practices for the assessment and treatment of schizophrenia [19]. Since many standard practices at Ontario Shores were already aligned with the clinical practice guidelines (CPGs) for schizophrenia, the primary additions to the EMR were for antipsychotic prescribing (to promote adherence to antipsychotic monotherapy), metabolic monitoring, and referral to cognitive behavioral therapy for psychosis (CBT-P) and vocational rehabilitation. A tracking template was created for psychopharmacologic trials to document historical pharmacological trials, side effects, and outcomes with the intention of providing physicians with information that may help make decisions about antipsychotic prescribing. Exception handling was also added to require physicians to select a reason for ordering polypharmacy to prompt reflection and consideration for monotherapy. Since metabolic abnormalities are common side effects of antipsychotic medications, a reflex order set was created for metabolic monitoring triggered by the input of an order for antipsychotic medication. The order set included annual glucose and lipid panels and monthly blood

pressure, body weight, and waist circumference measurement as per guidelines. Reflex orders were also created to trigger referral to CBT-P and vocational rehabilitation services upon entering a diagnosis of schizophrenia.

To support adherence to CPG practices, adherence data were extracted from the EMR and shared with physicians, clinical practice leaders, leadership staff, and clinical teams using clinical scorecards built into a business intelligence tool. When adherence to guidance was suboptimal, quality improvement projects were initiated.

Infection Prevention and Control Surveillance Status Board

Due to an increasing number of outbreak days (ie, total number of days in which an outbreak was declared on any unit in the organization) at Ontario Shores, EMR functionality was leveraged to improve symptom surveillance and optimize the processes for initiating precautions and monitoring symptoms. The symptom surveillance assessment and communication tool was digitized and a field was added to allow nursing staff to initiate precautions electronically without consulting physicians. The infection prevention and control surveillance status board allowed infection prevention and control practitioners to monitor symptoms from any EMR-enabled workstation in the hospital, quickly identifying any need for follow-up.

Drug of Abuse Screening

In response to increased laboratory costs, a review of the quantity and expense of tests ordered was performed via the EMR. An expensive, broad-spectrum drug screen was being preferentially ordered rather than the drug of abuse screen. Discussion between the physician advisory group and the medical advisory committee established that the less expensive drug of abuse screen met clinical needs in the majority of cases. Thus, decision support was built into the EMR to support the physician order of the drug of abuse screen and protocols were added to identify the components of each test.

Patient Portal

In alignment with the recovery model, Ontario Shores' health check portal was implemented in 2014 to enhance patient access to their personal health information and promote patient activation and partnership between patients and clinicians. The portal was a built-in function of the EMR that was customized to Ontario Shores' needs. Full details have previously been reported [20]. Briefly, upon portal enrollment, patients were able to access parts of their medical record, view upcoming appointments, and communicate with their clinician from any device with Internet connectivity. Patients could choose to share information with family or community support workers.

Data Analysis

Monthly adherence to patient and medication scan rates was retrieved from the EMR from July 2013 to March 2015 (9 months of monitoring adherence prior to and 12 months of follow-up after quality improvement intervention). Monthly adherence to CPG practices (antipsychotic monotherapy, metabolic monitoring, and referral to CBT-P and vocational services) was retrieved from the EMR for the month prior to

and 12 months following CPG implementation (March 2014 to March 2015). Restraint and seclusion rates were used as a proxy for adherence to least restraint practices under the assumption that adherence to recovery-oriented practices would reduce the number of restraint and seclusion incidents. The number of restraint and seclusion incidents was retrieved from the EMR from April 2011 to March 2014. Finally, outbreak days were used as a proxy for adherence to infection prevention and control practices with the assumption that if procedures were followed, transmission rates, and, therefore, number of outbreak days would be reduced. Data were compiled from infection prevention and control documentation (April 2013 to March 2015), as these data were not entered into the EMR prior to implementation. For all other initiatives, anonymous summary data were retrieved from the corporate database. Descriptive analysis was used to examine changes in practice (Excel spreadsheet, Microsoft Corp). A crude return on investment summary was completed by documenting the initial investment, operational costs, and both hard and soft return on investments.

Results

Standard Modules: Closed Loop Medication Administration

Prior to the quality improvement initiative, patient and medication scan rates ranged from 80% to 95%. Following distribution of letters of expectation and follow-up with clinical managers (March 2014), scan rates were maintained above 95% (Figure 1). This was accompanied by a reduction in the number of moderate adverse drug events from 5 in 2011 (initial implementation) to 2 in each of 2012 and 2013 (prior to achievement of acceptable adherence rates) and 1 in 2014 (acceptable adherence rate).

Specialty Mental Health and Recovery Modules

Clinical Practice Guidelines

In the 12 months following CPG implementation, modest improvements were realized in CPG adherence [21]. Adherence to CBT-P and vocational rehabilitation guidance was increased from 6.5% to 11.4% and 36.6% to 49.1%, respectively. Adherence to antipsychotic monotherapy guidance increased initially from 53.4% to 62.7% but fell back to 55.1% by 12 months. Adherence to metabolic monitoring increased slightly from completing 76.7% of all required metabolic measurements to 81.6% [21].

Restraint and Seclusion

There was a 19.7% decrease in the number of restraint and seclusion incidents, a 42.3% decrease in the total restraint and seclusion hours, and a 38.9% decrease in the average hours per restraint or seclusion from the 2011/2012 fiscal year to the 2013/2014 fiscal year [18].

Custom Modules Developed by Ontario Shores: Infection Prevention and Control

There were 7 outbreak days in the year following infection prevention and control surveillance status board implementation compared to 47 outbreak days in the previous year (Figure 2).

Figure 1. Patient and medication scan rates. Closed squares, solid line: medication scan rate; open circles, dotted line: patient scan rate; solid horizontal line: target 95% adherence; dashed vertical line: implementation of quality improvement initiative.
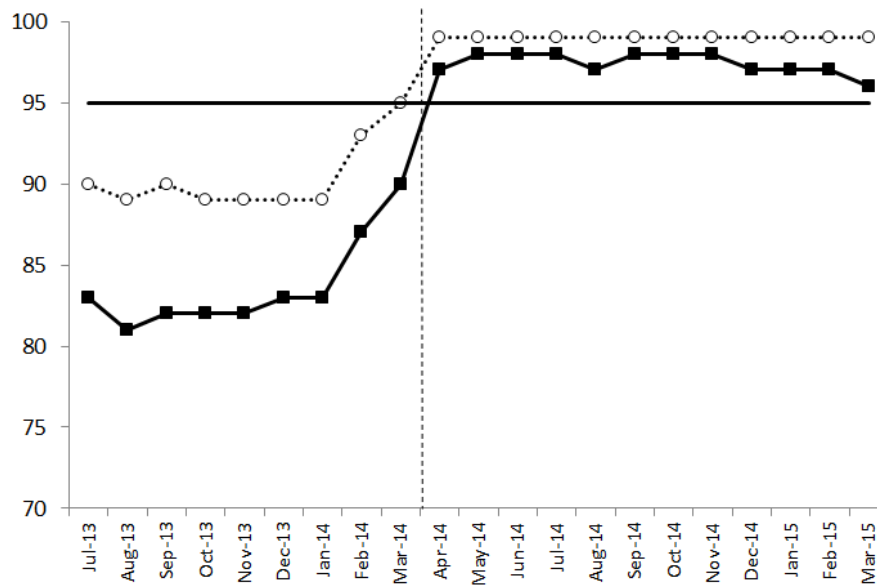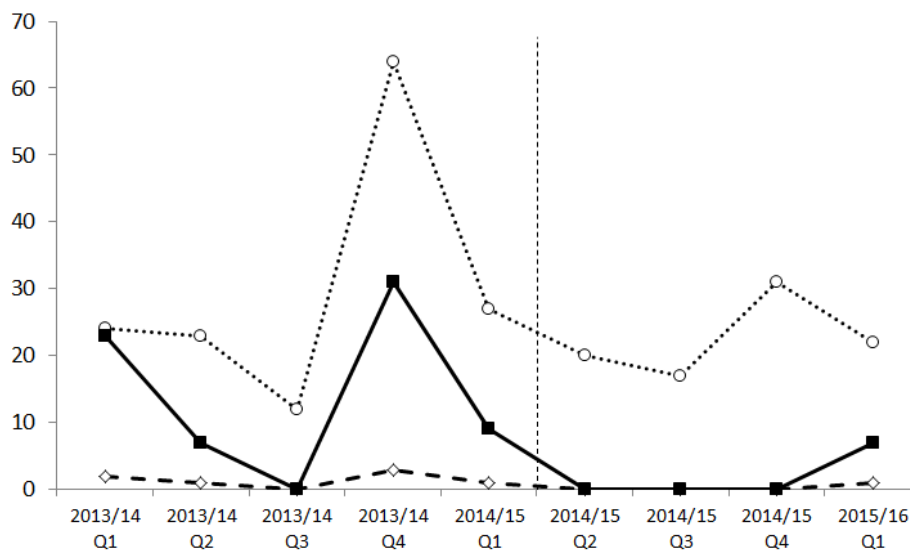

Figure 2. Number of precautions, outbreaks and outbreak days. Closed squares, solid line: outbreak days; open circles, dotted line: precautions; open diamonds, dashed line: outbreaks; dashed vertical line: implementation of infection prevention and control surveillance status board.

Return on Investment

As shown in Table 2, after EMR implementation in 2009/2010, costs were recovered by year 2013/2014. Much of the crude cost-savings were attributed to reduction in materials and salaries needed to support paper charting. Significant savings were also realized from the reduction in restraint and seclusion events following implementation of the restraint and seclusion modules in the EMR.

Table 2. Crude return on investment summary for electronic medical record implementation. All currency is presented in Canadian dollars.

	Initial	2010/11	2011/12	2012/13	2013/14	2014/15
Capital/one-time investment/implementation						
Initial investment and implementation cost	6,546,274	—	—	—	—	—
Staffing	4,135,900	—	—	—	—	—
Total initial investment	10,682,174	—	—	—	—	—
Operational costs (incremental)						
Incremental IT/CI[a] staffing	—	563,768	871,960	881,933	873,036	864,333
Maintenance and support—EMR[b]	—	173,093	289,059	270,811	329,293	333,388
Incremental hardware, software, and licensing	—	24,864	14,215	77,512	17,264	428,481
Total expenses	—	761,725	1,175,234	1,230,256	1,219,594	1,626,202
Hard return on investment						
Savings/efficiencies						
HIM[c] paper chart savings	—	15,000	20,000	25,000	25,000	25,000
HIM chart control staff Reduction	—	—	185,865	200,630	209,831	341,885
Transcription service savings	—	—	101,280	107,955	97,969	118,071
Lab test utilization	—	—	—	—	16,372	16,372
Reduced clerical and admin overhead	—	150,000	153,000	156,060	159,181	162,365
Reduction in medication cost due to unit dose and utilization	—	196,078	200,000	204,000	208,080	212,242
Staff savings from restraint and seclusion prevention	—	—	—	—	776,633	666,957
Reduction in antipsychotics due to CPG[d]	—	—	—	—	—	12,240
Revenue/incentives						
Grants	—	—	—	—	50,000	216,563
Soft return on investment						
Cost savings/avoidance						
Adverse drug event	—	—	—	38,100	38,100	50,800
Annual benefits	—	361,078	660,145	731,745	1,581,166	1,822,494
Total annual cash flow	(10,682,174)[e]	(400,647)[e]	(515,0789)[e]	(498,511)[e]	361,573	196,293

[a]IT/CI: information technology/clinical informatics.

[b]EMR: electronic medical record.

[c]HIM: health information management.

[d]CPG: clinical practice guideline.

[e]Parentheses indicate deficit.

Discussion

Overview

Implementation of a fully integrated EMR and the use of data analytics to identify quality improvement projects provided value at the clinical, organizational, and societal levels. Customization for mental health care facilitated the adoption of recovery-oriented practices. One of the strengths of this project was that corporate data were used to capture the outcomes of the relevant population in its entirety rather than consenting a subset of participants, which may have introduced selection bias because more motivated and well patients would

be more likely to volunteer to participate. Additionally, this project incorporated a number of implementation facilitators identified in the literature, such as active involvement and support of management, inclusion of clinical staff in the implementation and decision-making processes, end-user training and real-time support, and the identification of physician champion and point-of-care staff champions to provide support to peers [14]. Importantly, it was recognized that implementation was not purely a technical project and that behavior change of staff would be a critical element to the success of the project [14]. Relevant change models or frameworks were embedded in the implementation plan to guide the behavior change process [15,16].

Comparison With Prior Work

The integration of clinical decision support tools into standard workflows was an important facilitator to the adoption of new workflows [6-9] and has been identified as particularly important to the adoption of evidence-based medicine [8,22]. Our CPG for schizophrenia adoption strategy relied heavily on clinical decision support. Adherence to practices, however, was not greatly changed, with the exception of referrals to CBT-P and vocational rehabilitation services [21]. Adherence to metabolic monitoring, for example, only increased slightly from 76.7% in the month prior to CPG implementation to 81.6% at 1-year follow-up [21]. This is in contrast to a previous study, which showed that when clinical decision support to alert the need for metabolic monitoring was used, the odds were 3.51 times greater that clinicians would complete the recommended measurements compared to a control group [7]. Differences may be due to baseline adherence, which was only around 3% [7] compared to our baseline of 76.7% [21]. As Ontario Shores adopts and integrates other CPGs within the EMR (eg, dementia, depression), appropriate coordination and use of clinical decision support will be required to maintain patient safety and a patient-centered approach to care. Barriers contributing to the failure of clinical decision support to aid in adoption include perception of usefulness and over-triggering [8]. Thus, careful planning and consideration is needed to optimally integrate clinical decision support in the EMR as functions become more complex. Indeed, it has been identified as an important component to appropriately integrate patient-centered approaches, combinations of CPGs, self-management interventions, and continuity of care in patients with multiple morbidities [23]. These may be especially important in recovery-focused mental health care to support patients who may have multiple diagnoses—both psychological and physical—and be working toward achieving individualized goals on their recovery journey.

Careful implementation of the EMR and additional modules resulted in clinical, organizational, and, potentially, societal value. Clinical value was seen across the 6 dimensions of quality care [3]. Notably, patient safety was improved through the reduction in restraint, seclusion, and associated adverse events; reduced infection transmission; and the prevention of medication errors from reaching the patient. Other studies examining the impact of EMRs in acute and general health care settings have shown that well-planned and implemented systems have reduced mortality and medical complications [4] but simple

implementation has shown mixed results [1]. As in this study, others have demonstrated reduced medical errors and redundant diagnostic tests [5,24,25], thereby increasing treatment effectiveness and care efficiency along with safety. Implementation of CPGs ensured that all patients diagnosed with schizophrenia were provided the full spectrum of evidence-based assessments and treatments [6,7,9], ensuring the timely and equitable delivery of services while also contributing to improvements in the 3 aforementioned dimensions of quality care.

A number of initiatives implemented through the EMR to improve patient-centered care are poorly described in the literature but especially important to recovery-oriented mental health care [11-13]. The plan of care was implemented to lead clinicians through a method of collaborative treatment planning with patients and their families. In accordance with recovery principles, patient goals, strengths, and treatment preferences were documented with the intention of using those at the basis for planning care. The plan of care is reviewed once per shift to encourage staff to consider patients individually throughout their treatment. Modules for restraint and seclusion included patient de-escalation preferences and reflex orders to ensure timely debriefing following events with the intention of repairing the therapeutic relationship, which is often compromised [26]. Restraint data were reviewed extensively and informed initiatives implemented as Ontario Shores strove to minimize restraint and seclusion as per best practice guidelines. Ontario Shores' health check portal was another important patient-centered initiative enabled by the EMR. It promoted patient activation by enabling patients to access their personal health information. A full benefits evaluation of the portal was previously reported [20] that showed improvements in patient activation and recovery over the year following implementation. Notably, a subset of patients who responded to a survey reported an increased sense of autonomy [20], which is an important component to mental health recovery [11-13]. Indeed, the EMR has played an important role in the incorporation of recovery-oriented services into practice and will remain a central component of future initiatives to promote patient-centered recovery.

Organizational and societal value was seen through the standardization of workflows, facilitating the fulfillment of legal and regulatory obligations. A benefits evaluation of the patient portal showed improved administrative efficiencies and productivity in the year following compared to the year prior to portal implementation [20]. Internally, the EMR data were presented to highlight trends, providing information for clinicians to support informed clinical decision making at the direct care level. Additionally, discrete data created opportunities to evaluate adherence to and effectiveness of treatment and to select appropriate quality improvement initiatives. An overall culture change has been realized along with greater efficiencies for clinical and nonclinical roles. Furthermore, the potential of the EMR to support research has been recognized, and it will play a strong role in the future.

It should be noted that a significant upfront monetary investment was needed. Similar to other studies, cost benefits were not realized immediately [27] but were seen after the EMR data

XSL·FO

RenderX

were used to identify the need for and support the implementation of quality improvement initiatives. Most of the direct cost savings (Table 2) are a result of reduced costs in the health information management department, including transcription cost savings exceeding Can $425,000 since 2011/2012 and the reduction of supplies associated with paper filing. Paper communications with external entities and patients have been reduced through use of the health information exchange and health check portal, respectively, decreasing the preparation time for release of information. Many manual processes related to the collection of data for internal and external reporting, such as wait time data and workload statistics, have been eliminated, and the turnaround time has significantly decreased. Cost reductions were realized with decreased restraint and seclusion events and reduced antipsychotic medication prescribing (Table 2). It should be noted that a simple return on investment calculation for this mental health facility may underestimate the full benefit of EMR implementation. For example, while implementation of metabolic monitoring according to CPGs provides no cost savings to Ontario Shores, the anticipated prevention or delay of onset of cardiovascular diseases and type 2 diabetes mellitus has the potential to reduce the economic burden on the Ontario health care system. The EMR also enables health information exchange initiatives to integrate electronic patient information from across the care continuum to improve timely access at the point of care, ultimately improving the patient and clinician experience and reducing overall costs to the health care system.

The EMR has enabled a number of new initiatives at Ontario Shores. Recently (November 2016), a partnership was formed with another mental health facility to create a shared EMR using a single system and database. The purpose of this partnership is to continue to improve clinical outcomes, advance evidence-based practices and clinical standards, advance mental health research, and realize cost effectiveness and efficiencies. Future evaluation and study will be needed to determine the effectiveness of this partnership, which may set the stage for increased technology-sharing partnerships to increase EMR adoption in Canada.

Limitations

Despite the above-mentioned strengths, this project was not without limitations. Since EMR implementation and modules were designed to optimize recovery-based practices at a specialized, tertiary care mental health facility, results may not be generalizable to all other settings. However, the framework and examples should provide a guideline for leveraging EMR functionality to optimize practices. Since this case study reports organizational practices, outcomes were limited to metrics available through corporate data. As such, many outcomes are proxy measures of adherence to practices. Explanatory variables such as staff perspectives were not collected and could not be included in this study. As Ontario Shores is continually striving to improve the quality of care, a number of concurrent non–EMR-based projects were initiated during the same time period. Therefore, some of the improvements may not have been directly related to the EMR or may have been further improved by other initiatives. Clear benefits of the EMR have, however, been described.

Conclusions

This case report described the implementation of an EMR system in a tertiary-level mental health care facility to increase clinical, organizational, and societal value. Customization for mental health services enabled the provision of recovery-oriented services to patients overall improving delivery across the 6 dimensions of quality care. Future initiatives to further leverage the EMR potential for evaluation, quality improvement, research, and collaboration are underway.

Authors' Contributions

SR, IF, PK, and JC conceptualized the study and were involved in the design and implementation. MS interpreted data and drafted the manuscript. All authors critically reviewed the manuscript and approved the final version.

Conflicts of Interest

All authors were employed by Ontario Shores Centre for Mental Health Sciences at the time of this study. Neither financial incentives nor conditions of employment were related to outcomes of this study.

References

1. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy 2011;4:47-55 [FREE Full text] [doi: 10.2147/RMHP.S12985] [Medline: 22312227]
2. Health Information Management Systems Society. Provider Solutions. URL: http://www.himssanalytics.org/provider-solutions [accessed 2016-11-25] [WebCite Cache ID 6mHtpAD1U]
3. Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Institute of Medicine 2001. [Medline: 25057539]

XSL•FO
RenderX

4.   Amarasingham R, Plantinga L, Diener-West M, Gaskin DJ, Powe NR. Clinical information technologies and inpatient outcomes: a multiple hospital study. Arch Intern Med 2009 Jan 26;169(2):108-114. [doi: 10.1001/archinternmed.2008.520] [Medline: 19171805]

5.   Bates DW, Kuperman GJ, Rittenberg E, Teich JM, Fiskio J, Ma'luf N, et al. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. Am J Med 1999 Feb;106(2):144-150. [Medline: 10230742]

6.   Janssen B, Ludwig S, Eustermann H, Menke R, Haerter M, Berger M, et al. Improving outpatient treatment in schizophrenia: effects of computerized guideline implementation—results of a multicenter study within the German research network on schizophrenia. Eur Arch Psychiatry Clin Neurosci 2010 Feb;260(1):51-57. [doi: 10.1007/s00406-009-0016-2] [Medline: 19876665]

7.   Lai C, Chan H, Pan Y, Chen C. The effectiveness of a computer reminder system for laboratory monitoring of metabolic syndrome in schizophrenic outpatients using second-generation antipsychotics. Pharmacopsychiatry 2015 Jan;48(1):25-29. [doi: 10.1055/s-0034-1394396] [Medline: 25350604]

8.   McGinn T. Putting meaning into meaningful use: a roadmap to successful integration of evidence at the point of care. JMIR Med Inform 2016;4(2):e16 [FREE Full text] [doi: 10.2196/medinform.4553] [Medline: 27199223]

9.   Sheldon TA, Cullum N, Dawson D, Lankshear A, Lowson K, Watt I, et al. What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. BMJ 2004 Oct 30;329(7473):999 [FREE Full text] [doi: 10.1136/bmj.329.7473.999] [Medline: 15514342]

10.   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]

11.   Jacobson N, Greenley D. What is recovery? A conceptual model and explication. Psychiatr Serv 2001 Apr;52(4):482-485. [doi: 10.1176/appi.ps.52.4.482] [Medline: 11274493]

12.   Anthony WA, Blanch A. Research on community support services: what have we learned. Psychosoc Rehabil J 1989;12(3):55-81. [doi: 10.1037/h0099533]

13.   Oades L, Deane F, Crowe T, Lambert WG, Kavanagh D, Lloyd C. Collaborative recovery: an integrative model for working with individuals who experience chronic and recurring mental illness. Australas Psychiatry 2005 Sep;13(3):279-284. [doi: 10.1111/j.1440-1665.2005.02202.x] [Medline: 16174202]

14.   Boonstra A, Versluis A, Vos JFJ. Implementing electronic health records in hospitals: a systematic literature review. BMC Health Serv Res 2014;14:370 [FREE Full text] [doi: 10.1186/1472-6963-14-370] [Medline: 25190184]

15.   Kotter J. Leading Change. Boston: Harvard Business School Press; 1996.

16.   Canada Health Infoway. A Framework and Toolkit for Managing eHealth Change: People And Processes. 2013. URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/toolkits/change-management/563-managing-ehealth-change-people-and-processes [accessed 2016-12-27] [WebCite Cache ID 6n4s1JR7q]

17.   Macdonald A, Riahi S. EHR implementation: one organization's road to success. Nurs Inform 2012;2012:258 [FREE Full text] [Medline: 24199098]

18.   Riahi S, Dawe IC, Stuckey MI, Klassen PE. Implementation of the six core strategies for restraint minimization in a specialized mental health organization. J Psychosoc Nurs Ment Health Serv 2016 Oct 1;54(10):32-39. [doi: 10.3928/02793695-20160920-06] [Medline: 27699424]

19.   National Institute for Clinical Excellence. Psychosis and schizophrenia in adults: prevention and management. URL: https://www.nice.org.uk/guidance/cg82 [accessed 2016-11-26] [WebCite Cache ID 6mIc0oIcE]

20.   Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A Web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]

21.   Fischler I, Riahi S, Stuckey MI, Klassen PE. Implementation of a clinical practice guideline for schizophrenia in a specialist mental health center: an observational study. BMC Health Serv Res 2016;16(a):372 [FREE Full text] [doi: 10.1186/s12913-016-1618-9] [Medline: 27514778]

22.   Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 1;90(5):357-364 [FREE Full text] [doi: 10.2471/BLT.11.099424] [Medline: 22589569]

23.   Fraccaro P, Arguello CM, Ainsworth J, Buchan I. Adoption of clinical decision support in multimorbidity: a systematic review. JMIR Med Inform 2015 Jan 07;3(1):e4 [FREE Full text] [doi: 10.2196/medinform.3503] [Medline: 25785897]

24.   Bates DW, Leape LL, Cullen DJ, Laird N, Petersen LA, Teich JM, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA 1998 Oct 21;280(15):1311-1316. [Medline: 9794308]

25.   Bates DW, Teich JM, Lee J, Seger D, Kuperman GJ, Ma'Luf N, et al. The impact of computerized physician order entry on medication error prevention. J Am Med Inform Assoc 1999;6(4):313-321 [FREE Full text] [Medline: 10428004]

26.   Huckshorn KA. Reducing seclusion restraint in mental health use settings: core strategies for prevention. J Psychosoc Nurs Ment Health Serv 2004 Sep;42(9):22-33. [Medline: 15493493]

27.   Jang Y, Lortie MA, Sanche S. Return on investment in electronic health records in primary care practices: a mixed-methods study. JMIR Med Inform 2014;2(2):e25 [FREE Full text] [doi: 10.2196/medinform.3631] [Medline: 25600508]

Abbreviations

CBT-P: cognitive behavior therapy for psychosis
CI: clinical informatics
CPG: clinical practice guideline
EMR: electronic medical record
HIM: health information management
HIMSS: Health Information Management Systems Society
IT: information technology
Ontario Shores: Ontario Shores Centre for Mental Health Sciences
RAI-MH: Resident Assessment Instrument—Mental Health

XSL•FO
RenderX

<u>Original Paper</u>

Checking Questionable Entry of Personally Identifiable Information Encrypted by One-Way Hash Transformation

Xianlai Chen[1], PhD; Yang C Fann[2], PhD; Matthew McAuliffe[3], PhD; David Vismer[4], PhD; Rong Yang[5], BN

[1]Institute of Information Security and Big Data, Central South University, Changsha, China

[2]Intramural IT and Bioinformatics Program, Division of Intramural, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, United States

[3]Division of Computational Science, Center for Information Technology, National Institutes of Health, Bethesda, MD, United States

[4]Sapient Government Services, Arlington, VA, United States

[5]7th Ward, Xiangya Hospital, Central South University, Changsha, China

Corresponding Author:
Rong Yang, BN
7th Ward
Xiangya Hospital
Central South University
87 Xiangya Road
Changsha, 410008
China
Phone: 86 73189753007
Fax: 86 73189753007
Email: cxlyr0576@163.com

Abstract

Background: As one of the several effective solutions for personal privacy protection, a global unique identifier (GUID) is linked with hash codes that are generated from combinations of personally identifiable information (PII) by a one-way hash algorithm. On the GUID server, no PII is permitted to be stored, and only GUID and hash codes are allowed. The quality of PII entry is critical to the GUID system.

Objective: The goal of our study was to explore a method of checking questionable entry of PII in this context without using or sending any portion of PII while registering a subject.

Methods: According to the principle of GUID system, all possible combination patterns of PII fields were analyzed and used to generate hash codes, which were stored on the GUID server. Based on the matching rules of the GUID system, an error-checking algorithm was developed using set theory to check PII entry errors. We selected 200,000 simulated individuals with randomly-planted errors to evaluate the proposed algorithm. These errors were placed in the required PII fields or optional PII fields. The performance of the proposed algorithm was also tested in the registering system of study subjects.

Results: There are 127,700 error-planted subjects, of which 114,464 (89.64%) can still be identified as the previous one and remaining 13,236 (10.36%, 13,236/127,700) are discriminated as new subjects. As expected, 100% of nonidentified subjects had errors within the required PII fields. The possibility that a subject is identified is related to the count and the type of incorrect PII field. For all identified subjects, their errors can be found by the proposed algorithm. The scope of questionable PII fields is also associated with the count and the type of the incorrect PII field. The best situation is to precisely find the exact incorrect PII fields, and the worst situation is to shrink the questionable scope only to a set of 13 PII fields. In the application, the proposed algorithm can give a hint of questionable PII entry and perform as an effective tool.

Conclusions: The GUID system has high error tolerance and may correctly identify and associate a subject even with few PII field errors. Correct data entry, especially required PII fields, is critical to avoiding false splits. In the context of one-way hash transformation, the questionable input of PII may be identified by applying set theory operators based on the hash codes. The count and the type of incorrect PII fields play an important role in identifying a subject and locating questionable PII fields.

XSL•FO
RenderX

Introduction

Background

To accelerate biomedical discovery, it is critical for researchers to collaborate, especially to share their study data with each other. After announcing the Big Data Research and Development Initiative to explore how big data could be used to address important problems faced by the government in 2012, Obama's administration proposed Precision Medicine Initiative [1] in 2015. The latter will seek to collect data from large populations and integrate biomedical research with health care. In general, subject data is collected from multiple sites. There needs to be a link between the data from those different sites on the same subject. Personally identifiable information (PII) is often used to identify and aggregate different types of data (eg, laboratory, imaging, genetic, clinical assessment data) of the same subject collected from multiple sites [2]. Generally PII includes an ID (eg, patient ID, social security number, or national ID), name, birth date, birth place, address, postcode, and so on [3]; however, sharing PII may lead to disclosing privacy of an individual. Therefore, when medical data is shared, privacy protection is a very important task of biomedical research [4,5], especially when PII is a concern [6]. Patient data must be protected before they are transferred [7,8]. In the United States, sharing health information must comply with the Standards for Privacy of Individually Identifiable Health Information and the Common Rule [9,10].

There are various methods to protect a patient's privacy, including data anonymization [10,11], deidentification [12-14], depersonalization [15], limited dataset [16], and hash transformation [17,18]. Among the unique ID methods of protecting patient privacy, the global unique identifier (GUID) algorithm is an effective solution. It transforms combination patterns of PII fields into hash codes by a one-way hash algorithm. It can be used to identify a participant across sites or studies, without transferring any portion of PII. Multiple PII fields can be gathered and combined in different patterns, facilitating matching even in the face of variations across collection sites. As part of the GUID algorithm, the identifying information undergoes one-way hash before being transferred to the central system, so that PII is never transmitted or stored outside collection sites.

For the GUID system [18] to work properly, PII must be collected with a high degree of accurate entry. If there are many errors in the items captured, none of the hash codes may match and there will be a false split (ie, where the same subject is given 2 different GUIDs). Although several methods, including double data entry, were proposed to improve data entry accuracy, the most effective way is prompting questionable fields during data entry. Therefore, while registering a subject, the client application of the GUID system would ideally check the PII input to allow the user to correct them, if any errors are found. This task must depend on the information stored on the GUID server; however, only the GUID and its related hash codes are stored on the GUID server (ie, no portion of PII is stored on the server). In addition, a GUID is a random code that is not directly generated from PII or hash codes. Hash codes are related to PII, but they have been mapped by a one-way hash algorithm, and it is impossible to reidentify PII fields. Thus, it is problematic to find exact questionable inputs while registering a subject. Fortunately, in the GUID system, there are multiple hash codes, which are transformed from combinations of PII fields and where some of the PII fields are overlapping within different hash codes. Therefore, it is possible to identify and reduce data entry error based on matching hash codes and its corresponding PII fields. Our study will explore it based on set theory.

Before exploring the analysis of questionable data input while registering a subject in the GUID system, it is necessary to review the principle of the system.

The GUID System

PII Fields and Its Combination Patterns

The GUID system [18] uses 17 PII fields for identifying a subject, including 8 required fields and 9 optional fields (Table 1). Generally, they are unique for the subject and do not change in the lifetime of the subject. Each PII field has its associated approximated probability such that 2 different individuals can randomly be identified within the subject population of the system sharing the same value for that field.

Each PII field is programmatically normalized to have only uppercase letters and numbers, no spaces, and no punctuation. For each subject, these PII fields are combined with 5 patterns (Table 2) according to their combined inverse probability that ensures a high degree of subject separation. Each combination pattern is converted into a 64-byte hash code by a one-way hash algorithm. An additional byte is appended to each resulting code to indicate the count of missing PII fields for the hash code. Each combination is sufficient to discriminate confidently subjects. In turn, a random unique GUID code will be generated and associated with that subject. The GUID and its linked hash codes are stored on the GUID server and used for anonymously identifying the subject in a clinical study. Because PII fields are not sent to the GUID server, and therefore are not stored in the server, privacy protection is maintained.

Table 1. Personally identifiable information (PII) fields used in global unique identifier (GUID) system.

Type	Name	Meaning
Required	FN	Complete legal given (first) name at birth
	LN	Complete legal family (last) name at birth
	MN	Complete legal additional (middle) name
	SEX	Physical sex at birth (male or female)
	COB	Country of government issued or national ID
	DOB	Day of birth
	MOB	Month of birth
	YOB	Year of Birth
Optional	GIID	Government issued or national ID
	MFN	Mother's complete legal given (first) name at her birth
	MLN	Mother's complete legal family (last) name at her birth
	FFN	Father's complete legal given (first) name at his birth
	FLN	Father's complete legal family (last) name at his birth
	MDOB	Mother's day of birth
	MMOB	Mother's month of birth
	FDOB	Father's day of birth
	FMOB	Father's month of birth

Table 2. Personally identifiable information (PII) combination patterns for hash cod.

Hash code	Combinations patterns
1	$YOB + DOB + SEX + GIID^a$
2	$FN + MN + LN + COB + DOB + MOB$
3	$FN + YOB + MFN^a + MLN^a + FFN^a + FLN^a$
4	$FN + LN + COB + SEX + MDOB^a + MMOB^a + FDOB^a + FMOB^a$
5	$FN + MN + MOB + MFN^a + FFN^a + MLN^a$

[a]The field that is optional.

Match Rule of Hash Code and Subject in GUID System

As part of the GUID system, each hash code consists of 64-bytes hash value, which is computed from PII combination pattern using a one-way hash algorithm, and 1 additional byte is added to hold the count of missing PII fields in the hash code (Figure 1). So, any error with PII fields used in a combination will result in a failure to match a hash code.

The GUID system has 3 types of hash codes: perfect, good, and bad. For each hash code, 2 parameters are used to determine its type: a lower threshold (L) and an upper threshold (U) (Table 3). A perfect hash code requires that the count of missing PII fields is equal to or less than L. The count of missing PII fields for generating a good hash code is limited to the interval (L,U). If the count of missing PII fields is greater than U, its related hash code will be defined as a bad one. The match between 2 perfect hash codes is called a perfect match, and the match between 2 good hash codes is considered a good match.

Table 3. Thresholds of missing fields to determine type of hash code.

| Parameters	Hash code 1	Hash code 2	Hash code 3	Hash code 4	Hash code 5
Lower threshold	0	1	1	1	1
Upper threshold	1	2	3	3	3

Once PII is inputted while registering a subject, the system will calculate the count of perfect matches or good matches. In turn, it will determine if there exists a matched subject based on matched hash codes. There are 3 parameters to determine if a subject is matched: threshold for a perfect match (P), threshold for a good match (G), and threshold for a mixed match (X).

Two subjects match each other when the count of perfect matches ≥ P, or the count of good matches ≥ G, or the sum of the count of perfect matches and good matches ≥ X. In this system, the thresholds are set to P=1, G=2, and X=2. In the context of the above GUID system, correct PII is critical for uniquely identifying a subject. Therefore, before requesting a randomly assigned GUID from the server, checking the input value of the PII fields is essential; however, since hash code is the only information related to PII in the GUID system, a process for checking questionable PII input must depend on the hash codes.

Figure 1. Components of hash code.
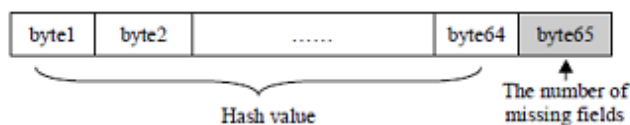
Methods

Study Design

Hash codes are generated from the combinations of PII fields in GUID system, so each one can be considered as a set of transformed PII fields. In addition, there are overlapping PII fields populated within different hash codes. Therefore, set theory may be used to systematically validate questionable PII fields. As long as a hash code is matched, its corresponding PII fields may be eliminated from questionable PII fields by set operations. Because missing values of optional PII fields are permitted, first all probable combination patterns of PII fields for perfect or good hash codes need to be analyzed and then the algorithm for checking questionable PII input might be designed.

Probable PII Combination Patterns for Perfect or Good Hash Codes

According to the principle of the GUID system, there are 3 types of hash codes and a subject is identified only with perfect or good hash codes. Missing fields may affect the match of a hash code. While registering a subject, if missing fields are considered, some improper mismatching will be avoided. For example, hash code 4 from Table 2 (Figure 2) is generated from the combination of required fields FN, LN, COB, and SEX and optional fields MDOB, MMOB, FDOB, and FMOB. Assuming that a subject was registered for the first time, the MDOB field was missed, and the other fields were correctly inputted, it would generate hash code 4^0. But when the subject is registered again on another site, and the correct value of all the above PII fields including MDOB is provided, the system will produce hash code 4'. Because field MDOB was missed in hash code 4^0, hash code 4' will not match with hash code 4^0. However, there is a perfect match between hash code 4' and hash code 4. If field MDOB is supposed as missing field to generate hash code 4'', hash code 4'' will be a perfect match with the previous hash code 4^0 and thus will avoid improper mismatching of hash code 4. So all perfect or good hash codes of a subject, which are registered, should be analyzed for identifying the subject and checking questionable PII fields.

Each hash code is generated from different combination patterns of PII fields, which are optional or required. Based on the combination patterns, the match rule of hash code and the type of PII fields, all probable perfect or good hash codes of the GUID system can be analyzed and identified (Figure 3 and Table 4). For example, hash code 3 is generated from a combination pattern of fields MFN, MLN, FFN, FLN, FN, and YOB. Of them, fields FN and YOB are required fields and the other 4 fields are optional. According to match rules of hash codes, a perfect hash code 3 may have 1 missing field and a good hash code 3 may have 2 or 3 missing fields. That is, a perfect hash code 3 may contain 1 missing field from MFN, MLN, FFN, or FLN and a good hash code 3 may use only 1 or 2 of those PII fields. So there are 5 probable perfect and 10 probable good hash code 3.

Table 4. Probable personally identifiable information (PII) combinations for hash codes with different matching types.

Index	Hash code	Combinations of personally identifiable information fields								Missed fields	Type of hash code
1	1	GIID	SEX	DOB	YOB						Perfect
2		a	SEX	DOB	YOB					GIID	Good
3	2	FN	LN	MN	DOB	MOB	COB				Perfect
4	3	MFN	MLN	FFN	FLN	FN	YOB				Perfect
5		MLN	a	FFN	FLN	FN	YOB			MFN	Perfect
6		MFN	FFN	a	FLN	FN	YOB			MLN	Perfect
7		MFN	MLN	FLN	a	FN	YOB			FFN	Perfect
8		MFN	MLN	FFN	FN	a	YOB			FLN	Perfect
9		a	a	FFN	FLN	FN	YOB			MFN, MLN	Good
10		a	MLN	a	FLN	FN	YOB			MFN, FFN	Good
11		a	MLN	FFN	a	FN	YOB			MFN, FLN	Good
12		MFN	a	a	FLN	FN	YOB			MLN, FFN	Good
13		MFN	a	FFN	a	FN	YOB			MLN, FLN	Good
14		MFN	MLN	a	a	FN	YOB			FFN, FLN	Good
15		a	a	a	FLN	FN	YOB			MFN, MLN, FFN	Good
16		MFN	a	a	a	FN	YOB			MLN, FFN, FLN	Good
17		a	MLN	a	a	FN	YOB			MFN, FFN, FLN	Good
18		a	a	FFN	a	FN	YOB			MFN, MLN, FLN	Good
20	4	MDOB	MMOB	FDOB	FMOB	FN	LN	SEX	COB		Perfect
19		a	MMOB	FDOB	FMOB	FN	LN	SEX	COB	MDOB	Perfect
21		MDOB	a	FDOB	FMOB	FN	LN	SEX	COB	MMOB	Perfect
22		MDOB	MMOB	a	FMOB	FN	LN	SEX	COB	FDOB	Perfect
23		MDOB	MMOB	FDOB	a	FN	LN	SEX	COB	Perfect / FMOD	
24		a	a	FDOB	FMOB	FN	LN	SEX	COB	MDOB, MMOB	Good
25		a	MMOB	a	FMOB	FN	LN	SEX	COB	MDOB, FDOB	Good
26		a	MMOB	FDOB	a	FN	LN	SEX	COB	MDOB, FMOB	Good
27		MDOB	a	a	FMOB	FN	LN	SEX	COB	MMOB, FDOB	Good
28		MDOB	a	FDOB	a	FN	LN	SEX	COB	MMOB, FMOB	Good
29		MDOB	MMOB	a	a	FN	LN	SEX	COB	FDOB, FMOB	Good
30		a	a	a	FMOB	FN	LN	SEX	COB	MDOB, MMOB, FDOB	Good
31		a	a	FDOB	a	FN	LN	SEX	COB	MDOB, MMOB, FMOB	Good
32		a	MMOB	a	a	FN	LN	SEX	COB	MDOB, FDOB, FMOB	Good
33		MDOB	a	a	a	FN	LN	SEX	COB	MMOB, FDOB, FMOB	Good

XSL•FO
RenderX

Index	Hash code	Combinations of personally identifiable information fields						Missed fields	Type of hash code
34	5	FN	MN	MFN	FFN	MLN	MOB		Perfect
35		FN	MN	a	FFN	MLN	MOB	MFN	Perfect
36		FN	MN	MFN	a	MLN	MOB	FFN	Perfect
37		FN	MN	MFN	FFN	a	MOB	MLN	Perfect
38		FN	MN	MFN	a	a	MOB	FFN, MLN	Good
39		FN	MN	a	a	MLN	MOB	MFN, FFN	Good
40		FN	MN	a	FFN	a	MOB	MFN, MLN	Good
41		FN	MN	a	a	a	MOB	MFN, FFN, MLN	Good

[a]The optional field that may be missed while being collected.

Figure 2. An example for match among hash codes.
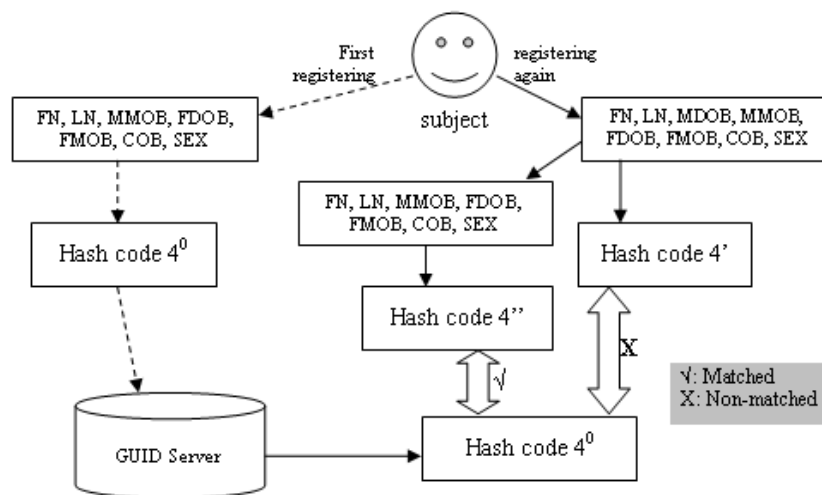
Figure 3. The count of probable perfect or good hash codes.

Hash code	Count of optional fields	Lower threshold	Upper threshold	Count of perfect hash code	Count of good hash code
1	1	0	1	$C_1^1=1$	$C_1^0=1$
2	0	1	2	$C_0^0=1$	-
3	4	1	3	$C_4^4 + C_4^3 =5$	$C_4^2 + C_4^1 =10$
4	4	1	3	$C_4^4 + C_4^3 =5$	$C_4^2 + C_4^1 =10$
5	3	1	3	$C_3^3 + C_3^2 =4$	$C_3^1 + C_3^0 =4$
Total				16	25

Set Theory and Checking Questionable Fields

Set theory is one of the most important theories of information processing. A set is a collection of a type of objects, and its basic operations include subtraction, union, intersection, subset, and so on. To eliminate some elements from a collection, the set operation (ie, subtraction) is a good solution. Since a hash code is transformed from a combination of PII fields, it must

XSL•FO
RenderX

be related to a set of PII fields. Once it matches with one of the hash codes of an identified subject, a corresponding set of PII fields also must match with each other and those PII fields will be considered validated. So using set theory, with the match rule of hash codes and subject in the GUID system, some PII input errors are likely to be located. For example, assuming that while registering a subject, it is found that the PII fields for hash codes 3, 4, and 5 are without missing fields and those hash codes match perfectly with the corresponding hash codes of the identified subject in the server. In addition, hash codes 1 and 2 do not match with the corresponding hash codes of the identified subject. According to the matching rules of the subject, it may be deduced that the subject has been registered in the system. The PII fields related to hash codes 3, 4, and 5 can be eliminated from questionable PII fields. That is,

{GIID, FN, LN, MN, DOB, MOB, YOB, SEX, COB, MFN, MLN, FFN, FLN, MDOB, MMOB, FDOB, FMOB}

/{FN,YOB, MFN, MLN, FFN, FLN} //PII related to hash code 3

U {FN, LN, MDOB, MMOB, FDOB, FMOB, COB, SEX} //PII related to hash code 4

U {FN, MN, MOB, MFN, FFN, MLN} //PII related to hash code 5

={GIID, DOB}

Therefore, the result suggests that questionable fields may be located at PII fields DOB and GIID (Figure 4). Figure 5 shows another example of locating questionable PII fields. In this case, only hash code 3 and 5 are good matches and PII fields FFN, FLN, and MLN are missed. There are 2 good hash code matches (G=2), so the subject has been registered according to the match rules. Due to the missing of fields FFN, FLN, and MLN, here the PII fields corresponding to hash code 3 are FN, YOB, and MFN and those to hash code 5 are FN, MN, MOB, and MFN. Therefore,

{GIID, FN, LN, MN, DOB, MOB, YOB, SEX, COB, MFN, MLN, FFN, FLN, MDOB, MMOB, FDOB, FMOB}

{ FN, YOB, MFN } U {FN, MN, MOB, MFN } //PII related to hash code 3,5

={ GIID, LN, SEX, COB, DOB, MLN, FFN, FLN, MDOB, MMOB, FDOB, FMOB}

It may be deduced that data entry error exists within PII fields GIID, LN, SEX, COB, DOB, MLN, FFN, FLN, MDOB, MMOB, FDOB, and FMOB.

Based on set theory and the principle of the GUID system, while registering subjects, the algorithm checking questionable PII fields can be described as following.

Step 1 Input PII of subject S_r being registered;

Step 2 Generate all probable perfect or good hash codes HC_{pg} of S_r, HC_{pg}={HC_1, HC_2, …, HC_{41}}, and store temporarily their corresponding set of PII field name, PII_1, PII_2, ..., PII_{41}, to HC_1, HC_2, ... and HC_{41} on the local site as described in Table 4:

PII_1={GIID, SEX, DOB, YOB}

PII_2={SEX, DOB, YOB}

…

PII_{41}={ FN,MN,MOB}

Step 3 Find matched subjects, S_m, with S_r from the GUID server according to match rules and HC_{pg};

Step 4 If count of $S_m > 1$ then

S_r is not unique;

else if S_m is empty then

S_r is a new subject;

else

Find hash codes in HC_{pg} that match with those of S_m

and get their set of PII fields, $PII_1{'}$, $PII_2{'}$, ...;

Step 5 Calculate union U_{PII} of $PII_1{'}$, $PII_2{'}$, ...;

$U_{PII} = PII_1{'}$ U $PII_2{'}$, U ...

Step 6 Calculate subtraction between U_{PII} and all PII fields;

R_{PII} = {GIID, FN, LN, MN, DOB, MOB, YOB, SEX, COB, MFN, MLN, FFN, FLN, MDOB, MMOB, FDOB, FMOB} - U_{PII}

Step 7 Return remaining PII fields R_{PII} which are questionable.

Figure 4. An example for locating questionable personally identifiable information (PII) fields while hash codes are perfect match.
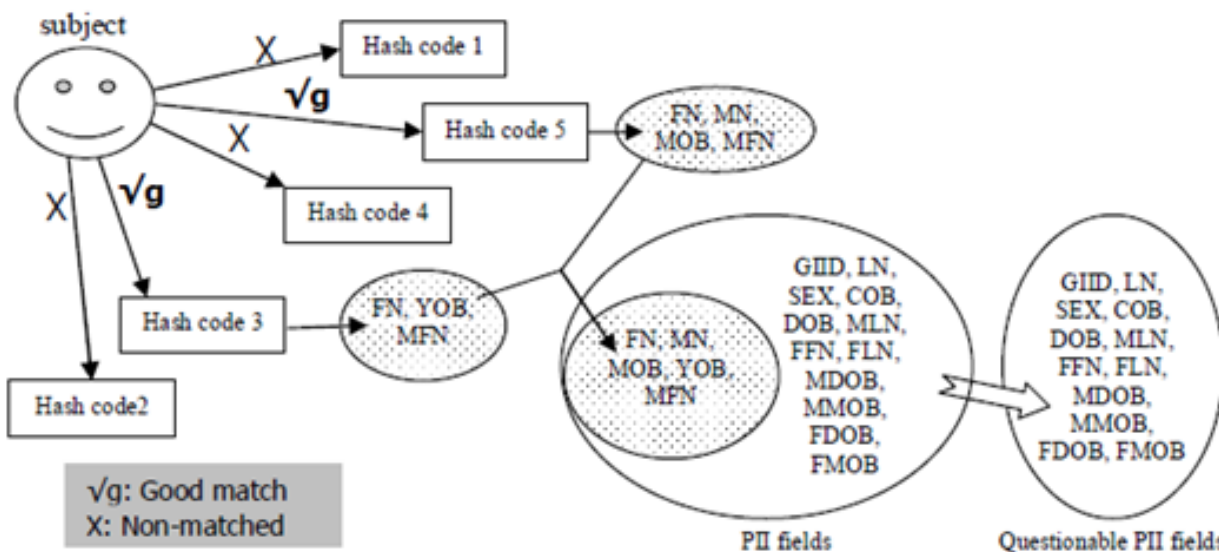
Figure 5. An example for locating questionable personally identifiable information (PII) fields while hash codes are good match.
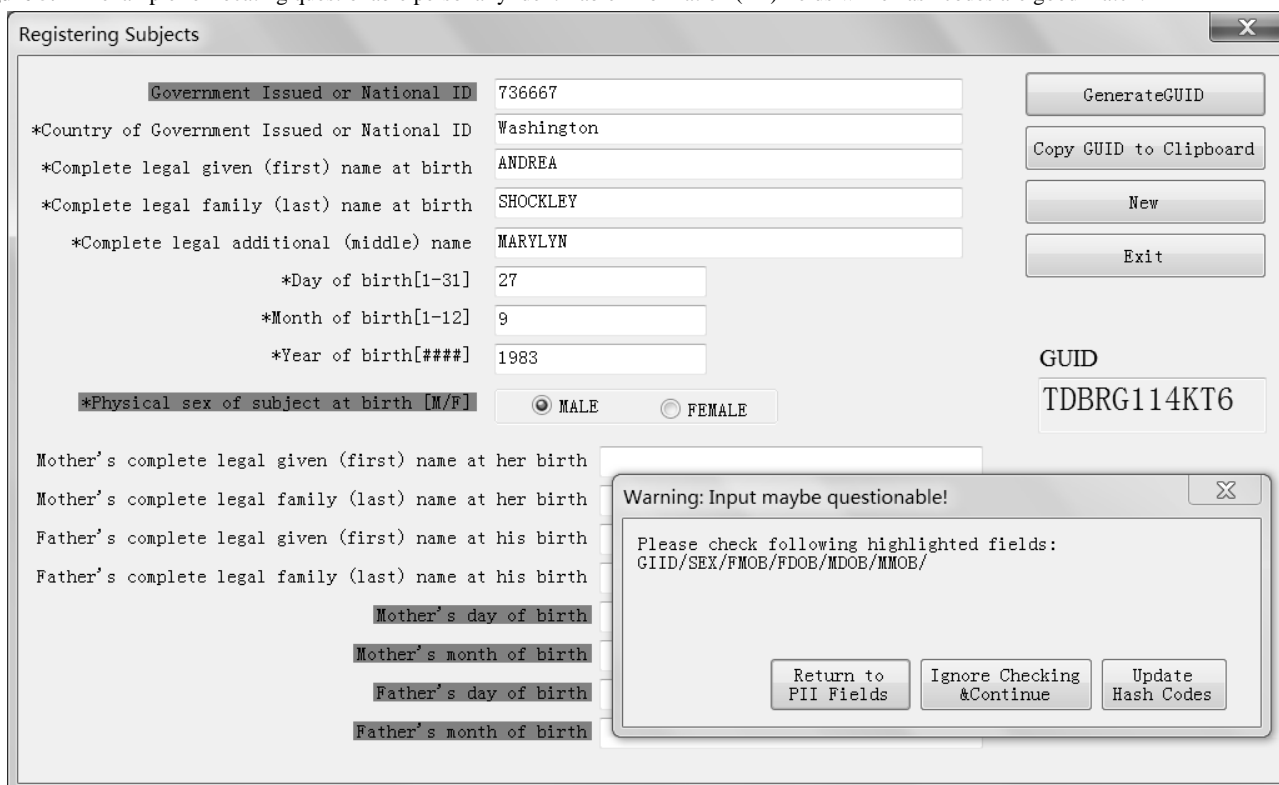
Simulations

For evaluating the proposed algorithm, the mailing list information [18] has been used as simulation data. Of mailing list information on 1 million individuals, first name (FN), last name (LN), and middle name (MN) were kept and the city of residence was used as city of birth (COB). Dates of birth (YOB, MOB, and DOB) were randomly generated. Individuals were assigned parents' information (MFN, MLN, FFN, FLN, MDOB, MMOB, FDOB, and FMOB) to be logically consistent with the family structure. The values of field GIID are replaced with the index of subjects. Randomly emptying is used to simulate

missing of optional fields. From the included pretreated subjects, we randomly selected 200,000 subjects for the simulation study of our method. Their original hash codes were generated and stored on the GUID server.

Then we randomly planted 200,000 errors into the simulation data, including emptying, inserting, deleting, and replacing. In any given field of the same hash code, the count of planted error is not more than one. After planting errors, out of 200,000 subjects, there are 127,700 subjects with errors and 72,300 subjects with no error. In 1 subject, the maximum for planted

errors is 8. The count (N_Err) and percent of planted errors by   PII fields is shown in Table 5.

Table 5. Distribution of planted errors by personally identifiable information (PII) fields.

PII[a] fields	N_Err	Percent (%)
Required fields		
FN	12,937	6.47
LN	14,166	7.08
MN	10,234	5.12
COB	12,954	6.48
DOB	10,440	5.22
MOB	12,645	6.32
YOB	11,578	5.79
SEX	11,587	5.79
Optional fields		
GIID	7980	3.99
MFN	12,984	6.49
MLN	10,504	5.25
FFN	10,823	5.41
FLN	11,656	5.83
MDOB	13,603	6.80
MMOB	11,301	5.65
FDOB	11,188	5.59
FMOB	13,420	6.71
Total	200,000	100

[a]PII: personally identifiable information.

After the dataset is treated, only error-planted subjects are used for simulating input while registering from the client application. The proposed algorithm is applied to validate and locate these planted errors.

Applications

When reregistering a subject in a GUID system, the proposed methods may be used to perform the following 2 tasks:

1. Checking questionable PII fields to ensure correct input. If any of the PII fields of the subject are improperly input, the client application will prompt the user to recheck the specified PII without revealing actual input value by using the proposed method.

2. Updating hash codes. If the client ensures that input of PII fields are correct and more complete than before, the application will allow the system to update hash codes.

For the above 2 tasks, we have developed an application program and integrated it into current GUID registering operation. Registered subjects are selected to confirm its value.

Results

Matching of Subjects

Due to planted errors, the values of some PII fields have changed. As shown in Table 6, of 127,700 error-planted subjects, 89.63% (114,464/127,700) are still identified by the hash codes from their remaining correct PII fields. The other 10.37% (13,236/127,700) subjects cannot match with their previous entries and are identified as new subjects. 83.16% (65,383/78,619) of the subjects with errors in required fields are still identified. All unidentified subjects have the required PII fields with errors. Additionally, of all identified subjects, 57.13% (65,383/114,464) have the required fields with errors.

Table 6. Identifying of error-planted subjects.

Matching type	Rec_{erf}[a]	Rec_{nerf}[b]	Subtotal
Unidentified	13,236	0	13,236
Identified	65,383	49,081	114,464
Total	78,619	49,081	127,700

[a]Rec_{erf}: the count of subjects with errors in required fields. [b]Rec_{nerf}: the count of subjects with no error in required fields.

Simulation results show that the average errors planted into the identified subjects is 1.48 and that planted into the unidentified subjects is 2.29. Table 7 lists the count of errors planted into 1 subject (n_{Err}), the count of subjects with n_{Err} errors (n_{Rec_Err}), the count of identified subjects with n_{Err} error, and the ratio of $n_{Rec\_Err\_Mtch}$ to n_{Rec_Err} ($n_{Rec\_Err\_Mtch}$). Table 8 displays the count of incorrect required fields in 1 subject (n_{Err_ReqF}), the count of subjects with n_{Err_ReqF} incorrect required fields ($n_{Rec\_Err\_ReqF}$), the count of identified subjects with n_{Err_ReqF} incorrect required fields ($n_{Rec\_Err\_ReqF\_Mtch}$), and the ratio of $n_{Rec\_Err\_ReqF\_Mtch}$ to $n_{Rec\_Err\_ReqF}$.

Table 7. Identifying of subjects with different count of planted errors.

n_{Err}	n_{Rec_Err}	$n_{Rec\_Err\_Mtch}$	Ratio
1	74,883	71,796	95.88
2	37,327	32,104	86.01
3	12,143	8798	72.45
4	2792	1545	55.34
5	476	199	41.81
6	69	18	26.09
7	8	4	50.00
8	2	0	0.00

Table 8. Identifying of subjects with different count of error required fields.

n_{Err_ReqF}	$n_{Rec\_Err\_ReqF}$	$n_{Rec\_Err\_ReqF\_Mtch}$	Ratio
0	49,081	49,081	100.00
1	62,716	56,750	90.49
2	14,026	8038	57.31
3	1740	569	32.70
4	132	25	18.94
5	5	1	20.00

Recalling of Planted Errors

Simulation results show that PII errors may be found and located within the limited fields. The best situation is to precisely locate an error at 1 PII field. The worst situation is to reduce the questionable scope of errors down to a set of 13 PII fields. According to the simulated results, the mean questionable scope of errors is shrunk to a set of 5.64 PII fields, 3.59 times as many as the average of errors planted into a subject. It suggests that the mean questionable scope of errors can be limited to a set of less than 4 PII fields.

For identified subjects, the count of analyzed questionable PII fields (n_{cqf}) is related to the count of planted errors in a subject (Table 9). For example, for subjects with only 1 error, the average of questionable PII is shrunk to 4.27 fields. For those with 7 errors, it is limited to 13 fields.

Table 10 lists the count of analyzed questionable fields by PII fields (n_{cqf_PII}). The subjects with error field FN has the maximum mean analyzed questionable PII (13 fields) and the subjects with error field GIID has the minimum mean analyzed questionable PII (3.74 fields). The subjects with other error PII fields have no significant difference.

If only 1 error is planted into a subject, the count of analyzed questionable PII fields (n_{cqf_1}) depends on the type of error PII field (Table 11). For example, it is 1 for the error field GIID, 13 for the error field FN, and 1 or 4 for the error field MDOB.

Table 9. The count of analyzed questionable fields by count of errors.

Count of planted errors in a subject	n_{cqf}		
	Minimum	Maximum	Average
1	1	13	4.27
2	2	13	7.39
3	3	13	9.42
4	4	13	10.86
5	6	13	11.67
6	11	13	11.83
7	13	13	13.00

Table 10. The count of analyzed questionable fields by personally identifiable information (PII) fields.

PII[a] fields with planted errors	n_{cqf_PII}		
	Minimum	Maximum	Mean
Required fields			
FN	13	13	13
LN	6	13	7.65
MN	2	13	5.56
SEX	6	12	7.30
COB	6	13	7.67
DOB	2	11	5.69
MOB	2	13	5.53
YOB	3	11	5.28
Not required fields			
GIID	1	11	3.74
MFN	1	13	6.48
MLN	1	13	6.51
FFN	1	13	6.59
FLN	1	13	4.84
MDOB	1	13	6.12
MMOB	1	13	6.11
FDOB	1	13	6.09
FMOB	1	13	6.06

[a]PII: personally identifiable information.

Table 11. The count of analyzed questionable personally identifiable information (PII) fields from subjects with only one error.

PII[a] fields with planted errors	n_{cqf_1}
Required fields	
FN	13
LN	6
MN	2
SEX	6
COB	6
DOB	2
MOB	2
YOB	3
Optional fields	
GIID	1
MFN	1/4
MLN	1/4
FFN	1/4
FLN	1
MDOB	1/4
MMOB	1/4
FDOB	1/4
FMOB	1/4

[a]PII: personally identifiable information.

Applications

The proposed hash code analysis scheme is integrated into the GUID application to enhance GUID accuracy. While registering a subject, who has been previously registered in the system, it analyzes the questionable PII fields, highlights them, and requests the client to correct them (Figure 6).

When the application finds the questionable PII fields, it will give a hint regarding possible PII errors. If it is confirmed that the input of all PII fields are proper, the user may select "update hash codes" function and the application will update the hash codes in the server based on user's input.

Figure 6. The application of checking questionable personally identifiable information (PII) fields.

Discussion

Identifying of Subject

In the GUID system [18], there are 17 PII fields, including 8 required fields and 9 optional fields. PII fields are combined into 5 patterns, which are processed into hash codes by a one-way hash algorithm. For privacy protection, only hash codes and its related random GUID code are stored on the server. In this case, it is impossible to directly identify a subject by PII and hash codes are the key to identifying a subject. One perfect hash code or 2 good hash codes is sufficient to identify a subject and the system has better error tolerance. A subject with error PII fields may still be identified and it is confirmed by the simulation result of this study. As shown in Table 6, 89.63% of subjects with error PII fields do still match with their previous entries.

In addition, simulation results also show that the count and type of error PII fields in a subject have great effect on identifying the subject. In Table 7, it can be found that the probability of identifying the subject is reversely related to the count of planted errors. That is, the more errors that are planted into a subject, the lower is probability of identifying the subject. Table 6 shows that all unidentified subjects have the errors within its required PII fields. It can also be deduced that the subject without error within required PII fields must be correctly identified. That is, if all required PII fields of a subject are correctly entered, the subject must be identified well. Table 8 indicates that when more errors are planted into required PII fields of a subject, the probability of identifying the subject is lower. Therefore, it suggests that required PII fields are vital to identifying a specific subject. According to the principles of the GUID system, we can also find that the match criteria and find important PII fields

based on the composition of hash codes. For example, PII field FN is a required field for hash code 2, 3, 4, and 5. Once this PII field of a subject is incorrect, those 4 hash codes will not be matched. In turn, it will significantly reduce the probability of identifying the subject. So to ensure correct registration of a subject, especially with required PII fields, correct data entry is critical to avoiding false splits.

Reducing PII Entry Errors

Hash codes are generated from PII, but it is an irreversible process and a hash code cannot be transformed back into PII. Therefore, it is impossible to validate questionable input by reversing hash codes to PII, which is intended by design. Additionally, missing values of PII fields make it more difficult to validate questionable PII fields. Fortunately, there exists a map between combinations of PII fields and hash codes and there are overlapping PII fields among hash codes of a subject. Each hash code represents a set of PII fields and all probable perfect or good hash codes (Figure 3 and Table 4) may be analyzed and produced for a subject being registered. Therefore, set theory can be used for analyzing questionable PII fields. For example, while registering a subject, if its hash code 1 is perfectly matched, then its PII fields GIID, SEX, DOB, and YOB can be eliminated from questionable PII fields. Simulation results confirm that the questionable PII fields of all identified subjects may be found and located. The best situation is to locate an error at one exact PII field; the worst situation is to reduce the scope of possible errors in a subject down to a set of 13 PII fields. The mean scope of possible errors in a subject is shrunk to a set of 5.64 PII fields, 3.59 times as many as the average of errors planted into a subject.

The simulation results also show that the count of analyzed questionable PII fields is closely related to the count of actual

errors. The greater the count of actual errors, the more the questionable PII fields to be evaluated (Table 9). For subjects with only 1 error, the scope of questionable inputs can be limited to an average set of 4.27 PII fields. For subjects with 7 errors, it could be a set of 13 PII fields. The type of PII fields with error is also associated with the count of analyzed questionable PII fields. For subjects with only 1 error, if the error is for an optional PII field, it can be located at 1 or upto 4 PII fields. If the error is for a required field, it cannot be limited to such narrow scope (Table 11). For example, the error in the FN field will result in the failed matching of hash codes 2, 3, 4, and 5 no matter whether there are other errors. Thus, at most, only hash code 1 is a perfect match and fields GIID, SEX, DOB, and YOB can be eliminated from questionable fields. The remaining 13 PII fields will be evaluated as questionable fields (Tables 10 and 11). Fortunately, the accuracy of first name is very high [18].

By using the proposed method in this study, while registering a subject, the application may give a proper hint to the user about questionable PII input. If the user assures that input of PII fields are correct, the hash codes in the system may be updated to improve from the previous entry error, thus improving the robustness of the GUID system.

Conclusions

In summary, a subject with PII errors may still be identified in the GUID system but it depends on the number and type of PII errors. Using set operations, questionable PII fields from the client application may be analyzed based on hash codes but it is difficult to find the exact location of an error because hash codes come from combinations of PII fields and it cannot be reversed to PII. If questionable PII fields need be precisely located, all probable perfect or good hash codes must be stored on the server or the generating mechanism of hash codes in the system must be redesigned.

Conflicts of Interest

None declared.

References

1. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795. [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
2. Al-Shahi R, Warlow C. Using patient-identifiable data for observational research and audit. Br Med J 2000 Oct 28;321(7268):1031-1032 [FREE Full text] [Medline: 11053151]
3. HHS. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule URL: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [accessed 2016-12-22] [WebCite Cache ID 6mwQOTn19]
4. Veerle V, Louise C, Matthew W. Managing and sharing data: best practice for researchers. 3rd edition. Colchester, Essex: UK Data Archive; 2011:1-904059.
5. De Clercq E, Van Casteren V, Bossuyt N, Moreels S, Goderis G, Bartholomeeusen S, et al. Nation-wide primary healthcare research network: a privacy protection assessment. Stud Health Technol Inform 2012;174:23-28. [Medline: 22491104]
6. Frost J, Vermeulen IE, Beekers N. Anonymity versus privacy: selective information sharing in online cancer communities. J Med Internet Res 2014 May 14;16(5):e126 [FREE Full text] [doi: 10.2196/jmir.2684] [Medline: 24828114]
7. Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J Investig Med 2010 Jan;58(1):11-18 [FREE Full text] [doi: 10.2310/JIM.0b013e3181c9b2ea] [Medline: 20051768]
8. Haynes CL, Cook GA, Jones MA. Legal and ethical considerations in processing patient-identifiable data without patient consent: lessons learnt from developing a disease register. J Med Ethics 2007 May;33(5):302-307 [FREE Full text] [doi: 10.1136/jme.2006.016907] [Medline: 17470509]
9. Kushida CA, Nichols DA, Jadrnicek R, Miller R, Walsh JK, Griffin K. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. Med Care 2012 Jul;50 Suppl:S82-101. [doi: 10.1097/MLR.0b013e3182585355] [Medline: 22692265]
10. Privacyruleandresearch.nih. Clinical research and the HIPAA Privacy Rule URL: https://privacyruleandresearch.nih.gov/pdf/clin_research.pdf [accessed 2016-12-22] [WebCite Cache ID 6mwQm1x6P]
11. Ohno-Machado L, Bafna V, Boxwala AA, Chapman BE, Chapman WW, Chaudhuri K, iDASH team. iDASH: integrating data for analysis, anonymization, and sharing. J Am Med Inform Assoc 2012;19(2):196-201 [FREE Full text] [doi: 10.1136/amiajnl-2011-000538] [Medline: 22081224]

XSL·FO

RenderX

12. Freymann JB, Kirby JS, Perry JH, Clunie DA, Jaffe CC. Image data sharing for biomedical research--meeting HIPAA requirements for de-identification. J Digit Imaging 2012 Feb;25(1):14-24 [FREE Full text] [doi: 10.1007/s10278-011-9422-x] [Medline: 22038512]

13. Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. J Am Med Inform Assoc 2007;14(5):550-563 [FREE Full text] [doi: 10.1197/jamia.M2444] [Medline: 17600094]

14. El Emam K, Arbuckle L, Koru G, Eze B, Gaudette L, Neri E, et al. De-identification methods for open health data: the case of the Heritage Health Prize claims dataset. J Med Internet Res 2012 Feb 27;14(1):e33 [FREE Full text] [doi: 10.2196/jmir.2001] [Medline: 22370452]

15. Elger BS, Iavindrasana J, Lo Iacono L, Müller H, Roduit N, Summers P, et al. Strategies for health data exchange for secondary, cross-institutional clinical research. Comput Methods Programs Biomed 2010 Sep;99(3):230-251. [doi: 10.1016/j.cmpb.2009.12.001] [Medline: 20089327]

16. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. J Am Med Inform Assoc 2010;17(2):169-177 [FREE Full text] [doi: 10.1136/jamia.2009.000026] [Medline: 20190059]

17. Quantin C, Fassa M, Coatrieux G, Riandey B, Trouessin G, Allaert FA. [Linking anonymous databases for national and international multicenter epidemiological studies: a cryptographic algorithm]. Rev Epidemiol Sante Publique 2009 Feb;57(1):33-39. [doi: 10.1016/j.respe.2008.10.010] [Medline: 19162419]

18. Johnson SB, Whitney G, McAuliffe M, Wang H, McCreedy E, Rozenblit L, et al. Using global unique identifiers to link autism collections. J Am Med Inform Assoc 2010;17(6):689-695 [FREE Full text] [doi: 10.1136/jamia.2009.002063] [Medline: 20962132]

Abbreviations

CNOB: Country of birth
COB: Name of city or municipality in which subject was born
DOB: Day of birth
FDOB: Father's day of birth
FFN: Father's complete legal given (first) name at his birth
FLN: Father's complete legal family (last) name at his birth
FMOB: Father's month of birth
FN: Complete legal given (first) name at birth
GIID: Government Issued or national ID
GUID: Global Unique Identifier
LN: Complete legal family (last) name at birth
MDOB: Mother's day of birth
MFN: Mother's complete legal given (first) name at her birth
MLN: Mother's complete legal family (last) name at her birth
MMOB: Mother's month of birth
MN: Complete legal additional (middle) name
MOB: Month of birth
PII: Personally identifiable information
SEX: Physical sex at birth (male or female)
YOB: Year of Birth

Progress in the Enhanced Use of Electronic Medical Records: Data From the Ontario Experience

Mavis Jones[1], PhD; Chad Koziel[1], MMA; Darren Larsen[1], MD; Plumaletta Berry[2], MHI; Elena Kubatka-Willms[3], MPP

[1]OntarioMD, Toronto, ON, Canada
[2]Formerly with OntarioMD, Toronto, ON, Canada
[3]Formerly with the Ontario Medical Association, Toronto, ON, Canada

Corresponding Author:
Mavis Jones, PhD
OntarioMD
150 Bloor St W
Suite 900
Toronto, ON, M5S 3C1
Canada
Phone: 1 416 623 1248 ext 5388
Fax: 1 416 623 1249
Email: mavis.jones@ontariomd.com

Abstract

Background: This paper describes a change management strategy, including a self-assessment survey tool and electronic medical record (EMR) maturity model (EMM), developed to support the adoption and implementation of EMRs among community-based physicians in the province of Ontario, Canada.

Objective: The aim of our study was to present an analysis of progress in EMR use in the province of Ontario based on data from surveys completed by over 4000 EMR users.

Methods: The EMM and the EMR progress report (EPR) survey tool clarify levels of capability and expected benefits of improved use. Maturity is assessed on a 6-point scale (0-5) for 25 functions, across 7 functional areas, ranging from basic to more advanced. A total of 4214 clinicians completed EPR surveys between April 2013 and March 2016. Univariate and multivariate descriptive statistics were calculated to describe the survey results.

Results: Physicians reported continual improvement over years of use, perceiving that the longer they used their EMR, the better patient care they provided. Those with at least two years of experience reported the greatest progress.

Conclusions: From our analyses at this stage we identified: (1) a direct correlation between years of EMR use and EMR maturity as measured in our model, (2) a similar positive correlation between years of EMR use and the perception that these systems improve clinical care in at least four patient-centered areas, and (3) evidence of ongoing improvement even in advanced years of use. Future analyses will be supplemented by qualitative and quantitative data collected from field staff engagements as part of the new EMR practice enhancement program (EPEP).

Introduction

Electronic medical records (EMRs) have significant potential to support quality patient care for community-based practices through the following multiple functions: appointment scheduling, practice billing, communication and messaging, encounter documentation, data quality and nomenclature consistency, document management, results management, referral and consultation tracking, prevention and screening, complex care or chronic disease management.

EMR use across Canada has increased steadily over the past few years with growing awareness of practice-level benefits. In 2014, EMR use increased by 53%, with 77% of primary care physicians using EMRs—up from 24% in 2007 [1,2]. In Ontario,

over 13,000 physicians, representing an adoption rate of 71% of family doctors and 55% of community-based specialists, have EMRs in their practices. The challenge now is to move beyond the basic use of EMRs to a more advanced use in practices. EMRs can facilitate the collection of population health data for analytics, planning, and delivery, and evidence is also accumulating on financial benefits. In the United States, the case for promoting EMR "meaningful use" via the 2009 HITECH Act [3,4] was supported by early analyses suggesting reduced billing errors at the practice level [5] and an anticipated US $81 billion in annual savings that could eventually double through technology-enabled improvements to prevention, management of chronic disease, and subsequent social benefits [6]. Similarly, a Canada Health Infoway study estimated that annually, nationwide EMR use resulted in workflow savings valued at Can $177 million, reduction in duplicate testing and adverse drug events valued at Can $123 million, and emerging benefits from chronic disease management and preventive care (for example increases in vaccination rates associated with EMR-generated reminders) [7].

Despite anticipated benefits, barriers to achieving optimal EMR use prevent us from fully understanding the impact EMR use might have on upstream processes and downstream outcomes of care [8]. A systematic review by Boonstra and Broekhuis identifies 8 interrelated categories of barriers, including financial, technical, time, psychological, social, and legal. Two additional barriers—organizational and change—should be paid special attention, the authors argue, as they mediate the effects of the others and thereby most directly influence a project's success [9]. There is indeed a growing literature on EMR implementation as a complex change project [10-13] targeting factors related to cost, time, technical issues, or resistance (fear

of change, doubt the investment of resources will be worth it, and so on) [11,12,14-17] . EMR adoption and use operates in a complex adaptive system, highly sensitive to shifting politics and public policy. Larger, well-resourced physician practices and hospitals may be equipped to manage change associated with EMR adoption; others, however, may struggle without help. It is critical that barriers in all types of practices be systematically addressed [7,18] . Here, we describe OntarioMD's approach to supporting community-based physicians in the adoption and optimization of EMRs and analysis from self-assessment surveys of over 4000 EMR users.

Methods

Change Management Approach and Maturity Model

Recognizing that mature use does not necessarily result from the installation of hardware and software in clinics, OntarioMD (a subsidiary of the Ontario Medical Association, with funding from the Ontario Ministry of Health and Long Term Care) applies a practice level change management strategy to address variability in physician EMR use. Modeled on the awareness, desire, knowledge, ability, and reinforcement (ADKAR) framework (Prosci) [19], this approach identifies factors that facilitate or obstruct a user's capability to adapt to change, generate new knowledge, acquire skills, improve performance, and sustain momentum (Table 1). Multidisciplinary support is available from EMR vendors, peer leaders (experienced EMR users—physicians, clinic managers, and nurses), and the EMR practice enhancement program (EPEP). EPEP provides intensive assistance in the form practice advisors who conduct on-site assessments of current use, identify hidden gaps, and develop action plans to help physicians optimize their EMR use.

Table 1. OntarioMD's change management approach.

ADKAR[a] elements	OntarioMD supports
Awareness of the need for change	Partnerships and support from the Ontario Medical Association and Ontario government drive awareness of the need and support status as a trusted advisor
Desire to support and participate in the change	Peer leaders validate the practice benefits of EMR[b] use; self-assessment identifies priority areas
Knowledge of how to change	Practice advisors, peer leaders, and vendors instruct and advise on EMR use in the implementation phase
Ability to implement required skills and behaviours	Practice advisors, peer leaders and vendors support changes in practice workflow and specific functionalities
Reinforcement to sustain the change	Support from practice advisors, peer leaders and vendors continues; EPEP[c] provides hands-on support to plan concrete actions towards enhanced use

[a]ADKAR: awareness, desire, knowledge, ability, and reinforcement.

[b]EMR: electronic medical record.

[c]EPEP: EMR practice enhancement program.

To anchor this work, OntarioMD developed the EMR maturity model (EMM) and the EMR progress report (EPR) survey tool. Influenced by existing robust models [20-22], the EMM provides a basis for understanding differences in levels of EMR use and the benefits that can be expected with mature use. In the EMM:

- Each key measure is identified as a practice aspect where EMR solutions can have a significant impact, relevant to performance assessment at practice and population levels.
- Each measure can be assessed independently across the 6 maturity levels from 0 to 5, where 0 is paper-based.
- Each level of maturity builds upon the functionality or maturity state of the preceding level.

XSL•FO
RenderX

- EMR capability starts at level 1 based on the specification offerings and requirements of OntarioMD's EMR adoption program assuming adoption of that aspect of the EMR software.
- Levels 0 to 3 are within scope for an average community-based practice. Levels 4 and 5 mostly reflect potential capabilities (eg, population health) and connectivity, not available in all contexts.

Use of these tools provides insight into the reasons for variability in maturity across practices. Figure 1 shows the EMM, as updated in 2016 to coincide with the retirement of the EPR survey in favor of a new tool, with which we are now collecting data for future progress reports (more information on these updates can be found at ontariomd.ca or by contacting the authors). As originally conceived, the EPR allowed for assessment of against the 6 maturity levels (0–5), for 25 EMR functions across 7 functional areas, ranging from basic to more advanced (Table 2).

Table 2. Functional areas and corresponding functions mapped on the original electronic medical record (EMR) maturity model (EMM).

Functional area	Functions
Practice management	Appointment scheduling
	Practice billing
	Communication and coordination
	Business continuity planning
Information management	Registration information
	Encounter documentation
	Data quality management
	Nomenclature consistency
	Document management
	Privacy and security
Patient results management	Laboratory results
	Diagnostic image reports
	Hospital summary information
	Referrals and consults tracking
Diagnosis support	Patient assessment tools
	Preventive or follow-up care
	Evidence-based resources
Treatment planning support	Care planning and coordination
	Medication management
	Complex care or chronic disease management
Patient engagement and communication	Patient education
	Self-care or comanagement
Evaluation and monitoring	Health quality indicators
	Health outcome measures
	Public health reporting

Figure 1. OntarioMD's electronic medical record maturity model.

Level	Criteria	Capabilities
5	INTEGRATE	*Use of portals, hubs, attachment to provincial e-health platforms sharing data from the EMR.*
4	POPULATION DATA USE	*Dashboarding of whole populations, acting upon the whole, performing population analysis at the practice level.*
3	LOOK AHEAD/ PREDICT	*Reminders and alerts are used at the point of care. Searches are done regularly and scheduled for review.*
2	EARLY DATA USE	*Acting upon the output of episodic searches, quick entry tools, forms, calculators, etc.*
1	ENTER DATA	*Documentation occurs electronically. Progress notes, forms, and other documents are entered into the EMR.*
0	PAPER	*Processes are primarily paper-based.*

Survey Instrument

Data presented here were gathered via the EPR, the survey instrument in effect until summer 2016. The EPR was a Web-based self-assessment tool designed to help community-based family physicians and specialists enrolled in OntarioMD's EMR adoption program identify their current skill level and track progress over time. The survey tool was developed based on the review of the evidence on best practices and evaluation of EMR adoptions within the primary care environment (eg, Health Care Information and Management Systems Society). It was face-validated by the 30 (at that time) members of OntarioMD's peer leader program, who then pilot-tested the survey among their clinical associates. The tool was subsequently refined and was launched province wide in August 2013. Figure 2 shows a sample screenshot of a question in the EPR.

Since that time, with physicians across the province moving to mature use, both the EMM and survey were updated to reflect the evolving realities of practice across 3 broad functional areas associated with quality patient-centered care: practice management, information management, and diagnosis and treatment support. The EMR progress assessment (EPA), launched in summer 2016, is a more concise instrument for assessing maturity; but, as they essentially measure the same thing, data from the EPR and EPA have been blended to support longitudinal analysis. With the former EPR and now the EPA, physicians have immediate access to their own data on several measures and can compare their performance with the average of physicians surveyed across the province or those in the same practice type (eg, solo or group). In the clinical environment, the EPR tool facilitates benchmarking, gap analysis, customized goal setting, and improvement projects.

Figure 2. Sample electronic progress report question.

Data and Analysis

Of the 11,650 participants enrolled in the adoption program, 4214 completed at least one EPR. Analyses were run on the most recent EPR completed by each participant (some completed more than 1 EPR between April 2013 and March 2016, for annual assessments). Univariate and multivariate descriptive statistics were calculated to summarize responses, whereas multivariate linear regressions were developed to describe correlations and adjust for the competing effects of variables. Respondents were analyzed using the parameters "type of physician" (family physician or specialist) and "number of years of EMR use" (<2, 2-4, 4-6, >6). Of the 4214 who completed an EPR, almost half were between ages 45 and 64 years (2078); 1776 were 44 years or younger, and 360 were 65 years or older. Surveys were completed by clinicians province wide, across its range of population demographics and densities (urban, suburban, rural, or remote). Information was not collected on physicians' gender.

Limitations

The data analyzed here were subject to the following limitations:

- Financial incentives: EMR adoption funding agreements granted physicians a payment upon completing an EPR.
- Self-report: All data is self-reported, thus representing clinicians' perceptions of improvement rather than measurable improvements against health quality indicators (not standardized in Ontario at the time of launch).
- Technological limitations: Levels 4 and 5 (integrated care and population health impacts) may not be attainable on some measures due to interoperability issues, lack of availability in connected provincial assets, and EMR product-specific limitations.
- Pace of progress: Health information technologies are rapidly evolving and may quickly render the findings presented here outdated. Regular updates could address this problem.
- Variation across areas of specialization: Enrollment in OntarioMD's funding program was originally only open to family physicians; a limited number of specialists were included in 2009 and both cohorts have grown steadily since then. Rate of enrollment varies significantly across specialties; however, for reasons that are beyond the scope of our analysis, family physicians still comprise the majority of survey respondents.
- Generalizability concerns: The functional areas defined in the tool are broadly applicable—certainly in the Canadian context—as a means to establish within-practice baselines against which progress can be measured. However, caution should be exercised in using these tools to compare or benchmark maturity across regions in a larger geographic area. Adjustment should be made for factors that could influence maturity, such as infrastructural and demographic variations.

Results

Overall

Figure 3 shows the breakdown of respondents by type of practitioner (family practice and specialist) and years of EMR use (time since go-live date). The majority of responses were completed within 2 years of adopting a new EMR.

Figure 3. Breakdown of respondents by years of use (N=4214).
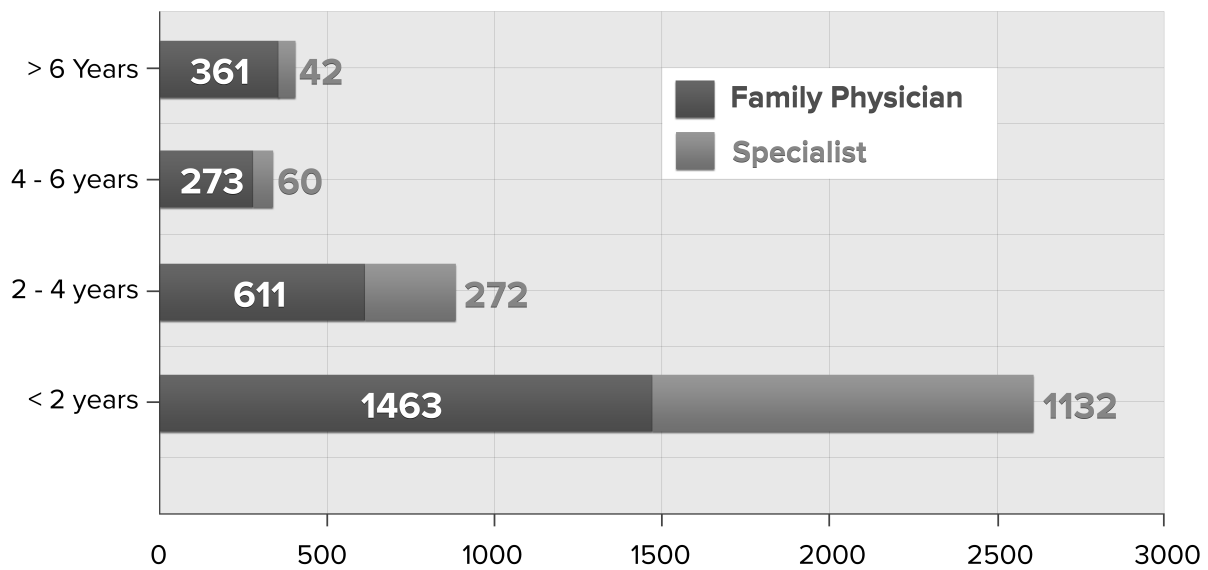
Self-Reported EMR Maturity by Years of Use

Ideally mature EMR use improves over time. Physicians generally need a minimum of 2 to 3 years of EMR data collection to ensure that their records are adequately populated to support advanced functions, for example, monitoring their patient population to identify those due for preventive screening. 73.94% (2426/3281) of physicians indicated that they were primarily paperless and 25.51% (837/3281) reported using both paper and electronic charts. Comparing data from this sample (all of whom have at least adopted an EMR) with the general population polled in the National Physician Survey [1,2] and bearing in mind semantic nuances in the questions asked, we still saw a steady shift from primarily paper to primary electronic charts between 2007 and 2014 (Table 3). Further, the percentage of Ontario physicians reporting using electronic charts in 2014 was among the highest in the country (tied with BC and Saskatchewan, and second only to Alberta).

Table 3. Extent of electronic versus paper-based workflow in physician practices, 2007 and 2014. (source: National Physician Survey; responding samples have been weighted to represent the population size. See www.nationalphysiciansurvey.ca)

Workflow description	Ontario only, 2007	Ontario only, 2014	National, 2007	National, 2014
Total n/N	2286/20,267	3883/26,238	7038/55,398	9711/68,177
Paper charts only	54.8%	16.6%	57.9%	21.3%
Combination of paper and electronic charts	29.8%	48.4%	26.1%	49.3%
Electronic charts instead of paper charts	9.9%	34.9%	9.8%	29.4%

Figure 4 shows physicians' self-reported maturity level across increments of years of use. (As noted, each increment represents a different cohort, rather than the same cohort progressing across increments.) Results indicate movement through maturity levels over time (as one might expect, even given the limitations of this analysis). Of those responses completed at 4 or fewer years of use, 45.55% (1582/3473) reported an overall maturity of less than level 2, with an additional 45.98% (1597/3473) achieving an overall maturity of level 2, and only 8.47% (294/3473) progressing beyond level 2. In contrast, of those responses completed at more than 6 years of use, 57.0% (228/400) had reported an overall maturity level of 2, and 21.0% (84/400) had progressed beyond level 2. No respondents reported achieving level 5 on these functions. Overall results reflected that physicians are now integrating the EMR as an essential tool and using the core functionalities to engage patients in their

day-to-day practice operation. For Figures 4-6, due to missing responses, the N is 4206 rather than 4214.

This snapshot of overall EMR maturity, although positive, masks the range of proficiency within each skill level as well as the impact of maturity on individual practices. Unraveling this aggregated information about multiple clinical functionalities could reveal a deeper understanding of issues that may limit mature EMR use. To this end, we selected 4 measures associated with patient care—continuity of care, quality of care, patient safety, and patient experience—to explore physicians' perceptions of how EMRs affect their capability in these areas. These measures reflect Ontario practice and health system priorities of patient-centered care, as outlined in Health Quality Ontario's primary care performance measurement framework [23] (informed by quality-driven frameworks such as Institute

for Health Care Improvement's Triple Aim) [24] and the Ontario government's Patients First Action Plan [25], emphasizing sustainability through access to care, care coordination and integration, patient safety, and improved outcomes.

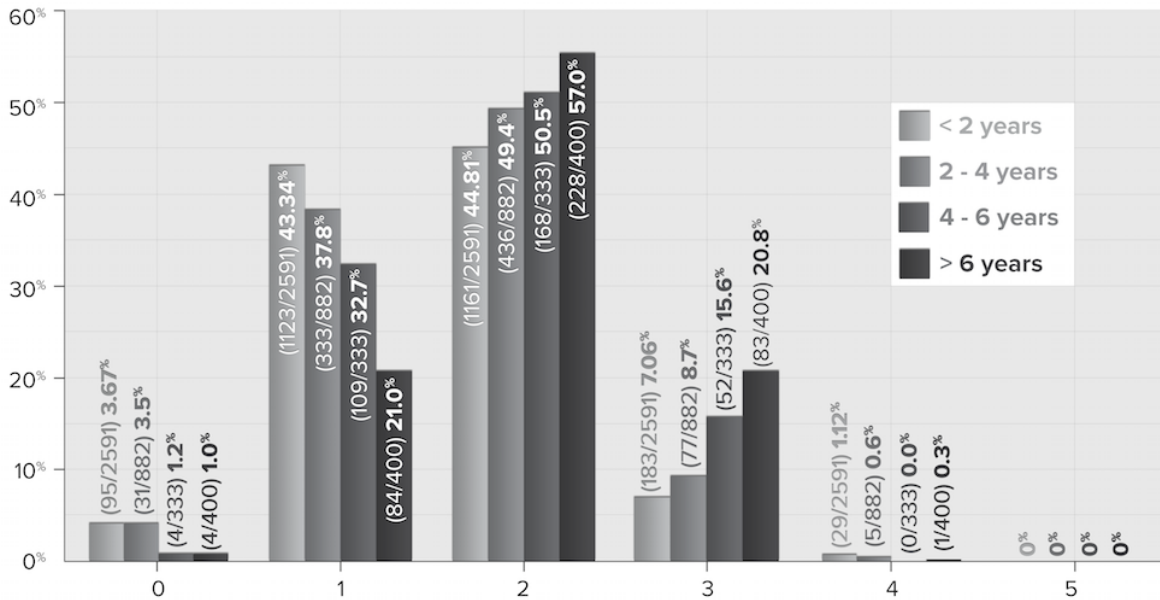**Figure 4.** Maturity by years of use (N=4206).

Figure 5. Percentage of physicians who indicate improvement on all patient care metrics by years of use (N=4206).
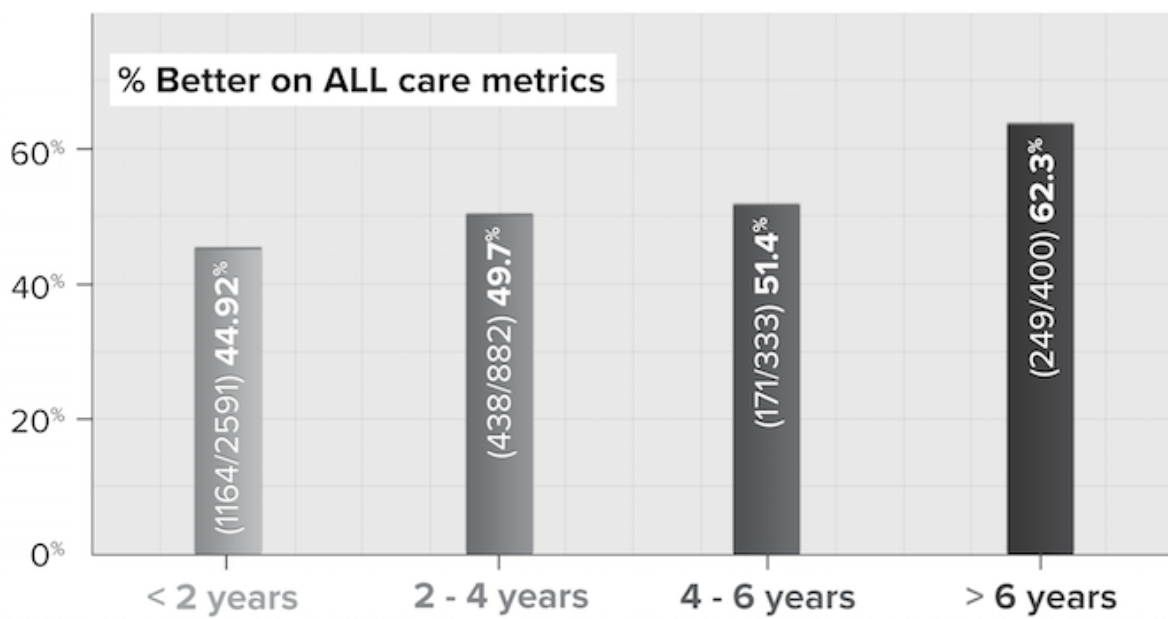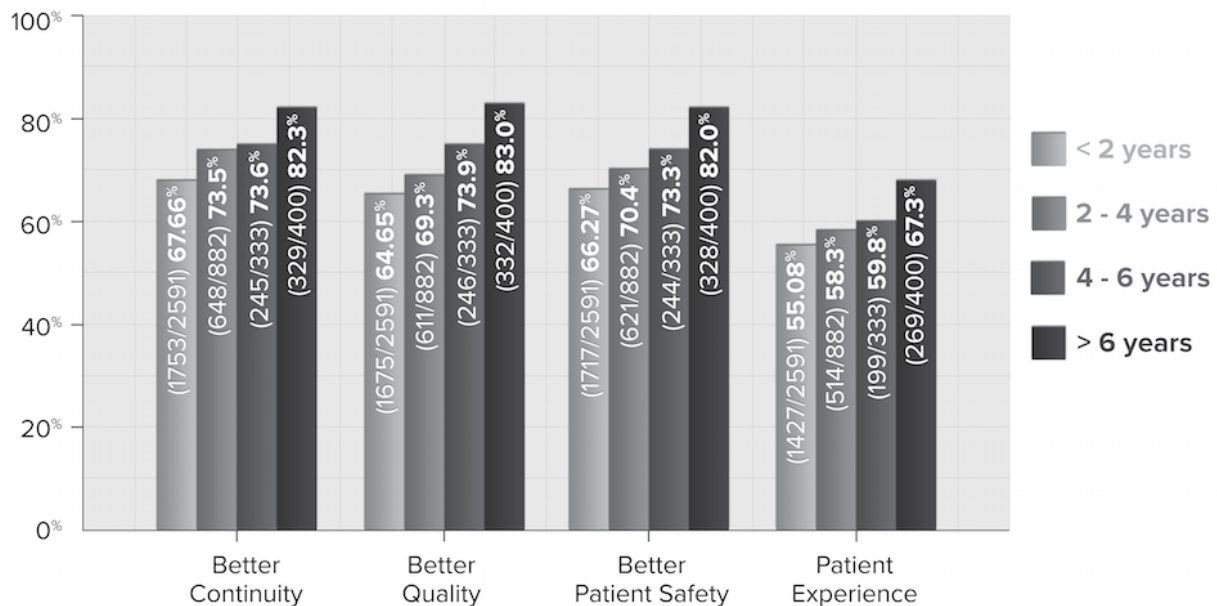
Figure 6. Physician perception across all patient care dimensions by years of use (N=4206).

Focus: Perception of Impact on Patient Care

The EPR asked physicians to rate their view of changes in their care approach since implementing an EMR if newly adopted, or over the last year if more experienced, on a 5-point scale: much worse, somewhat worse, about the same, somewhat better, or significantly better. The following 4 measures of patient care were addressed: quality of care, patient experience, continuity of care, and patient safety

Figure 5 shows the percentage of physicians whose response to that question was "somewhat better" or "significantly better" for all 4 measures. Physicians responding to the survey reported perceiving that the longer they used their EMR, the better patient care they provided. Those with at least two years of experience reported the greatest progress.

Furthermore, when the analysis was run for "same or better," the percentages climbed by 10 for every increment. While it cannot be concluded from this data that longer EMR use translates into greater expertise—nor improved patient care—the correlation is suggestive and worth further exploration, especially as it is consistent with other findings [26]. As shown in Figure 6, patient care is perceived as continually improving, as physicians accumulate more EMR use over time.

Discussion

While we are still accruing the longitudinal data to tell a more fulsome story about progress on enhanced use, our initial assessments suggest the following: (1) There is a direct correlation between years of EMR use and EMR maturity as measured in our model; (2) There is a similar positive correlation between years of EMR use and the perception that these systems improve clinical care in at least four patient-centered areas; and

(3) There is evidence of ongoing improvement even in advanced years of use (ie, we have not yet plateaued on the benefit of change management efforts and practice improvement support).

Future analyses will have the benefit of insights from our peer leaders and EPEP field staff, in particular regarding the value of these change management strategies in supporting our enrolled clinicians in advancing to mature EMR use. However, as we gain these insights, we will also gain a risk of selection bias. Physicians who have invested time, money, and energy into the implementation of their EMR tool may feel frustration due to the disruption in their practice workflow inherent in large scale change, and slow progress in the first year or two of their use. After they have adapted to their new workflows, they may become more neutral or increasingly satisfied, tending to perceive—and report—their situation more positively (ie, "if I've stuck with my EMR this long I must be satisfied" versus "I'm very satisfied with a top notch product and its impact on my day").

Nevertheless, we can conclude that there is utility in examining the interaction between an innovation, its intended adopters, and the particular context (here, community-based practice)—particularly in assisting the innovation's spread and impact [27]. The change management approach used here recognizes that there are different types of adopters [22] whose needs and concerns vary. A strategy that studies differences between users and their workflow contexts, monitors their successes and obstacles, and assesses the value of supports (such as training focused on process and outcome rather than narrowly prescribed goals), can easily be adapted to other health technology challenges and contexts [28]. These factors were taken into account in the design of EPEP, where field staff supports physicians through site engagements, to address

barriers, and optimize the value their EMR brings to their practice.

Whereas the findings presented here are based on EPR self-assessments completed from 2013-2016, we can expect to mine a richer collection of data going forward. We have already begun preliminary analyses of data collected with the new EPA tool and are supplementing this with quantitative and qualitative data from our teams in the field. From this we can expect to develop a more nuanced understanding of practice-level barriers to and facilitators of progress in enhanced use, which in turn will inform how our teams provide support to clinicians in order to sustain their progress—particularly as the digital health landscape evolves to realize better connectivity and access to a patient's record at all points of care. The work required to maintain momentum, it is hoped, will be rewarded by observable (by clinicians) improvements in the quality of care delivered and in patient outcomes. It is clear from our experience that structured and measurable processes are critical to provide practices with effective ongoing support and training during and after EMR adoption. Our updated tools and approach will help us identify greater opportunities to help EMR users develop more sophisticated EMR capability, sustain and improve their proficiency, and build a more comprehensive view of the full potential of their EMR to benefit both their practice and the larger community of care.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Nationalphysiciansurvey.ca. 2007. 2007 National Physician Survey URL: http://nationalphysiciansurvey.ca/surveys/2007-survey/ [accessed 2017-01-20] [WebCite Cache ID 6nex4NfyG]
2. Nationalphysiciansurvey.ca. 2014. 2014 National Physician Survey URL: http://nationalphysiciansurvey.ca/surveys/2014-survey/ [accessed 2017-01-20] [WebCite Cache ID 6nex8PaG5]
3. Shin P, Sharac J. Readiness for meaningful use of health information technology and patient centered medical home recognition survey results. Medicare Medicaid Res Rev 2013;3(4):E1-E9 [FREE Full text] [doi: 10.5600/mmrr.003.04.b01] [Medline: 24834365]
4. Dranove D, Garthwaite C, Li B, Ody C. Investment subsidies and the adoption of electronic medical records in hospitals. J Health Econ 2015 Dec;44:309-319. [doi: 10.1016/j.jhealeco.2015.10.001] [Medline: 26596789]
5. Wang SJ, Middleton B, Prosser LA, Bardon CG, Spurr CD, Carchidi PJ, et al. A cost-benefit analysis of electronic medical records in primary care. Am J Med 2003 Apr 1;114(5):397-403. [Medline: 12714130]
6. Hillestad R, Bigelow J, Bower A, Girosi F, Meili R, Scoville R, et al. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. Health Aff (Millwood) 2005;24(5):1103-1117 [FREE Full text] [doi: 10.1377/hlthaff.24.5.1103] [Medline: 16162551]
7. PwC. Infoway-inforoute. Ontario: Canada Health Infoway; 2013. The emerging benefits of electronic medical record use in community-based care URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/benefits-evaluation/1224-the-emerging-benefits-of-electronic-medical-record-use-in-community-based-care-full-report [accessed 2017-01-20] [WebCite Cache ID 6nexMkvR9]
8. Garber S, Gates S, Keeler E, Vaiana M, Mulcahy A, Lau C. RAND. Santa Monica, California: RAND Corporation; 2014. Redirecting innovation in U.S. health care URL: http://www.rand.org/pubs/research_reports/RR308.html [accessed 2017-01-20] [WebCite Cache ID 6nexaDI2S]
9. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010;10:231 [FREE Full text] [doi: 10.1186/1472-6963-10-231] [Medline: 20691097]
10. Ghazisaeedi M, Mohammadzadeh N, Safdari R. Electronic health record (EHR) as a vehicle for successful health care best practice. Med Arch 2014 Dec;68(6):419-421 [FREE Full text] [doi: 10.5455/medarh.2014.68.419-421] [Medline: 25648601]
11. Green LA, Potworowski G, Day A, May-Gentile R, Vibbert D, Maki B, et al. Sustaining "meaningful use" of health information technology in low-resource practices. Ann Fam Med 2015;13(1):17-22 [FREE Full text] [doi: 10.1370/afm.1740] [Medline: 25583887]
12. Madore A, Rosenberg J, Muyindike WR, Bangsberg DR, Bwana MB, Martin JN, et al. Implementation of electronic medical records requires more than new software: Lessons on integrating and managing health technologies from Mbarara, Uganda. Healthc (Amst) 2015 Dec;3(4):264-269. [doi: 10.1016/j.hjdsi.2015.08.006] [Medline: 26699355]
13. Shea CM, Malone R, Weinberger M, Reiter KL, Thornhill J, Lord J, et al. Assessing organizational capacity for achieving meaningful use of electronic health records. Health Care Manage Rev 2014;39(2):124-133 [FREE Full text] [doi: 10.1097/HMR.0b013e3182860937] [Medline: 23380882]

XSL•FO

RenderX

14. Peeters JM, Krijgsman JW, Brabers AE, Jong JD, Friele RD. Use and uptake of eHealth in general practice: a cross-sectional survey and focus group study among health care users and general practitioners. JMIR Med Inform 2016 Apr 06;4(2):e11 [FREE Full text] [doi: 10.2196/medinform.4515] [Medline: 27052805]

15. Bowens FM, Frye PA, Jones WA. Health information technology: integration of clinical workflow into meaningful use of electronic health records. Perspect Health Inf Manag 2010 Oct 01;7:1d [FREE Full text] [Medline: 21063545]

16. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]

17. Paré G, Raymond L, de Guinea AO, 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]

18. Friedberg M, Chen P, Van Busum KR, Aunon F, Pham C, Caloyeras J. RAND. Santa Monica, California: RAND Corporation; 2013. Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy URL: http://www.rand.org/pubs/research_reports/RR439.html [accessed 2017-01-20] [WebCite Cache ID 6ney7GPJ6]

19. Hiatt J. ADKAR: a model for change in business, government and our community. Loveland, Colorado: Prosci Research; 2006.

20. HIMSS. Amazonaws. 2008. Healthcare integration and connectivity: Results of a survey by the Enterprise Information Systems Steering Committee URL: http://s3.amazonaws.com/rdcms-himss/files/production/public/HIMSSorg/Content/files/IntegrationConnectivitySurvey.pdf [WebCite Cache ID 6neypc3rh]

21. Lau F, Price M, Keshavjee K. From benefits evaluation to clinical adoption: making sense of health information system success in Canada. Healthc Q 2011;14(1):39-45. [Medline: 21301238]

22. Canada Health Infoway. Pan-Canadian Change Management Framework. 2013. A framework and toolkit for managing ehealth change URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/toolkits/change-management/methodologies-and-approaches/1659-a-framework-and-toolkit-for-managing-ehealth-change-2?Itemid=101 [accessed 2017-01-20] [WebCite Cache ID 6nez3YxGs]

23. Health Quality Ontario. Hqontario. 2014. A primary care performance measurement framework for Ontario: report of the Steering Committee for the Ontario Primary Care Performance Measurement Initiative: Phase One URL: http://www.hqontario.ca/portals/0/Documents/pr/pc-performance-measurement-report-en.pdf [accessed 2017-01-20] [WebCite Cache ID 6neza4Pdf]

24. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769 [FREE Full text] [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]

25. Health.gov. Toronto, Ontario: Government of Ontario; 2015. Patients First: Action plan for health care URL: http://health.gov.on.ca/en/ms/ecfa/healthy_change/docs/rep_patientsfirst.pdf [accessed 2017-01-20] [WebCite Cache ID 6nezoUAQe]

26. Tharmalingam S, Hagens S, Zelmer J. The value of connected health information: perceptions of electronic health record users in Canada. BMC Med Inform Decis Mak 2016 Jul 16;16:93 [FREE Full text] [doi: 10.1186/s12911-016-0330-3] [Medline: 27422571]

27. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]

28. Canada Health Infoway. National Change Management Framework Internet URL: https://www.infoway-inforoute.ca/en/solutions/change-management/change-management-framework [accessed 2017-01-20] [WebCite Cache ID 6nezsu5wu]

Abbreviations

ADKAR: awareness, desire, knowledge, ability, and reinforcement
EMR: electronic medical record
EMM: electronic medical record (EMR) maturity model
EPA: EMR progress assessment
EPR: EMR progress report
EPEP: EMR practice enhancement program

Original Paper

The State of Open Source Electronic Health Record Projects: A Software Anthropology Study

Mona Alsaffar[1*], MSc; Peter Yellowlees[2*], MBBS, MD; Alberto Odor[3,4*], MSc, MD; Michael Hogarth[5*], MD

[1]University of California-Davis, Department of Health Informatics, Sacramento, CA, United States

[2]University of California-Davis, Department of Psychiatry, Sacramento, CA, United States

[3]Health Informatics Graduate Group, University of California-Davis, Sacramento, CA, United States

[4]Betty Irene School of Nursing, University of California-Davis, Sacramento, CA, United States

[5]University of California-Davis, Department of Pathology, Sacramento, CA, United States

[*]all authors contributed equally

Corresponding Author:
Mona Alsaffar, MSc
University of California-Davis
Department of Health Informatics
3700 Business Dr, Suite 130
Sacramento, CA, 95820
United States
Phone: 1 9167348710
Fax: 1 (916) 734 3486
Email: mona_alsaffar@hotmail.com

Abstract

Background: Electronic health records (EHR) are a key tool in managing and storing patients' information. Currently, there are over 50 open source EHR systems available. Functionality and usability are important factors for determining the success of any system. These factors are often a direct reflection of the domain knowledge and developers' motivations. However, few published studies have focused on the characteristics of free and open source software (F/OSS) EHR systems and none to date have discussed the motivation, knowledge background, and demographic characteristics of the developers involved in open source EHR projects.

Objective: This study analyzed the characteristics of prevailing F/OSS EHR systems and aimed to provide an understanding of the motivation, knowledge background, and characteristics of the developers.

Methods: This study identified F/OSS EHR projects on SourceForge and other websites from May to July 2014. Projects were classified and characterized by license type, downloads, programming languages, spoken languages, project age, development status, supporting materials, top downloads by country, and whether they were "certified" EHRs. Health care F/OSS developers were also surveyed using an online survey.

Results: At the time of the assessment, we uncovered 54 open source EHR projects, but only four of them had been successfully certified under the Office of the National Coordinator for Health Information Technology (ONC Health IT) Certification Program. In the majority of cases, the open source EHR software was downloaded by users in the United States (64.07%, 148,666/232,034), underscoring that there is a significant interest in EHR open source applications in the United States. A survey of EHR open source developers was conducted and a total of 103 developers responded to the online questionnaire. The majority of EHR F/OSS developers (65.3%, 66/101) are participating in F/OSS projects as part of a paid activity and only 25.7% (26/101) of EHR F/OSS developers are, or have been, health care providers in their careers. In addition, 45% (45/99) of developers do not work in the health care field.

Conclusion: The research presented in this study highlights some challenges that may be hindering the future of health care F/OSS. A minority of developers have been health care professionals, and only 55% (54/99) work in the health care field. This undoubtedly limits the ability of functional design of F/OSS EHR systems from being a competitive advantage over prevailing commercial EHR systems. Open source software seems to be a significant interest to many; however, given that only four F/OSS EHR systems are ONC-certified, this interest is unlikely to yield significant adoption of these systems in the United States. Although the Health Information Technology for Economic and Clinical Health (HITECH) act was responsible for a substantial

XSL•FO

RenderX

infusion of capital into the EHR marketplace, the lack of a corporate entity in most F/OSS EHR projects translates to a marginal capacity to market the respective F/OSS system and to navigate certification. This likely has further disadvantaged F/OSS EHR adoption in the United States.

Introduction

Background

The medical field has been using open source applications for almost 40 years [1]. Electronic health record (EHR) systems first appeared in the early 1960s [2]. The Computer Stored Ambulatory Record (COSTAR) system was the first F/OSS EHR system and was originally developed to be used by the Harvard Community Health Plan. Although COSTAR was implemented in a number of institutions, it did not result in broad national adoption of EHRs at the time. Only the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 and its financial incentive program have resulted in broad adoption of EHRs in the United States [3]. F/OSS EHR systems have been increasing in popularity over the period [4].

Although the HITECH incentive payments have increased adoption, EHR adoption continues to have obstacles [5,6]. One of the main obstacles continues to be affordability [5]. CDW Healthcare Physician Practice estimated the total cost of an EHR deployment at approximately USD $120,000 per physician in the first year after implementation, with annual recurring costs of USD $30,000 per physician [7]. Along with the financial cost, there is also the non-financial cost related to time spent to bring the system live and into full functional use [7].

Open source EHR may lessen financial barriers while also providing improved flexibility given that they can be "freely" modified [8]. Many of the prevailing EHRs do not adhere to minimal usability testing standards [9] requiring continuous customization to meet the needs of the organization [10]. A KLAS study of 128 physicians on the current state of acute care EHRs found that no vendor scored high in usability [11]. Since open source software can be freely modified and redistributed, this could reduce the cost of continuous customization to improve usability [12]. Open source projects tend to also benefit from a higher degree of transparency about software anomalies (software bugs), leading to a higher degree of reliability over time. A common belief across the open source community and often referred to as "Linus Law" states "given enough eyeballs, all bugs are shallow" [13]. Unlike organizations who are dependent on a commercial vendor's prioritization of features and software release schedules, those implementing F/OSS would have complete control over the timing of customizations and deployment, allowing them to choose what functionality is available and when it will be available to their users [14].

F/OSS does come with challenges as well. Although some commercial companies provide support for F/OSS EHRs, the majority of the F/OSS EHR projects do not have a support service one can purchase. This creates a major challenge in ensuring reliability, particularly when the original system has been customized by institutional programmers [15]. Those skeptical of F/OSS EHR systems often highlight the potential dependency on volunteer developers [16] who do not guarantee technical support [15]. In addition, identifying a reliable source for version updates can be challenging [15]. Many organizations also fear that open source projects can become inactive anytime, creating an acute need for substantial in-house software development expertise [17]. A majority of health care organizations do not typically have infrastructure to support software development, and they might not have information technology (IT) staff with expertise in managing the software development lifecycle (SDLC) for complex systems. Instead, the typical health care delivery organization's IT staff focuses on deploying and optimizing vendor software.

Despite these disadvantages, the F/OSS software has been growing in terms of the number of projects. The 8th Annual Future of Open Source Survey found that the number of F/OSS projects doubled between 2012 [18] and 2014 [19].

The core success of the open source movement depends on developers who contribute their knowledge and effort for free to the community. Developers are either unpaid volunteers, hobbyists [20], or employees who are paid to write code. A study of mainstream F/OSS projects categorized developers' contribution into eight different roles: project leader, core member, active developers, peripheral developer, bug fixer, bug reporter, reader, and passive user [4]. As reflected in this categorization, there are a number of different roles for contributors and a significant amount of resources required to support a high-quality project. In large part, the developer community and their motivations are a key determinant of success or failure of an open source project. Exploring these motivations is an important aspect of understanding a key success factor for F/OSS EHR systems. The motivation-affecting factors for open source developers can be categorized as internal (cognitive) and external (social). Internal factors are comprised of motivation, altruism, and community identification. External motivation factors include future rewards (eg, peer recognition), self-marketing, human capital, contribution as part of employment (ie, being paid to contribute), and revenue from related products and services [21]. Generally, anything related to the joy of coding is considered intrinsic motivation, whereas extrinsic motivation is associated with receiving some benefit for the contribution. These factors have been explored in mainstream F/OSS projects but have not yet been characterized in F/OSS health care projects.

Objectives

The objectives of this study were to canvass the current state of open source EHR systems and to characterize the motivations, knowledge, and demographics of the developers.

Methods

To find EHR F/OSS projects, we used SourceForge, a widely used open source project repository, and Google using the search terms "electronic patient record," "electronic health record," "electronic medical records," and "clinical information system." The search revealed hundreds of EHR F/OSS projects, but only 54 of them were EHRs according to our study inclusion criteria. The following are two fundamental inclusion criteria used in the study: (1) the software had to be defined as an EHR, such that the project had to adhere to the functional definition of EHR. The HealthIT.gov website defines an EHR as a "digital chart" containing at least the medical and treatment history of the patients; and (2) the software had to use an open source license. Open source software is defined as software without license restrictions on its redistribution and the software can be freely modified. [22].

The study was conducted for a 3-month period starting May 2014. To understand the characteristics of the various EHR projects, we looked at license type, downloads, programming languages, spoken languages, project age, development status, supporting materials, top downloads by country and whether they were "certified" EHRs.

License Type

Many SourceForge applications are defined by their license type on the application homepage. In this study, the licenses were classified into permissive, restrictive, or highly restrictive [22]. The highly restrictive licenses, such as a general public license (GPL), allow free modification but request that any modification should be contributed back to the community under the same license. Highly restrictive licenses are used more in applications geared toward the end user (eg, games) [23]. License restrictions tend to affect who contributes and accessibility of the source code [23].

Downloads

We made an assumption that download frequency reflects the popularity of the software. This assumption is a commonly held belief in this research domain [24]. We looked at the download number in the last 12-month period on SourceForge. Around 47 projects (87%, 47/54) in this study have information on downloads, a proxy for use of the software.

Development Status

Development status shows the readiness of software for day to-day use. This study utilized the SourceForge classification for software readiness. In this study, six software stages were used: planning, pre-alpha, alpha, beta, production, stable, and mature. The software's status can influence a project's success and affects the interest of the users and developers [24].

Project Age

Project age represents the number of years since the project development started. The project's age, in addition to other factors, is positively related to its ability to attract more users and/or resources, which affect the project's future sustainability [25].

Programming Language

Programming language for each software system was examined and classified according to whether one or multiple were used. One open source study suggests that using one common programming language affects the success of the software project [26].

Spoken Languages

The projects were classified according to their spoken languages. One study proves that open source software popularity is related to the number of language versions available [27].

Supporting Materials

The setup of the system is not always obvious, and in some cases requires IT administration skills. The top 10 downloaded projects were analyzed, as these materials make the installation and usage of the system easier. The supporting materials included user guides, installation guides, and version demonstrations.

Top Downloads According to Countries

SourceForge provides important information about the software and gives the highest number of downloads for each EHR software system. The top 10 most downloaded projects were identified, in addition to the country that made the greatest number of downloads, to see if the adoption rate affects the number of downloads.

Certified Open Source Electronic Health Record

Using a certified EHR is a requirement for payments through the HITECH Act's EHR-incentive program. Certification of an EHR under the ONC program is an important success factor in the United States. The certified open source EHR products and their specifications from The Centers for Medicare and Medicaid Services (CMS) website were examined (Table 1).

Table 1. The Centers for Medicare and Medicaid Services-certified free and open source software electronic health record applications.

Product name	Original practice type	Vender	Product version number	Product classification	Certification year	Certification body
OpenEMR	Ambulatory	OEMR	4.1	Complete EHR[a]	2011	ICSA labs
TolvenEMR	Ambulatory, inpatient	Tolven Inc.	2.1	Complete EHR, modular	2011	ICSA labs
WorldVista	Ambulatory, inpatient	WorldVistA	2.0	N/A	2011	InfoGrad
ClearHealth	Ambulatory	ClearHealth Inc.	3.1.5	Complete EHR	2011	InfoGrad

[a]EHR: electronic health record.

Survey Data Collection

A survey was conducted in 2014 using the commercial survey tool SurveyGizmo (Boulder, CO). The survey consisted of 20 questions and took approximately 5 minutes to complete (Multimedia Appendix 1). The target audience included anyone who self-identified as a developer of health care F/OSS. Our questionnaire was modeled after a similar survey developed to compare proprietary and open source software in 2003 [28].

An announcement of the survey was published on 10 websites that focused on health care open source news and targeted health care developers. The announcement contained a brief summary of the main goal of the survey along with the author's names and their affiliations. About 1 week after this first announcement, the survey was distributed to 54 open source project developers' mailing lists obtained from SourceForge. We reached out to project email addresses and asked them to distribute our survey to their mailing lists. In addition, we sent a survey personal invitation to specific developers who mentioned working in F/OSS health care projects in their LinkedIn profile. The survey was posted for 5 weeks and a total of 103 responses were collected.

Results

Application Data

The study revealed several key observations. At the time of the study, there were 54 open source EHR projects, but only four had been successfully certified under the Office of the National Coordinator for Health Information Technology (ONC Health IT) Certification Program. Nearly half of the projects (57%, 31/54) used a restrictive license type, and approximately 57% (30/54) used GPL. The data revealed that 52% (28/54) of the projects were in production/stable status, only 2 (4%, 2/54) were in mature status, while 1 (2%, 1/54) project was inactive (Table 2). There were 44 active projects at varying stages of development, while 10 had unspecified status.

Table 2. Applications development status (N=54).

Development status	Frequency, n (%)
Production/stable	28 (52%)
Undetermined	11 (20%)
Alpha	6 (11%)
Beta	6 (11%)
Mature	2 (4%)
Pre-alpha	1 (2%)

As one might expect, many open source projects (46%, 25/54) used one programming language. However, a large percentage (36%, 19/54) used multiple programming languages. Approximately 18% (10/54) of the projects did not indicate the use of a specific programming language. The analysis also showed that the number of downloads for the projects that were written using multiple programming languages (n=147,914) were higher than the projects using one programming language (n=115,299). Among those projects indicating a programming language, the "PHP Hypertext Preprocessor", commonly known as PHP, was the leader, and it was used in 31% (17/54) of the projects.

The OpenEMR project had the highest number of downloads (63,418 in a 12-month period) (Table 3). The data shows that the United States accounts for the majority of the downloads and constitutes 64.07% (148,666/232,034) of the total downloads of open source EHR projects on SourceForge. In total, 19% (10/54) of open source EHR systems have installation and user guides along with demonstration versions. The mean project age is 7 years with a range of 2 to 14 years.

Table 3. The characteristics of the top 10 downloaded free and open source software electronic health record systems.

Product name	Installation guide	Demonstration	User guide	Top downloads according to country	Development status	Start year	Age, years
OpenEMR	Yes	Yes	Yes	United States	Production/stable	2002	12
OpenMRS	No	Yes	No	United States	Production/stable	2010	4
Care2x	Yes	Yes	Yes	United States	Production/stable	2002	12
OpenClinic GA	Yes	Yes	Yes	India	Production/stable	2010	4
Open Hospital	Yes	N	Yes	India	Production/stable	2006	8
FreeMED	Yes	Yes	Yes	United States	Production/stable	2000	14
GNU Health	Yes	Yes	Yes	India	Production/stable	2006	8
HOSxP	Yes	Yes	No	Thailand	Production/stable	2002	12
Tolven Health Record	Yes	Yes	Yes	France	Production/stable	2006	8
OSCAR McMaster	Yes	Yes	Yes	Canada	Production/stable	2001	13

In terms of project (spoken) language, a large number of projects were written in English (46%, 25/54), but surprisingly, 28% (15/54) were being developed in one or more languages besides English. It was found that 11% (6/54) of the projects did not specify a language (Table 4). Data shows that the number of downloads for projects with multiple spoken languages was very high (n=187,933); with a download rate three times the one-language projects (n=48,402). These numbers indicate a global interest in F/OSS EHR development and prove that the number of language translations is positively related to project success and popularity.

Table 4. Frequency of spoken languages (N=54).

Language	Frequency, n (%)
English	25 (46%)
Non-English	4 (7%)
Multiple (English plus other)	15 (28%)
Multiple (non-English)	4 (7%)
Unspecified	6 (11%)

Survey Data

A total of 103 developer survey responses were successfully collected showing the developer's characteristics, background, and their reasons to contribute to health care open source projects. Survey respondents were primarily male (94%, 94/99) with an age range of 25 to 55 years old. The majority of developers were American (37%, 36/98) and 42% (41/98) lived in the United States. Employees made up 58% (58/100) of the sample, and self-employed developers constituted 33.0% (33/100). Approximately 58% (58/99) of developers were married and 18.0% (18/100) had children older than age six. As F/OSS developers tend to be highly educated, many had graduate level (43.0%, 43/100), professional level (18.0%, 18/100), or undergraduate level (36.0%, 36/100) education. Only 3.0% (3/100) of F/OSS developers reported not having formal education beyond high school.

As has been found in mainstream F/OSS projects, the majority of health care F/OSS developers participated in the projects as part of a paid activity. Nearly 34.7% (35/101) of the contributors received direct payment for developing F/OSS. Another 30.7% (31/101) received direct payment for either managing or supporting F/OSS while 34% (32/94) received no payment at all for contributing to the projects. Most of the developers contributed by writing code during their off-work hours (41.6%, 42/101) and (39.6%, 40/101) during their work hours. Almost half of the respondents reported employer awareness of their F/OSS work (42%, 42/99). A number of respondents (32%, 32/99) worked on F/OSS as part of their employment. Some employers (17%, 17/99) were unaware of the developers work on F/OSS, but very few (3%, 3/99) did not want them to contribute to F/OSS development.

Contributors' efforts have been measured by the number of hours spent on a project per week. Respondents' answers show an average of 9.8 hours per week spent on F/OSS projects. Health care open source developers were asked how many projects they had contributed to. Nearly 49.0% (50/102) worked on one project while 47% (48/94) had worked on 2 to 5 open source projects. The average number of projects was 1.6 with a maximum number of 5 projects.

A significant number of respondents (47.0%, 47/100) reported 75% to 100% of their code was included in a F/OSS project. Around one third (43.0%, 43/100) reported less than 25% of

their code was included in the final project. The majority of contributors (50%, 48/96) wrote between 500 and 5000 code lines, while only 30% (29/96) wrote more than 10,000 lines. The average lines of code were 1776 with a maximum of 5000 lines.

Respondents were asked to answer several questions regarding their F/OSS project to analyze their opinions, motivations, and habits toward F/OSS projects. Interestingly, developers had different motivations to participate in F/OSS. The top reasons for contributing to a health care open source project were based on enjoyment-related intrinsic motivation; the project was "important and visible" (47.5%, 48/101) or "technically interesting" (47.5%, 48/101) for them. Approximately 20.8% (21/101) indicated community-based related intrinsic motivation as their reason to contribute and stated that they knew people working on the project.

The survey results confirmed that many developers start developing F/OSS to give back to the community. Around 40% (39/97) considered it important or very important to give back to the community and 38% (37/96) considered the interaction with like-minded programmers to be important. A significant number of respondents were motivated to promote the mode of development and the ideal of freely modifiable software (47%, 47/99), while the remaining developers were motivated to provide alternatives to proprietary software (53%, 50/93). Some developers (27%, 25/97) began developing F/OSS software by modifying it to fit with their requirements or to fix the bugs in their existing software (32%, 32/97), and 26% (25/96) of them were interested in learning how the program worked.

Discussion

Principal Findings

Open source projects continue being started at a rapid pace, but sustainability of the projects appears to be a challenge. A study in 2005 found 45 F/OSS EHR applications [29] compared to the 54 found in this study. This represents a 16.6% increase in the past 9 years. However, most of the projects are different, suggesting sustainability is a significant challenge. Functional EHR systems are complex with many required subsystems and modules, requiring robust development and change management processes in order to achieve high-quality "production-grade" software for the mission-critical environment of a hospital or clinic. Certification is an added challenge in some markets [30]. Despite these challenges, F/OSS EHR systems have been viewed as excellent options for community health clinics, small practices, and hospitals that are under-capitalized in the United States and overseas [8].

In terms of open source licenses, the majority F/OSS projects use GPL, which allows free modification but requires that any modification be contributed back to that open source project under the same license. This creates a "poison pill," making it difficult for commercial entities to integrate GPL open source components into a module that also has proprietary software. This can be viewed as a "restriction" with regards to open source licensing. Restrictiveness of the license also affects a number of contributors and accessibility of the source code [23].

With regards to developers, this study found the average F/OSS contributor is well-educated, young, and male; the F/OSS community is male dominant, as women contributors are only 2%. The reasons for gender inequities remain unclear but may involve women facing hybrid discriminations from a F/OSS community [31]. Our research shows marital status and having kids older than 6 years old increases the probability a person will volunteer [32,33].

One of the most common aspects of F/OSS is that they are based on voluntary efforts of developers. The goal of this research was to understand what motivates these developers to contribute to a health care F/OSS project. Learning new skills and becoming a better programmer is one of the motivations this research found (35%, 34/96), which is similar to previous findings about mainstream F/OSS developers (36.5%) [28]. Learning and acquiring new skills appears to be another important motivation for many developers.

Intrinsic motivations, such as altruism, are high among health care developers; 47.5% (48/101) of developers worked on the programs for this reason. Only 16% of developers in other research papers stated that altruism was their prime motivator [21]. Contrary to our research findings, some previous research papers [34,35] found that only a minority were motivated by extrinsic motivation, which involved payments for their participation in F/OSS projects.

One of the unique findings in this research is that 74.2% (75/101) of health care F/OSS developers are not health care practitioners, and 45% (45/99) do not work in the health care field. Being a developer outside of the health care field can be a core problem for the development of usable clinical software with a high degree of functionality. This may serve to explain why open source EHRs have limited functionality today.

Although the open source projects solve licensing cost problems, there is a need for maintenance and implementation costs, which can be a barrier for organizations without health IT expertise. There are many open source EHR options on SourceForge, but not all of them can fit in the clinical workflow in the United States or in certain hospitals.

Challenges

The availability of motivated developers and the need to continuously improve EHR systems will likely mean F/OSS EHR will continue to be part of the health care software landscape despite the many challenges these projects face today. Characterizing the challenges and benefits of adopting open source EHR will be important in understanding the value of these systems.

Usability

This study shows a potential weakness for F/OSS EHR projects stemming from developer background, and their ability to understand some of the nuances in health care workflows. A large fraction of developers (74.2%, 75/101) were not health care practitioners and few developers (54%, 54/99) had worked in the health care field. A potential solution to this issue would be to have programs that give F/OSS EHR developers direct access to providers and care venues.

Interoperability

Many hospitals would benefit greatly from integrated software, rather than a disparate group of systems. However, the F/OSS option can be difficult to interface with commercial systems and may require personnel with multiple skill sets [36]

Privacy and Security

Maintaining patient privacy through robust security is a critical aspect of any EHR system. A study of F/OSS health care systems found the information is often not safeguarded with consent or privacy policies and offers limited protection against unauthorized access or release of information [37]. However, we did find one open source EHR system (Tolven eCHR), which supported encryption of health care data at the row level in a relational database. This would make data unreadable even to the database administrator of the system. This is a high degree of security and not typical among EHR systems at the time.

Lack of Financial and Professional Expertise

Low acquisition, installation, and ownership costs make the F/OSS system an excellent option for organizations with limited capital. However, in general, the health care environments that could benefit the most from F/OSS EHR systems tend to have low information technology capital budgets and very limited access to health informatics professionals. These challenges may prevent the successful implementation of any health IT technology, including F/OSS EHR [36]. Our study shows that developing countries adopted F/OSS EHR systems but the major adoptions were in North America after the HITECH act. This suggests that even F/OSS EHR system implementations require substantial financial and workforce resource capacity to succeed.

Cost

Although F/OSS software does not have licensing costs, effective implementation still requires skilled staff, time for installation, and time for learning the software. F/OSS EHR implementation cost can be as high as the proprietary software because of the add-ons, consultation costs, and need for assistance [36]. An important aspect of the total cost of ownership of an EHR system is usability. Poor usability can directly impact the productivity of expensive health care providers and support personnel. The ability to freely modify F/OSS EHR systems and optimize them for local use likely translates into lower overall cost.

Skilled Information Technology Personnel

Adopting F/OSS EHR requires a large number of IT employees with specific programming skills who understand the program well. A hospital that chooses to use F/OSS EHR will need to hire developer IT staff and contract with extra IT vendors or consultant support [36].

Limited Functionality of Clinical Decision Support

At an operational level, F/OSS software also presented reduced functions in decision support and knowledge management. Clinical reasoning, guidelines and protocols, quality assurance, and integrated care were rather limited or nonexistent in most applications [37]. Developers lack medical knowledge as the majority of them are not health care practitioners, which may affect the efficacy of clinical decision support functionality.

Lack of Liability and Accountability

One of the major shortcomings is the lack of liability and accountability in F/OSS. Our study did not address this issue. Few studies of F/OSS in health care have addressed the risks of using F/OSS. F/OSS EHR projects come without warranties regarding the development, release date, or fulfillment of functionality. F/OSS EHR systems do not have a commercial entity providing support and tend to rely on a volunteer community. This can be a significant risk for a software system that supports the core business of a hospital [36].

Scope and Limitations

Data on F/OSS EHR projects were primarily collected from SourceForge. Therefore, this study is a snapshot in time, with projects being added and deleted before or after data collection. This amount of "churn" reflects a dynamic software category where our findings may not apply in the future.

Not all F/OSS EHR projects were listed on SourceForge; some projects are managed through independent websites. Therefore, some of the projects on SourceForge were outdated, may have contained inaccurate information, and were a small representation of F/OSS EHR projects as a whole. SourceForge was the main source for most of the data because it tends to be one of the preferred project management platforms for the open source movement and has numerous projects and registered developers.

Survey questions were mostly closed-ended which restricted respondents' answer options. However, we did accommodate all possible answers by providing an additional "other" option for some questions. There is a possibility that respondents misunderstood questions as there was no usability testing done on the survey prior to using it in the general F/OSS EHR community; however, we did not receive any respondent requests to clarify any questions. Furthermore, the survey respondents do not reflect the entire population of open source developers, but we believe it is a reasonable representative of the target population.

Conclusions

This study highlights a number of important aspects of F/OSS EHR. Open source software systems seem to be important to some health care organizations; however, only four F/OSS EHR systems are ONC-certified in the United States, which creates a barrier to broader use. Health care open source software also currently lacks directed corporate or governmental support for sustainability and growth of these software programs. We hope this research underlines the challenges that hinder the future of F/OSS and provides avenues for future research to study and improve adoption of F/OSS systems in the United States.

Conflicts of Interest

None declared.

Multimedia Appendix 1

[PDF File (Adobe PDF File), 181KB - medinform_v5i1e6_app1.pdf]

References

1. McDonald CJ, Schadow G, Barnes M, Dexter P, Overhage JM, Mamlin B, et al. Open Source software in medical informatics--why, how and what. Int J Med Inform 2003 Mar;69(2-3):175-184. [Medline: 12810121]

2. Barnett GO, Justice NS, Somand ME, Adams J, Waxman B, Beaman P, et al. COSTAR— a computer-based medical information system for ambulatory care. Proc IEEE 1979 Sep;67(9):1226-1237. [doi: 10.1109/PROC.1979.11438]

3. Henry JW, Pylypchuk Y, Searcy T, Patel V. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015.: The Office of National Coordinator for Health Information Technology; 2016 May. URL: http://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php [accessed 2016-09-21] [WebCite Cache ID 6kgMi77LN]

4. Ye Y, Kishida K. Toward an understanding of the motivation open source software developers. New York: ACM Press; 2003 Presented at: ICSE '03 Proceedings of the 25th International Conference on Software Engineering; 2003 May 3-10; Portland, OR p. 416-429.

5. Gans D, Kralewski J, Hammons T, Dowd B. Medical groups' adoption of electronic health records and information systems. Health Aff (Millwood) 2005;24(5):1323-1333 [FREE Full text] [doi: 10.1377/hlthaff.24.5.1323] [Medline: 16162580]

6. Ajami S, Bagheri-Tadi T. Barriers for adopting electronic health records (EHRs) by physicians. Acta Inform Med 2013;21(2):129-134. [doi: 10.5455/aim.2013.21.129-134] [Medline: 24058254]

7. Fleming NS, Culler SD, McCorkle R, Becker ER, Ballard DJ. The financial and nonfinancial costs of implementing electronic health records in primary care practices. Health Aff (Millwood) 2011 Mar;30(3):481-489 [FREE Full text] [doi: 10.1377/hlthaff.2010.0768] [Medline: 21383367]

8. Aminpour F, Sadoughi F, Ahamdi M. Utilization of open source electronic health record around the world: a systematic review. J Res Med Sci 2014 Jan;19(1):57-64 [FREE Full text] [Medline: 24672566]

9. Ratwani RM, Benda NC, Hettinger AZ, Fairbanks RJ. Electronic health record Vendor adherence to usability certification requirements and testing standards. JAMA 2015 Sep 08;314(10):1070-1071. [doi: 10.1001/jama.2015.8372] [Medline: 26348757]

10. Briggs B, Carter-Templeton H. Electronic health record customization: a quality improvement project. OJNI 2014 Oct 14;18(3).

11. Tate C. AMDIS 2015. 2015 Jun. URL: http://amdis.org/wp-content/uploads/2015/06/Klas-Review_Tate1.pdf [accessed 2016-09-21] [WebCite Cache ID 6oK6PeQFA]

12. Yellowlees PM, Marks SL, Hogarth M, Turner S. Standards-based, open-source electronic health record systems: a desirable future for the U.S. health industry. Telemed J E Health 2008 Apr;14(3):284-288. [doi: 10.1089/tmj.2007.0052] [Medline: 18570554]

13. Raymond E. The cathedral and the bazaar. Know Techn Pol 1999 Sep;12(3):23-49. [doi: 10.1007/s12130-999-1026-0]

14. Kobayashi S. Open source software development on medical domain. In: Kalloniatis C, editor. Modern Information Systems. Croatia: Intech; Jun 13, 2012:174.

15. Maglogiannis I. Towards the adoption of open source and open access electronic health record systems. J Healthc Eng 2012 Mar;3(1):141-162. [doi: 10.1260/2040-2295.3.1.141]

16. Chełkowski T, Gloor P, Jemielniak D. Inequalities in open source software development: analysis of contributor's commits in Apache Software Foundation projects. PLoS One 2016;11(4):e0152976 [FREE Full text] [doi: 10.1371/journal.pone.0152976] [Medline: 27096157]

17. Goulde M, Brown E, Rymer J, Bough V, Hartzband D, PostTechnical Research, Forrester Research. Open Source Software: A Primer for Health Care Leaders. Oakland, California: Forrester Consulting; 2006 Mar. URL: http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/PDF%20O/PDF%20OpenSourcePrimer.pdf [accessed 2017-02-07] [WebCite Cache ID 6o6h742Ua]

18. The Future of Open Source Survey Results 2012. 2012 May 21. URL: http://www.northbridge.com/2012-open-source-survey [WebCite Cache ID 6oAu3zfwN]

19. The Future of Open Source Survey Results 2014. 2014. URL: http://www.northbridge.com/2014-future-open-source-survey-results-0 [accessed 2016-09-21] [WebCite Cache ID 6kgOkYqBZ]

20. Godfrey M, Tu Q. Evolution in open source software: a case study. 2000 Oct 11 Presented at: ICSM '00 Proceedings of the International Conference on Software Maintenance; 2000 Oct 11 - 14; San Jose, CA p. 131. [doi: 10.1109/ICSM.2000.883030]

21.   Hars A, Ou S. Working for free? Motivations for participating in open-source projects. 2002 Presented at: Proceedings of the 34th Annual Hawaii International Conference on System Sciences; 2001 Jan 6; Maui, HI. [doi: 10.1109/HICSS.2001.927045]

22.   Sen R, Subramaniam C, Nelson ML. Determinants of the choice of open source software license. J Manage Inform Syst 2009 Feb 10;25(3):207-240. [doi: 10.2753/MIS0742-1222250306]

23.   Lerner J, Tirole J. The scope of open source licensing. J Law Econ Organ 2005 Apr 01;21(1):20-56. [doi: 10.1093/jleo/ewi002]

24.   Janamanchi B, Katsamakas E, Raghupathi W, Gao W. The state and profile of open source software projects in health and medical informatics. Int J Med Inform 2009 Jul;78(7):457-472. [doi: 10.1016/j.ijmedinf.2009.02.006] [Medline: 19321384]

25.   Chengalur-Smith I, Sidorova A, Daniel S. Sustainability of free/libre open source projects: a longitudinal study. J Assoc Inf Syst 2010;11(11).

26.   Crowston K, Scozzi B. Open source software projects as virtual organisations: competency rallying for software development. IEE Proc Softw 2002;149(1):3. [doi: 10.1049/ip-sen:20020197]

27.   Midha V, Palvia P. Factors affecting the success of open source software. J Syst Softw 2012 Apr;85(4):895-905. [doi: 10.1016/j.jss.2011.11.010]

28.   Arora S, Waterman A, David P. Free/Libre/Open Source Software Group - Survey. URL: http://web.stanford.edu/group/floss-us/questionnaire/FLOSS-US-Questionnaire.html [accessed 2016-09-21] [WebCite Cache ID 6kgRUkKcM]

29.   Hogarth MA, Turner S. A study of clinically related open source software projects. AMIA Annu Symp Proc 2005:330-334 [FREE Full text] [Medline: 16779056]

30.   Karopka T, Schmuhl H, Demski H. Free/Libre open source software in health care: a review. Healthc Inform Res 2014 Jan;20(1):11-22 [FREE Full text] [doi: 10.4258/hir.2014.20.1.11] [Medline: 24627814]

31.   Mahmod M, Yusof S, Dahalin Z. Women contributions to open source software innovation: a social constructivist perspective. 2010 Presented at: 2010 International Symposium on Information Technology; 2010 Sept 2; Kuala Lumpur p. 1433-1438. [doi: 10.1109/ITSIM.2010.5561496]

32.   Selbee LK, Reed PB. Patterns of Volunteering Over the Life Cycle. Canada: Statistics Canada; 2000. URL: http://www3.carleton.ca/casr/Patterns%20of%20Vol.pdf [WebCite Cache ID 6oAw1ULg0]

33.   Hayghe HV. Volunteers in the U.S.: who donates the time? Mon Labor Rev 1991 Feb;114(2):17-23. [Medline: 10110196]

34.   Lakhani KR, Wolf R. Why hackers do what they do: understanding motivation and effort in free/open source software projects. In: Feller J, Fitzgerald B, Hissam S, Lakhani KR, editors. Perspectives on Free and Open Source Software. Cambridge, MA: MIT Press; 2005.

35.   Luthiger B, Jungwirth C. Pervasive fun. First Monday 2007 Jan 01;12(1).

36.   Schmuhl H, Heinze O, Bergh B. Use of open source software in health care delivery - results of a qualitative field study. Contribution of the EFMI LIFOSS Working Group. Yearb Med Inform 2013;8:107-113. [Medline: 23974556]

37.   Flores Zuniga AE, Win KT, Susilo W. Functionalities of free and open electronic health record systems. Int J Technol Assess Health Care 2010 Oct;26(4):382-389. [doi: 10.1017/S0266462310001121] [Medline: 20974022]

Abbreviations

COSTAR: Computer Stored Ambulatory Record
EHR: electronic health record
F/OSS: free and open source software
GPL: general public license
HITECH: Health Information Technology for Economic and Clinical Health
IT: information technology
ONC: Office of the National Coordinator

Original Paper

Ontology-Driven Search and Triage: Design of a Web-Based Visual Interface for MEDLINE

Jonathan Demelo[1*], BSc; Paul Parsons[2*], PhD; Kamran Sedig[1*], PhD

[1]Insight Lab, Department of Computer Science, Western University, London, ON, Canada

[2]Purdue Polytechnic Institute, Department of Computer Graphics Technology, Purdue University, West Lafayette, IN, United States

[*]all authors contributed equally

Corresponding Author:
Paul Parsons, PhD
Purdue Polytechnic Institute
Department of Computer Graphics Technology
Purdue University
Knoy Hall
401 N Grant St
West Lafayette, IN, 47907
United States
Phone: 1 765 494 0511
Fax: 1 765 494 9267
Email: parsonsp@purdue.edu

Abstract

Background: Diverse users need to search health and medical literature to satisfy open-ended goals such as making evidence-based decisions and updating their knowledge. However, doing so is challenging due to at least two major difficulties: (1) articulating information needs using accurate vocabulary and (2) dealing with large document sets returned from searches. Common search interfaces such as PubMed do not provide adequate support for exploratory search tasks.

Objective: Our objective was to improve support for exploratory search tasks by combining two strategies in the design of an interactive visual interface by (1) using a formal ontology to help users build domain-specific knowledge and vocabulary and (2) providing multi-stage triaging support to help mitigate the information overload problem.

Methods: We developed a Web-based tool, Ontology-Driven Visual Search and Triage Interface for MEDLINE (OVERT-MED), to test our design ideas. We implemented a custom searchable index of MEDLINE, which comprises approximately 25 million document citations. We chose a popular biomedical ontology, the Human Phenotype Ontology (HPO), to test our solution to the vocabulary problem. We implemented multistage triaging support in OVERT-MED, with the aid of interactive visualization techniques, to help users deal with large document sets returned from searches.

Results: Formative evaluation suggests that the design features in OVERT-MED are helpful in addressing the two major difficulties described above. Using a formal ontology seems to help users articulate their information needs with more accurate vocabulary. In addition, multistage triaging combined with interactive visualizations shows promise in mitigating the information overload problem.

Conclusions: Our strategies appear to be valuable in addressing the two major problems in exploratory search. Although we tested OVERT-MED with a particular ontology and document collection, we anticipate that our strategies can be transferred successfully to other contexts.

(JMIR Med Inform 2017;5(1):e4) doi:10.2196/medinform.6918

KEYWORDS

Introduction

Overview and Significance

Seeking information within the published medical literature is important in many domains and contexts [1,2]. Diverse users need to search the literature including physicians [3], medical students [4], cytogeneticists [5], and patients and their relatives [6]. Searches can be roughly categorized into 2 types: *lookup* and *exploratory* [7]. Lookup searches are closed-ended, having precise results and little need for examining and comparing result sets. Exploratory searches, however, are open-ended, having imprecise results and often requiring significant time and effort to work with result sets in order to satisfy the original information need. Examples of exploratory searches with open-ended goals include making evidence-based decisions and updating knowledge to stay abreast of current research findings [2,8]. Although significant progress has been made in supporting lookup searches, exploratory searches are still not well supported, and open-ended search goals are often quite difficult to achieve [2,9,10]. Common barriers to finding relevant medical information include the time it takes to perform searches [3,11], the increasing scope of topical coverage [2], and the information overload that arises from dealing with large result sets [2,3,11-13].

One of the most popular collections of published medical literature is MEDLINE, which comprises more than 25 million documents and is growing every year. The most common means of searching MEDLINE is PubMed, a free search engine and Web interface [14]. Although the search capabilities in PubMed have improved in recent years, there can still be a considerable burden on users when seeking information in the context of exploratory search, due to at least two major problems: (1) the difficulty in articulating information needs using accurate vocabulary and (2) the large number of documents that can be returned from searches. Many users do not have the proper vocabulary to construct effective queries [15,16], which is especially true in medical and health contexts [17-20]. When uncontrolled vocabularies are used, there is no guarantee that concepts are expressed with the same terms in different contexts [13,21]. For instance, if an article contains the term *eye hamartoma*, and a user searches for the vaguer term *eye growth*, there may not be a close match. Thus, without proper terminological knowledge, effective searching can be quite difficult. Adding to the difficulty of searching effectively is the large number of documents that can be returned, which leads to information overload problem [9,22,23]. Dogan et al [2] note

that at least one-third of PubMed searches return 100 or more documents. In our own testing, searches for common terms (eg, "breast cancer" or "brain tumor") returned many thousands of documents.

Interfaces to most search engines, including PubMed, use simple text boxes into which users enter query terms. This interface style does not assist users in articulating their information needs [24] and works well only for lookup search tasks [25,26]. For example, if a user is interested in finding information about "liver," but is not sure what terms are relevant in articulating a query, he or she must simply enter "liver" into the search box. As the query is vague, a very large set of documents is returned—almost one million documents spanning over 4900 pages when using PubMed (Figure 1).

Multiple strategies have been employed to help support query formation in exploratory search contexts by replacing the standard text box, including faceted search [27], visualization widgets [28], query previews [29], and hierarchical presentation of expansion terms [30]. The common theme among these strategies is that meaningful information is extracted from the document collection and then represented in a manner that can help the searcher recognize terms that will more accurately describe the information they are seeking. Such strategies promote recognition over recall, not relying on users having to know and retrieve correct vocabulary from memory [24].

We present Ontology-Driven Visual Search and Triage Interface for MEDLINE (OVERT-MED), a Web-based visualization tool that addresses two major difficulties in searching large document collections: (1) the difficulty in articulating information needs with useful vocabulary and (2) the difficulty in dealing with large search result sets. To address the first difficulty, we propose the idea of using a formal ontology to help users build domain-specific knowledge and vocabulary. To test this, we have implemented a searchable index of the Human Phenotype Ontology (HPO) that provides users with suggestion terms that are related to their information needs. To address the second difficulty, OVERT-MED supports multistage interactive triaging of search results using interactive visualization techniques. We use a custom-built index of MEDLINE, which comprises approximately 25 million documents, as our searchable collection of medical literature. Although OVERT-MED has been initially developed for use with a particular ontology and document collection, we expect that our design ideas will transfer to other contexts. The following subsections provide background information and discuss related work.

XSL•FO

RenderX

Figure 1. A screenshot of PubMed showing results from searching for "liver."
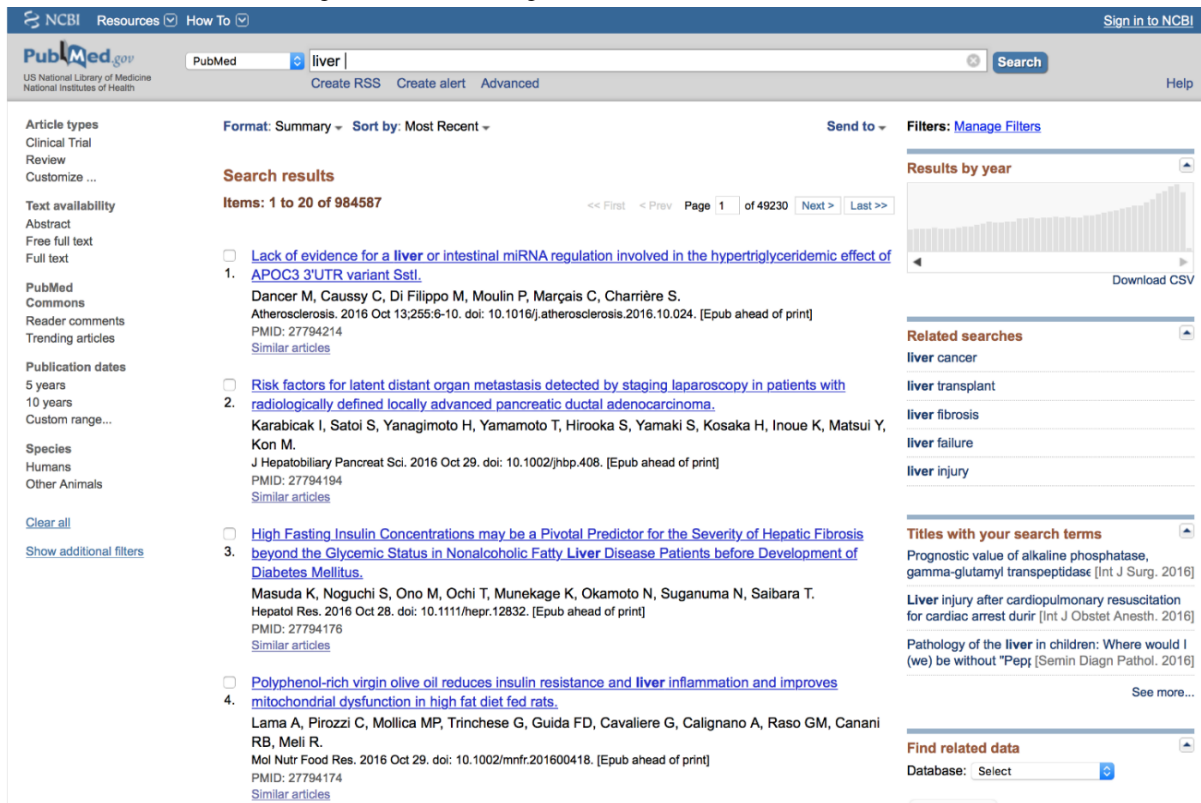
Ontologies

One way to meaningfully extract and model information from a domain is to construct an ontology [31,32]. An ontology represents concepts and their relationships using a standard vocabulary [32]. Ontologies serve many practical functions, including clarifying the structure of knowledge within a domain, providing a common vocabulary, enabling computational analysis, and supporting knowledge sharing [31-33]. Ontologies often capture concepts within a domain at multiple levels of abstraction. For instance, an anatomy ontology may have a concept *body*, a sub-concept *face*, a further sub-concept *nose*, and so on. The concepts in an ontology can be represented using many different structures, including trees and different types of graphs.

The ontology we are using, HPO, has been curated by domain experts in an attempt to capture all phenotypic abnormalities that are commonly encountered in human monogenic disease [34]. In our previous work with genomics researchers, we learned of the importance of HPO in their workflow, including in activities involving literature search [5]. HPO is widely used in the biomedical field, is regularly updated, and has a high level of quality control. It is also available for download in the popular Open Biomedical Ontologies (OBO) and Web Ontology Language (OWL) formats. For these reasons, we believe HPO is ideal for testing our proposal of using ontologies to address the vocabulary problem. It should be noted that we are not suggesting HPO is better than other ontologies or that it should be used in all contexts. HPO is only one of the many ontologies that could be used to support exploratory search, and search systems should make use of whichever ontologies are most appropriate for given contexts.

Document Triage

Triaging is an activity that involves determining the relevance of documents to an information need [35]. Triaging activities are often time-constrained and require quick assessment of relevance with incomplete knowledge. For example, a search may return hundreds or thousands of potentially relevant documents. As it is not feasible to read each one in detail, users must sort through the documents and quickly assess their relevance based on incomplete knowledge of their contents. Research suggests that triaging takes place in 3 successive stages: (1) the "multiple document" stage, where initial relevance judgments are made to select documents from a set without careful examination; (2) the "individual document" stage, where individual documents are examined in more detail and categorized (eg, kept or rejected); and (3) the "further reading" stage, where a small set of documents are read in depth to extract relevant information and satisfy the original information need [36]. In addition, research shows that triaging often occurs in a cyclical and iterative fashion, where the above stages are revisited multiple times [37].

Search Result Visualization

Most search interfaces present results in a traditional list-based manner, where documents are ranked and textually represented using a title and various metadata. While not a problem for simple lookup search tasks, traditional list-based representations are not effective in supporting exploratory search tasks, which are typically open-ended and involve complex information needs [38]. Although lists are familiar and simple, studies show that users rarely examine lists fully or carefully [39] and seldom venture past the first few pages of results [40]. Scanning through long lists can be tedious and cognitively demanding.

Visualizations of search results can overcome some of the problems associated with textual list-based representations by shifting cognitive burden onto the perceptual system. For instance, whereas visualizations can be scanned freely by the eyes, text must be scanned sequentially, requiring more time and cognitive effort to detect patterns and relationships [41,42]. In addition, visualizations can encode a significant amount of information within a small space, removing the need to navigate multiple pages to view search results. Previous work has demonstrated the utility of visualizations in document search, exploration, and analysis [43,44].

Related Work

Some researchers have recognized the value of using ontologies to better support search activities (eg, [13,45]). The central focus of this research is term extraction and mapping, which is done using text mining and natural language processing techniques. In this body of work, ontologies are used to improve search performance computationally without involving users. The fundamental difference compared with our work is that we use ontologies to help users develop knowledge and domain-specific vocabulary—that is, the focus is on the user rather than on algorithms and other computational processes. Our approach is important in contexts where users have valuable knowledge and context-specific goals that cannot be replaced by computation—in other words, users need to be kept "in the loop."

Other researchers have focused on developing interfaces to MEDLINE as alternatives to PubMed. For example, Wei et al have developed PubTator, a PubMed replacement interface that uses multiple text mining algorithms to improve search results [46]. PubTator also offers some support for document triaging. Whereas PubTator appears interesting and useful, it relies on queries being input into the standard text box, and it presents results in a typical list-based fashion. Thus, it is not aimed at addressing either of the two problems we are attempting to address with OVERT-MED—that is, the vocabulary problem and the information overload problem. Other alternative interfaces that offer interesting features but do not address either of the two problems include SLIM [47] and HubMed [48]. An alternative interface that potentially provides support in addressing the first problem is iPubMed [49], which provides fuzzy matches to search results. An alternative interface that may provide support in addressing the second problem is refMED [50], which provides minimal triaging support through relevance ranking. A for-profit private tool, Quertle, appears to use visualizations to mitigate the information overload problem, although very few details are publicly available. Lu [51] provides a detailed survey that includes many other alternative interfaces to MEDLINE, although none are aimed at solving either of the two problems that we are addressing here.

In summary, no extant research explores the combination of (1) ontologies to help build domain-specific knowledge and vocabulary when users need to be kept "in the loop" and (2) triaging support using interactive visualizations to help mitigate the information overload problem. The following sections provide details about our approach to addressing these issues.

Methods

Overview

We developed OVERT-MED to test our proposed solutions to the two problems described hereinbefore. To anchor our research in a specific context, we chose MEDLINE as our document collection. MEDLINE offers an interesting testbed because of its popularity and size. We developed a custom index of MEDLINE so that it can be queried from the front end of OVERT-MED. We have also indexed HPO to help users build knowledge and domain-specific vocabulary.

Indexing of MEDLINE and HPO

We downloaded the entire MEDLINE database, which has been made freely available by the National Library of Medicine (NLM) for research purposes. The MEDLINE database consists of article "citations," which are essentially article metadata, including authors, journal title, Medical Subject Heading (MeSH) keywords, publication date, and other fields. Also included in each citation is the abstract text. We developed a custom index using the open-source Apache Solr and Lucene projects. Lucene supports full-text indexing and search functionality, and Solr is a search platform that runs on the Lucene index. To rank documents, Lucene uses the well-known term frequency-inverse document frequency (tf-idf) scheme [52]. Lucene also ranks results based on an internal similarity measure that generates a vector space model (VSM) score [53], using index terms as dimensions and tf-idf values as weights. We have described our indexing strategy in greater detail earlier [5].

HPO is a formal ontology of human phenotypic abnormalities found in human disease [34]. Each entry in HPO describes a phenotypic abnormality such as melanoma or hepatoblastoma. HPO is under active development and currently contains more than 11,000 terms. We have also indexed HPO in our Lucene index. HPO contains multiple fields for each phenotype in the ontology, including name, definition, id, synonyms, and commentary from domain experts. We index all fields to provide robust vocabulary suggestions—when a user enters a term, all fields in the index are examined, which provides much more useful information than would result from looking for only exact matches on the phenotype name. This is described using an example in greater detail in the following.

Development and Architecture

We developed OVERT-MED as a Web-based tool that runs in any modern browser. It connects to a Web server that stores our indices and handles search requests (via our Solr search server). We have developed a series of scripts to retrieve MEDLINE updates from the NLM public ftp site and to construct the indices for MEDLINE and HPO in our Lucene index. We have also developed an application programming interface (API) that handles requests for searches and other basic functions. The front-end has been developed using HTML5, CSS, and JavaScript. The visualizations have been developed using D3.js [54], a popular JavaScript visualization library. Figure 2 provides a diagrammatic overview of the architecture of the OVERT-MED system.
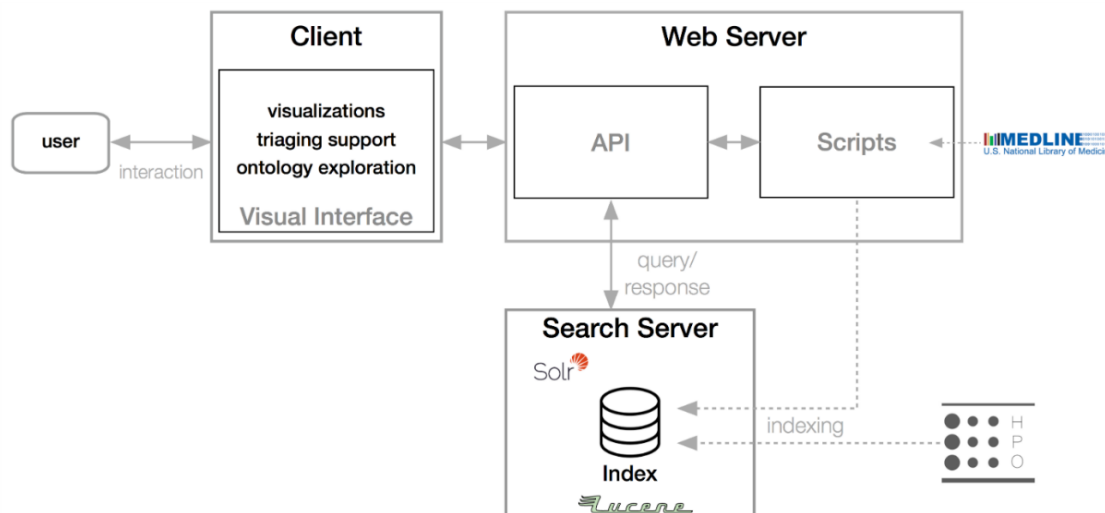
Results

Ontology Term Suggestion

OVERT-MED uses HPO to help users better articulate their search needs through a technique we call *ontology term suggester*. Users enter terms into a text box, and a set of suggestions (phenotypes) are provided. The suggestions are updated in real-time as a user types each character. In addition, to providing better terminological support, we look for matches on both the phenotype names as well as descriptions and expert commentary on the phenotypes (these are not shown to users, but are indexed on our server). For example, a user may be interested in finding articles related to the term "liver," but may not have sufficient vocabulary to articulate a useful query involving relevant terms. Figure 3 shows the ontology term suggester after typing "liver" into the search box. Phenotypes related to the liver are displayed. Results such as "Growth hormone deficiency" and "Ascites" are displayed because they have a connection to the liver—the effects of growth hormone are mediated by insulin-like growth factor, which is produced primarily in the liver; and ascites is commonly associated with liver disease. Many of the returned phenotypes do not have the term *liver* in their name, but are related to the liver. In a traditional search interface, there is no way for a user to get from "liver" to "ascites" or "growth hormone deficiency." Finally, because users may not understand a particular phenotype (eg, congenital diaphragmatic hernia), selecting the "?" button will open a new tab and load the official entry in the HPO Web browser. From there users can find more details, including associated genes and diseases. This search strategy can help users build knowledge of the domain and vocabulary that can be used to enhance cognitive performance and exploration.

Figure 3. The ontology term suggester, showing results from typing "liver."

Sensitivity Encoding for Query Refinement

A well-known problem in open-ended search tasks is that potentially relevant results may not be displayed if they do not meet the specified search criteria. For example, when searching for a house to buy, users often have ill-formed criteria, such as price range, number of bedrooms and bathrooms, yard size, location, and so on. Although certain search criteria may be specified (eg, 4 bedrooms, under $200,00), results that do not meet the criteria may also be relevant, such as a house that has only 3 bedrooms but is a great price. When using visualizations to support such search tasks, certain criteria can be relaxed and results that do not meet certain criteria can be visually encoded in different ways. For instance, results that do not meet number of bedrooms can be encoded with 1 color; results that do not meet yard size can be encoded with another; and so on. Visually encoding this type of information can provide cues to users to adjust their search criteria so that potentially relevant results are included. This visualization strategy, known as sensitivity encoding, has been shown to be beneficial in a number of contexts [55,56].

Although OVERT-MED supports the selection of precise phenotype names, the exact combination of words in a name may be too restrictive, and may not provide the most relevant results. For example, a user may select the phenotype *progressive external ophthalmoplegia*. Our index shows 811 articles associated with this specific phenotype. However, users may be interested in articles associated with different variations of the words—for example, *progressive opthalmoplegia* or *external opthalmoplegia*. We use a set of *Sensitivity Encoded Query Selectors* in OVERT-MED to handle this issue. When a phenotype is selected, we perform searches on our index using all possible combinations of the words and then visually encode the size of the result set. Figure 4 shows the result of a user selecting "progressive external opthalmoplegia." The number of matching articles for each combination is provided numerically and encoded visually using the length of the bar next to each combination. From Figure 4, we can see that if the user relaxes the term to "progressive ophthalmoplegia," an additional 104 articles show up in the index and with "external opthalmoplegia," an additional 418 articles show up. Without such a sensitivity encoding strategy, many of these potentially relevant results would not be made available. As users are often interested in more than 1 phenotype, multiple phenotypes can be selected, each of which is subjected to the same sensitivity encoding process. Figure 5 shows a second phenotype, congenital fibrosis of extraocular muscles, being added.

Figure 4. A set of sensitivity-encoded query selectors for "progressive external ophthalmoplegia."
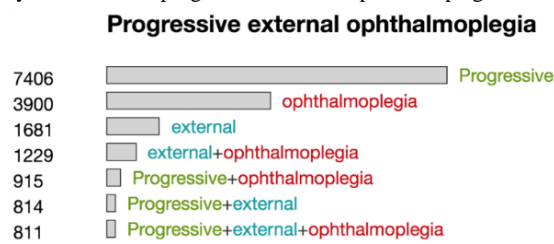
Figure 5. The result of adding a second phenotype via the ontology term suggester, which leads to more sensitivity-encoded query selectors.
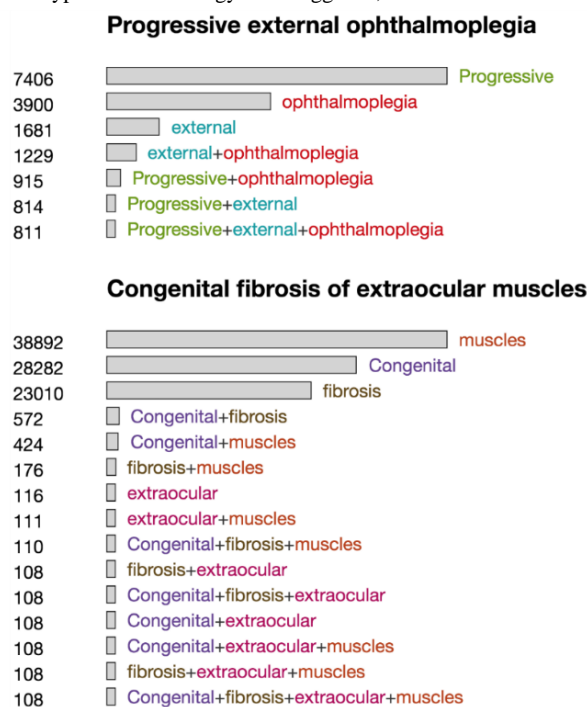
Interactive Triaging Support to Mitigate Information Overload

OVERT-MED provides multistage triaging support to mitigate the information overload problem. Multiple design strategies support the first stage of triaging—the "multiple document" stage. First, when a specific set of terms is chosen, the metadata from up to 250 documents are visualized. Each document is encoded using a small bar, and the presence of each term is encoded using a section of the bar. Figure 6 shows how 6 documents are represented in the case of 3 terms (progressive external opthalmoplegia). Within the visualization, each row represents a document, and each column represents one of the phenotype words. The words are color coded—in this case, green for progressive, teal for external, and red for opthalmoplegia. A white cell indicates no occurrence of the word. The visualization functions as a type of heatmap [57], where the color saturation encodes the frequency of a term within a document. We call this technique the *query result heatmap*. In Figure 6, a darker red means higher occurrence of the word opthalmoplegia. This type of encoding can aid in rapid visual scanning and identification of potentially relevant documents [43,58].

To further support the triaging activity, OVERT-MED allows users to interactively explore metadata associated with the matching documents. Figure 7 shows the state of the interface after a user has selected "progressive+opthalmoplegia." The first 250 documents (ranked by our indexing algorithm) are encoded in the Query Result Heatmap. Each row functions as an individual document heatmap, showing the occurrence of the 7 phenotype terms within the document. Because the user has selected "progressive" and "opthalmoplegia," all documents indicate occurrences of both terms. It is readily apparent that most of the documents also contain the term "external." Approximately 20 also contain "muscles," 4 contain "extraocular," 1 contains "fibrosis," and 1 "congenital."

OVERT-MED also provides a *Term Distribution Matrix* to help users quickly determine document relevance while browsing the Query Result Heatmap. Within the term distribution matrix, users can see the occurrence of terms in 4 places within the document metadata: (1) title, (2) journal name, (3) MeSH terms, and (4) abstract text. The document title, journal, year, and MeSH terms are also displayed. This representation helps users make decisions about relevance via quick visual scanning. For example, if a term appears only in the journal name it may not be very relevant, but if a term appears 5 times in the abstract text it is more likely to be relevant. Users can perceive this type of information quickly due to the categorical color encodings. Figure 8 shows the term distribution matrix for 2 different documents within the same result set. Through rapid visual scanning, even without reading the text, it is apparent that the terms are quite important in the document on the right.

To support rapid exploration—a fundamental goal of triaging—the keyboard arrow keys can be used to move quickly through the documents while the metadata is dynamically updated. If a relevant document is detected, users can hit the "enter" key or click the button to add the document to a pile for subsequent investigation (this stage is explained in greater detail in the following). This stage of triaging also allows for quick comparison of cooccurring phenotypes within documents. For example, Figure 9 shows the result of a user adding documents containing "congenital" and "fibrosis." It is immediately clear through quick visual scanning that not many documents contain both "congenital fibrosis" and "opthalmoplegia."

While browsing the query result heatmap, it may be difficult to remember which documents have been visited previously. This is especially true in the context of iterative triaging, where users may return to the heatmap after being away for some time. In OVERT-MED, when users pause on a document for 5 s or more, a small mark is placed beside the document to serve as a visual reminder (Figure 10). When revisiting the heatmap, users can quickly recognize which documents they have previously examined. We assume that 5 s is a reasonable threshold for determining when a user has examined the *term distribution matrix*.

Figure 6. The query result heatmap: 6 documents are represented by 6 rows, where each column represents a term (progressive external opthalmoplegia).
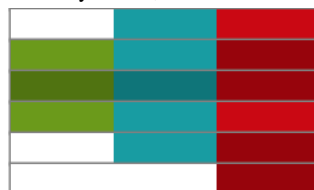
Figure 7. State of the interface after a user has selected "progressive+opthalmoplegia."
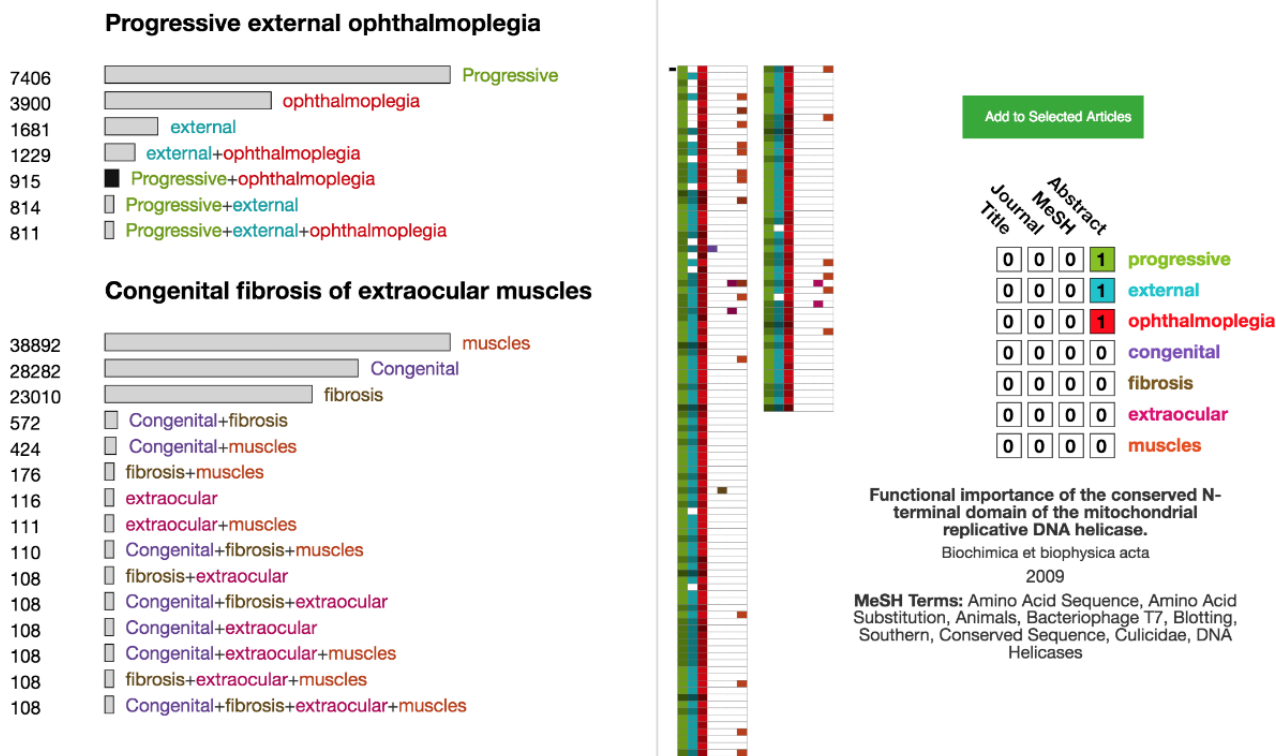

Figure 8. The term distribution matrix for 2 different documents within the same result set.
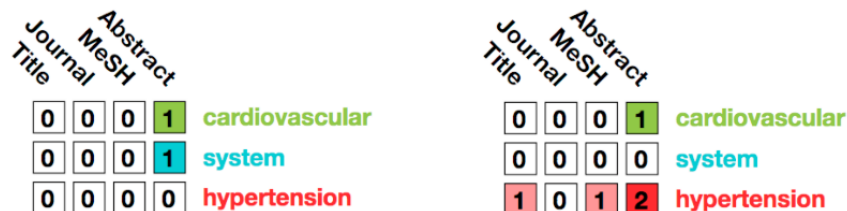

Figure 9. The result of a user adding documents containing "congenital" and "fibrosis" for comparison.
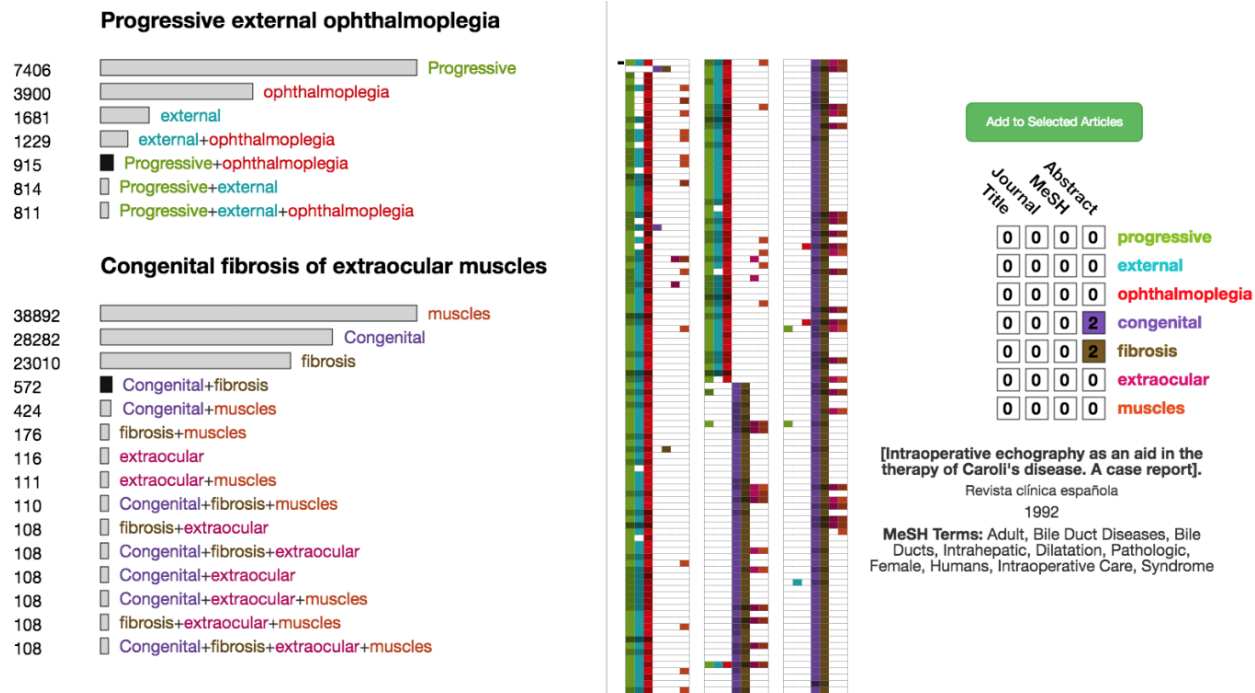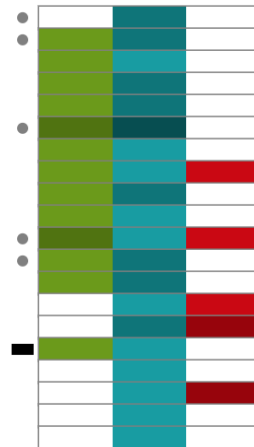
Figure 10. Closeup view of the query result heatmap.

The next stage in the triaging activity—the "individual document" stage—involves examining individual abstracts of previously chosen articles. At this stage, users are likely to have narrowed down the number of documents significantly. Documents are encoded via a *Selected Pile Heatmap* in the same manner as in the query result heatmap, and each can be selected to view its abstract. In this *term-encoded abstract*, matching terms are color coded to facilitate quick identification, especially within the abstract text. Figure 11 shows an example in which the user has selected 29 documents, which are encoded in the selected pile heatmap and the term-encoded abstract is displayed for the first document. Even before reading the text in detail, it is easy to see that "renin" and "hypertension" both appear frequently, indicating that they are important. Thus, users can scan the text quickly to get a sense of the appearance of the query terms, without having to necessarily read the text sequentially. An important aspect of this stage of triaging is the ability to quickly categorize documents. In OVERT-MED, users can quickly reject a paper by selecting the orange "x" button, or can quickly add a paper to the next stage by selecting the green button or pressing the "enter" key.

The final stage of triaging is the "further reading" stage, where a small set of documents are read in-depth to extract relevant information and satisfy the original information need. Although this stage could be supported in various ways, we support this stage in OVERT-MED by presenting a PubMed entry for a selected document in an embedded frame directly within the interface of OVERT-MED. This allows for quick inspection of any PubMed details that are important to the user, such as full-text links, citation details, and PubMed Commons links,

and also allows users to login to their *National Center for Biotechnology Information* (NCBI) account to save the article to a collection, compare with other saved articles, and so on. There is also a button to open the PubMed link in a new browser tab if a user needs more space. Figure 12 shows a full-screen capture of OVERT-MED in which a user has traversed all stages of a search and triaging activity.

As research shows that triaging activities are cyclical and iterative, we have designed OVERT-MED to be flexible in this regard. At any point during an activity, users may adjust their query or document selections, and each component of the interface will dynamically reflect any changes. For example, a user may reach the final stage of triaging and find a term within a document that seems relevant to the original information need. The user can return to the initial stage of entering the term and selecting phenotypes. In doing so, the rest of the interface remains stable and the user can proceed through any of the triaging stages. Figure 13 shows the interface after a user has examined a document in detail in the final stage, discovered a link between renin level (the original phenotype of interest) and arterial pressure, and has returned to the initial stage to find a phenotype related to arterial pressure. The user discovers a phenotype named "elevated mean arterial pressure" and selects it. At this stage, the user is not particularly interested in whether the arterial pressure is elevated, and simply wants to explore the relationship between renin level and arterial pressure. Due to our sensitivity encoding strategy, the user can select "arterial+pressure" to add documents with those 2 terms. From this point, the user can continue through the triaging stages or return to the initial stage again.

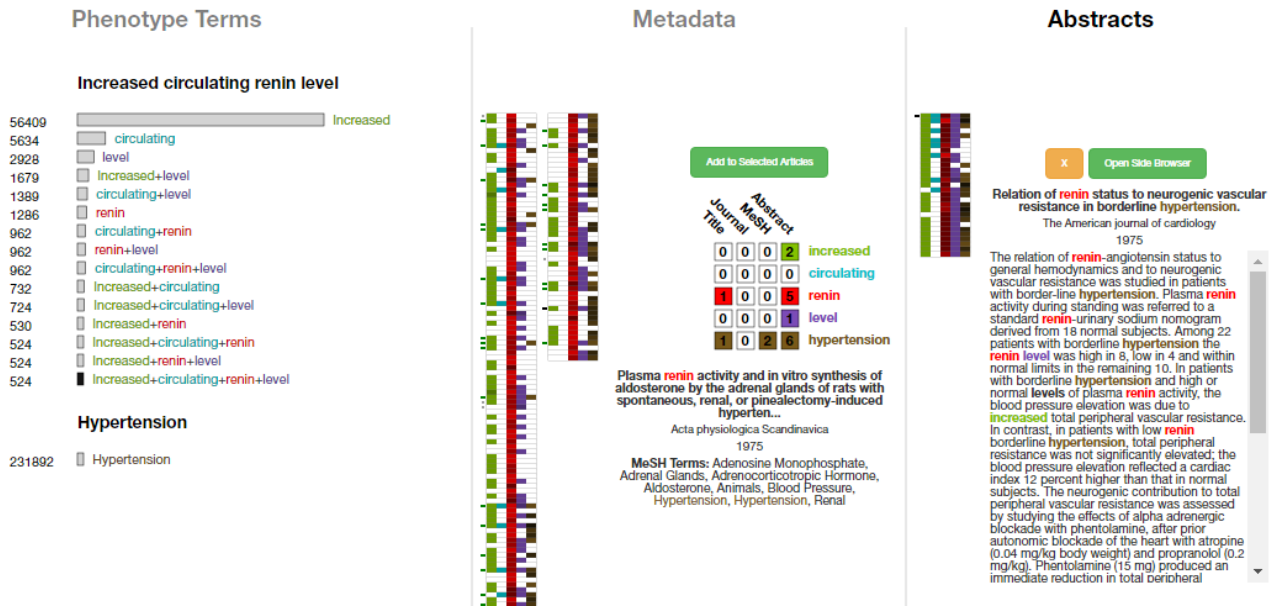**Figure 11.** Twenty-nine documents have been selected to examine in closer detail.

Figure 12. Full-screen capture showing all components of OVERT-MED where a user has traversed all stages of a search and triaging activity. OVERT-MED: Ontology-Driven Visual Search and Triage Interface for MEDLINE.
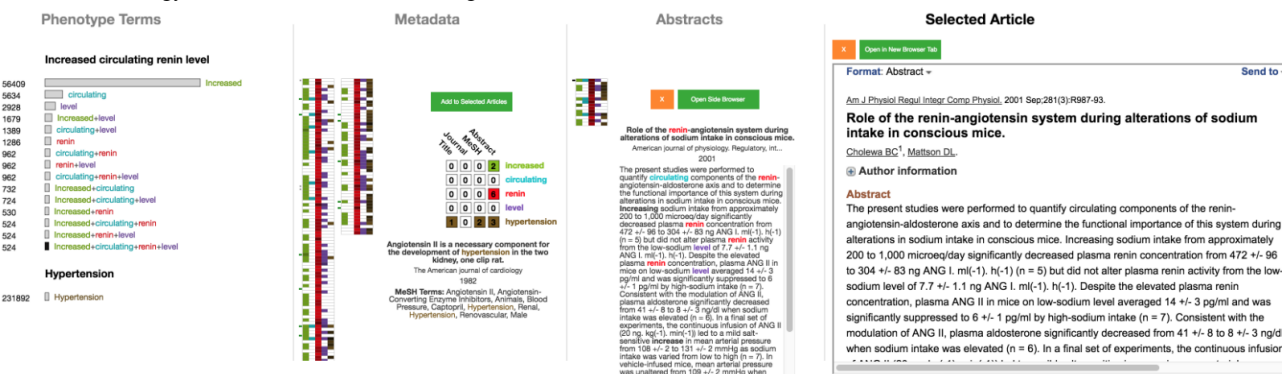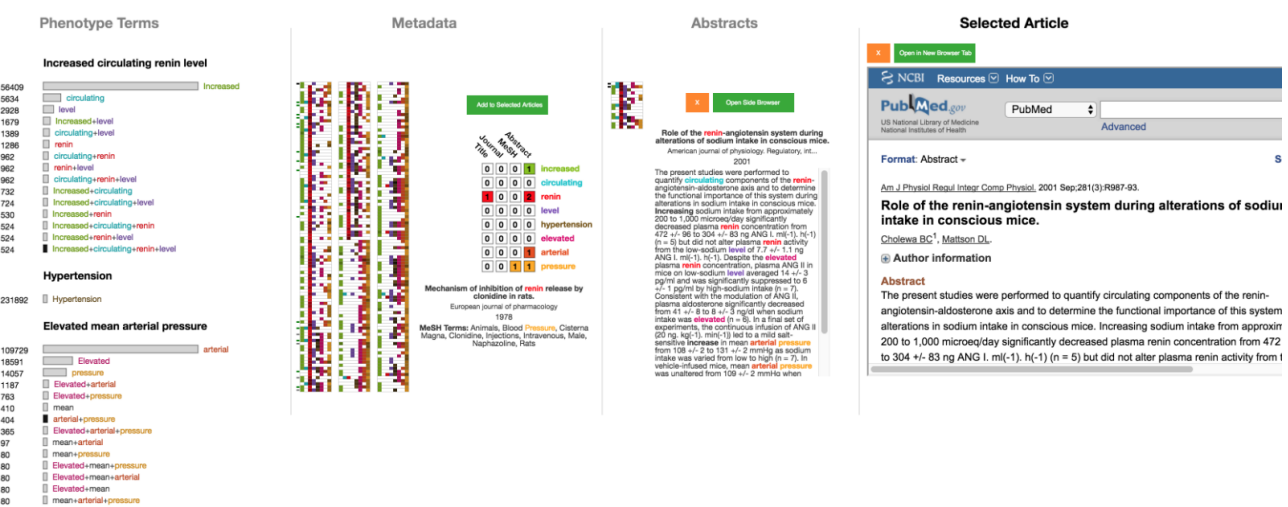

Figure 13. The interface after a user has examined a document in detail in the final stage, discovered a link, and has returned to the initial stage with a new information need.

Discussion

Overview

OVERT-MED was developed to address two major problems that are known to exist in complex, exploratory search activities:

(1) the difficulty in articulating information needs due to insufficient knowledge and domain-specific vocabulary, and (2) the difficultly in dealing with information overload due to the large number of results returned. To address the first difficulty, we proposed the idea of using a formal ontology to

XSL•FO

RenderX

help users build domain-specific terminology and knowledge for constructing search queries. To assist in this process, we indexed HPO and provided a search feature that provides robust results to terms that are entered. To address the problem of search criteria being too restrictive in open-ended contexts, we used a visual sensitivity encoding strategy to help users see possibilities with different combinations of terms.

There are 7 main steps that users take when performing search and triaging tasks with OVERT-MED—the first 2 within a vocabulary building phase and the next 5 within a triaging phase. The triaging phase can be broken down into the 3 key stages. Figure 14 provides an overview of this process and shows the techniques we use to help users at each step. To help users build vocabulary and generate queries, we use an *ontology term suggester* and *sensitivity encoded query selectors*. After selecting a query, users move to the triaging phase, where they traverse through 3 stages. During the first stage—the multi-document stage—users are presented with a *query result heatmap* that encodes the appearance and frequency of query terms within the document result set. A keyboard interaction technique enables rapid navigation through the documents. To facilitate assessment at this stage, a *term distribution matrix* provides more information about each document within the heatmap. Together these techniques allow for rapid scanning to assess relevance and select documents for the next stage. During the second triaging stage—the individual document stage—users are presented with a *Selected Pile Heatmap* that encodes only the selected documents from the previous stage. As users browse the heatmap, they can inspect a *term-encoded abstract* of each individual document. The term-encoding supports quick detection of the appearance of query terms within the document abstract. After assessing the relevance of individual documents, users select documents to move to the next stage. During the third triaging stage—the further reading stage—users focus on a single document by viewing details in depth. Here, the PubMed entry for a document can be retrieved directly within OVERT-MED or within a new browser tab. At any point in the overall activity, users can return to any step and continue from there, which supports the iterative and cyclical nature of search and triaging tasks.

Figure 14. Overall search and triage process supported by OVERT-MED. Users take 7 main steps—the first 2 within a vocabulary building phase, and the next 5 within a triaging phase. OVERT-MED: Ontology-Driven Visual Search and Triage Interface for MEDLINE.
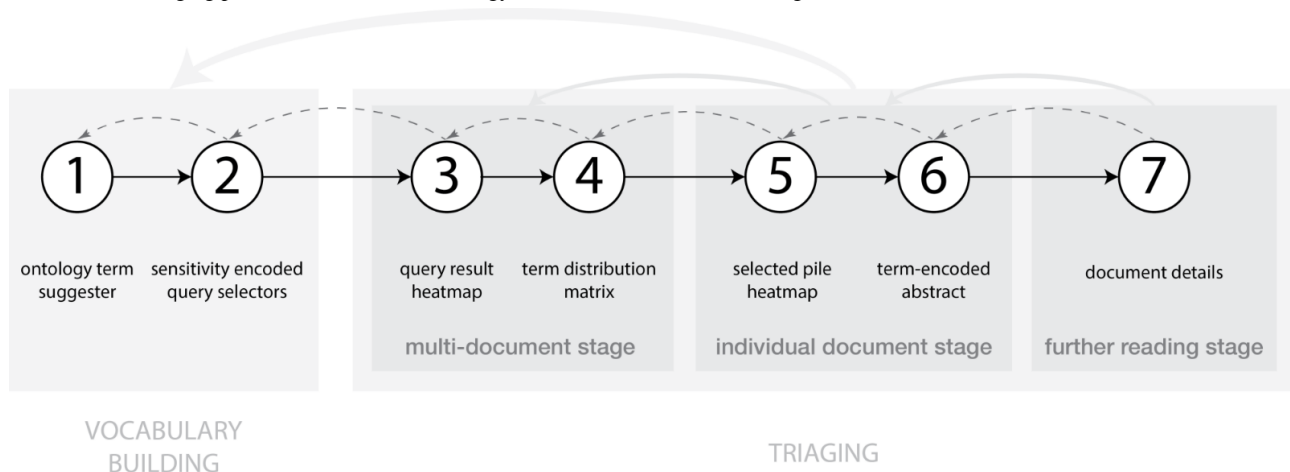
Validation

Ongoing formative evaluation suggests that the design features in OVERT-MED can mitigate the two problems mentioned above. We tested OVERT-MED with a small group of users who are not domain-experts, and our proposal to use a formal ontology to help users articulate their information needs does seem to be useful. As mentioned previously, different types of users are known to search the scientific literature, many of which are not domain experts. For example, pediatricians often try to identify abnormal phenotypes in patients before referring them to a clinical geneticist. However, because they are not domain experts, pediatricians may not have very extensive knowledge and vocabulary of phenotypes. Even if they search the literature to identify phenotype names (eg, via PubMed), they may still not find phenotypes that are related to one another. As another example, patients are known to search the literature to learn more about their own conditions. As they are not domain experts, patients could also benefit from having access to an ontology such as HPO to help them build domain-specific knowledge and vocabulary. Thus, testing with users who are not domain experts can give an indication of the usefulness of our design strategies.

In our testing, we noticed that although an ontology can help users develop more appropriate vocabulary, users do not necessarily develop a good understanding of the ontology itself. As a robust mental model of the ontology may lead to even better search performance (eg, by knowing which entities are highly connected to others, knowing relationships among entities at multiple levels of abstraction, and so on), we have decided to pursue a solution to this as future work (see Future Work section). In addition, our multistage triaging shows promise in mitigating the information overload problem. Users were able to go back and forth through the triaging stages to satisfy information needs without being overwhelmed by long lists of documents.

Limitations

There is 1 current limitation of OVERT-MED that should be noted: the MEDLINE data are limited to metadata and abstract text only, and do not include full texts. This is simply because

the NLM does not release full-texts due to copyright issues. There is little we can do to address this issue. Empirical evidence, however, does suggest that the document title and abstract are among the most important features of a document in determining its relevance [37], so perhaps it is not a critical limitation.

Future Work

We envision at least three lines of valuable future research:

First, developing interactive visualization techniques to support ontology sensemaking. The intention behind the current version of OVERT-MED is to help address the common problem of lack of adequate vocabulary. Although OVERT-MED appears to support users in improving their search terms and potentially developing some domain knowledge, it does not necessarily support users in making sense of the ontology itself—that is, understanding its size, organization, types of relationships, significant and insignificant entities, and so on. Interactive visualizations of ontologies may enhance search and triaging activities. Second, testing OVERT-MED with different ontologies in different contexts. This will help assess the transferability of the design features of OVERT-MED. Third, conducting formal testing of OVERT-MED. Although our informal testing has been useful, more formal testing will provide validation of the design strategies.

Conclusions

We have developed a Web-based interactive visualization tool, OVERT-MED, to address two common problems in exploratory search—namely, the lack of adequate vocabulary to construct useful queries and the difficulty of dealing with very large result sets. The novelty of our approach is in the combination of (1) using an ontology to help build domain-specific knowledge and vocabulary when users need to be kept "in the loop" and (2) providing multistage triaging support using interactive visualizations to help mitigate the information overload problem. We anticipate these ideas can be applied successfully in other contexts where either of these issues exists.

Conflicts of Interest

None declared.

References

1. Krupski TL, Dahm P, Fesperman SF, Schardt CM. How to perform a literature search. J Urol 2008 Apr;179(4):1264-1270. [doi: 10.1016/j.juro.2007.11.087] [Medline: 18280516]

2. Islamaj DR, Murray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. Database (Oxford) 2009 Nov 27;2009:bap018 [FREE Full text] [doi: 10.1093/database/bap018] [Medline: 20157491]

3. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013 Jun 26;15(6):e122 [FREE Full text] [doi: 10.2196/jmir.2436] [Medline: 23803299]

4. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Friedman CP, Tidmarsh P, et al. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. J Am Med Inform Assoc 2002;9(3):283-293 [FREE Full text] [Medline: 11971889]

5. Parsons P, Sedig K, Mercer R, Khordad M, Knoll J, Rogan P. Visual analytics for supporting evidence-based interpretation of molecular cytogenomic findings. In: Proceedings of the 2015 Workshop on Visual Analytics in Healthcare. New York, New York, USA: ACM Press; 2015 Oct 25 Presented at: IEEE VIS; 2015; Chicago. [doi: 10.1145/2836034.2836036]

6. Palotti J, Hanbury A, Müller H, Kahn C. How users search and what they search for in the medical domain. Inf Retrieval J 2015 Oct 24;19(1-2):189-224. [doi: 10.1007/s10791-015-9269-8]

7. Marchionini G. Exploratory search: from finding to understanding. In: Communications of the ACM - Supporting exploratory search. New York, NY, USA: ACM; Apr 01, 2006:41-46.

8. Hersh WR, Hickam DH. How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. J Am Med Assoc 1998 Oct 21;280(15):1347-1352. [Medline: 9794316]

9. Cui L, Carter R, Zhang G. Evaluation of a novel conjunctive exploratory navigation interface for consumer health information: a crowdsourced comparative study. J Med Internet Res 2014;16(2):e45 [FREE Full text] [doi: 10.2196/jmir.3111] [Medline: 24513593]

10. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]

11. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: obstacles and potential solutions. J Am Med Inform Assoc 2005;12(2):217-224 [FREE Full text] [doi: 10.1197/jamia.M1608] [Medline: 15561792]

XSL•FO

RenderX

12. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007 Jun;24(2):78-94 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00713.x] [Medline: 17584211]

13. Dietze H, Alexopoulou D, Alvers MR, Barrio-Alvers L, Andreopoulos B, Doms A, et al. GoPubMed: Exploring PubMed with Ontological Background Knowledge. In: Bioinformatics for Systems Biology. Bioinforma Syst Biol Totowa, NJ: Humana Press; 2009:385-399.

14. NCBI.NLM. Home-PubMed URL: https://www.ncbi.nlm.nih.gov/pubmed [accessed 2016-10-14] [WebCite Cache ID 6lGAJxbQD]

15. Furnas GW, Landauer TK, Gomez LM, Dumais ST. The vocabulary problem in human-system communication. Commun ACM 1987:964-971.

16. Belkin NJ. Helping people find what they don't know. Commun ACM 2000;43(8):58-61. [doi: 10.1145/345124.345143]

17. Patrick TB, Monga HK, Sievert ME, Houston HJ, Longo DR. Evaluation of controlled vocabulary resources for development of a consumer entry vocabulary for diabetes. J Med Internet Res 2001;3(3):E24 [FREE Full text] [doi: 10.2196/jmir.3.3.e24] [Medline: 11720966]

18. Plovnick RM, Zeng QT. Reformulation of consumer health queries with professional terminology: a pilot study. J Med Internet Res 2004 Sep 03;6(3):e27 [FREE Full text] [doi: 10.2196/jmir.6.3.e27] [Medline: 15471753]

19. Sievert M, Patrick T, Reid J. Need a bloody nose be a nosebleed? or, lexical variants cause surprising results. Bull Med Libr Assoc 2001 Jan;89(1):68-71 [FREE Full text] [Medline: 11209803]

20. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]

21. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. J Am Med Assoc 1994 Apr 13;271(14):1103-1108. [Medline: 8151853]

22. Malhotra A, Gündel M, Rajput AM, Mevissen H, Saiz A, Pastor X, et al. Knowledge retrieval from PubMed abstracts and electronic medical records with the Multiple Sclerosis Ontology. PLoS One 2015;10(2):e0116718 [FREE Full text] [doi: 10.1371/journal.pone.0116718] [Medline: 25665127]

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

24. Hoeber O, Khazaei T. Evaluating citation visualization and exploration methods for supporting academic search tasks. Online Information Review 2015 Apr 13;39(2):229-254. [doi: 10.1108/OIR-10-2014-0259]

25. Hoeber O. Visual Search Analytics: Combining Machine Learning and Interactive Visualization to Support Human-Centred Search. 2014 Presented at: Pro-ceedings of the MindTheGap'14 Workshop; March 4 2014; Berlin, Germany p. 37-43.

26. Hearst M, Elliott A, English J, Sinha R, Swearingen K, Yee K. Finding the flow in web site search. Commun ACM 2002;45(9):42-49. [doi: 10.1145/567498.567525]

27. Yee K, Swearingen K, Li K, Hearst M. Faceted metadata for image search and browsing. New York, NY, USA: ACM Press; 2003 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 5-10 2003; Ft Lauderdale, FL, USA p. 401-408. [doi: 10.1145/642611.642681]

28. Dork M, Williamson C, Carpendale S. Towards Visual Web Search?: Interactive Query Formulation and Search Result Visualization. 2009 Presented at: WSSP 2009: WWW Workshop on Web Search Result Summarization and Presentation; April 20 2009; Madrid, Spain p. 5.

29. Diriye A, Tombros A, Blandford A. A Little Interaction Can Go a Long Way: Enriching the Query Formulation Process. In: Lect Notes Comput Sci. 2012 Presented at: European Conference on Information Retrieval; April 1-5 2012; Barcelona, Spain p. 531-534. [doi: 10.1007/978-3-642-28997-2_57]

30. Joho H, Coverson C, Sanderson M, Beaulieu M. Hierarchical presentation of expansion terms. New York, NY, USA: ACM; 2002 Presented at: ACM symposium on Applied computing; March 11-14 2002; Madrid, Spain p. 645-649. [doi: 10.1145/508791.508916]

31. Gruber T. The role of common ontology in achieving sharable, reusable knowledge bases. 1991 Presented at: Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning; 1991; Cambridge, MA, USA.

32. Chandrasekaran B, Josephson JR, Benjamins VR. What are ontologies, and why do we need them? IEEE Intell Syst 1975 Oct 01;14(1):20-26.

33. Guarino N, Oberle D, Staab S. What Is an Ontology? In: Handbook on Ontologies. Handb Ontol Berlin, Heidelberg: Springer Berlin Heidelberg; 2009:1-17.

34. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 2008 Nov;83(5):610-615 [FREE Full text] [doi: 10.1016/j.ajhg.2008.09.017] [Medline: 18950739]

35. Mavri A, Loizides F, Photiades T, Zaphiris P. We Have the Content…Now What?: The role of Structure and Interactivity in Academic Document Triage Interfaces. Inf Des J 2013;20(3):247-265. [doi: 10.1075/idj.20.3.05mav]

36. Loizides F, Buchanan G. Towards a Framework for Human (Manual) Information Retrieval. In: Multidisciplinary Information Retrieval. 2013 Presented at: Information Retrieval Facility Conference; October 7-9; Limassol, Cyprus p. 87-98. [doi: 10.1007/978-3-642-41057-4_10]

37. Loizides F, Buchanan G. An empirical study of user navigation during document triage. 2009 Presented at: 13th European Conference; September 27 - October 2; Corfu, Greece p. 138-149.

38. Khazaei T, Hoeber O. Supporting academic search tasks through citation visualization and exploration. Int J Digit Libr 2016 Apr 26:1-14. [doi: 10.1007/s00799-016-0170-x]

39. Spink A, Wolfram D, Jansen M, Saracevic T. Searching the web: the public and their queries. J Am Soc Inf Sci 2001;52(3):226-234. [doi: 10.1002/1097-4571(2000)9999:9999<::AID-ASI1591>3.0.CO;2-R]

40. Silverstein C, Marais H, Henzinger M, Moricz M. Analysis of a very large web search engine query log. SIGIR Forum 1999 Sep 01;33(1):6-12. [doi: 10.1145/331403.331405]

41. Scaife M, Rogers Y. External cognition: how do graphical representations work? Int J Hum Comput Stud 1996 Aug;45(2):185-213. [doi: 10.1006/ijhc.1996.0048]

42. Larkin J, Simon H. Why a diagram is (sometimes) worth ten thousand words. Cogn Sci 1987;11(1):65-100. [doi: 10.1111/j.1551-6708.1987.tb00863.x]

43. Hearst M. TileBars: Visualization of Term Distribution Information in Full Text Information Access. 1995 Presented at: Proc SIGCHI Conf Hum factors Comput Syst; 1995; Denver, CO, USA. [doi: 10.1145/223904.223912]

44. Gorg C, Liu Z, Stasko J. Reflections on the evolution of the Jigsaw visual analytics system. Inf Vis 2013 Jul 23;13(4):336-345. [doi: 10.1177/1473871613495674]

45. Thomas W, Alexopoulou D, Dietze H, Schroeder M. Searching biomedical literature with anatomy ontologies. Anatomy Ontologies for Bioinformatics 2009;6:177-194. [doi: 10.1007/978-1-84628-885-2_9]

46. Wei C, Kao H, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. Nucleic Acids Res 2013 Jul;41(Web Server issue):W518-W522 [FREE Full text] [doi: 10.1093/nar/gkt441] [Medline: 23703206]

47. Muin M, Fontelo P, Liu F, Ackerman M. SLIM: an alternative Web interface for MEDLINE/PubMed searches - a preliminary study. BMC Med Inform Decis Mak 2005 Dec 01;5:37 [FREE Full text] [doi: 10.1186/1472-6947-5-37] [Medline: 16321145]

48. Eaton AD. HubMed: a web-based biomedical literature search interface. Nucleic Acids Res 2006 Jul 01;34(Web Server issue):W745-W747 [FREE Full text] [doi: 10.1093/nar/gkl037] [Medline: 16845111]

49. Wang J, Cetindil I, Ji S, Li C, Xie X, Li G, et al. Interactive and fuzzy search: a dynamic way to explore MEDLINE. Bioinformatics 2010 Sep 15;26(18):2321-2327 [FREE Full text] [doi: 10.1093/bioinformatics/btq414] [Medline: 20624778]

50. Yu H, Kim T, Oh J, Ko I, Kim S. RefMed: relevance feedback retrieval system fo PubMed. New York, NY, USA: ACM; 2009 Presented at: 18th ACM conference on Information and knowledge management; November 2-6 2009; Hong Kong, China p. 2099-2100. [doi: 10.1145/1645953.1646322]

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

52. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Inf Process Manag 1988 Jan;24(5):513-523. [doi: 10.1016/0306-4573(88)90021-0]

53. Salton G, Wong A, Yang CS. A vector space model for automatic indexing. Commun ACM 1975;18(11):613-620 [FREE Full text] [doi: 10.1145/361219.361220]

54. Bostock M, Ogievetsky V, Heer J. D³: data-driven documents. IEEE Trans Vis Comput Graph 2011 Dec;17(12):2301-2309. [doi: 10.1109/TVCG.2011.185] [Medline: 22034350]

55. Spence R, Tweedie L. The Attribute Explorer: information synthesis via exploration. Interact Comput 1998 Dec;11(2):137-146. [doi: 10.1016/S0953-5438(98)00022-8]

56. Spence R. Sensitivity encoding to support information space navigation: a design guideline. Inf Vis 2002;1(2):120-129. [doi: 10.1057/palgrave.ivs.9500019]

57. Wilkinson L, Friendly M. The History of the Cluster Heat Map. Am Stat 2009;63(2):179-184. [doi: 10.1198/tas.2009.0033]

58. Hoeber O, Yang X. HotMap: supporting visual exploration of web search results. J Am Soc Inf Sci 2009 Jan;60(1):90-110. [doi: 10.1002/asi.20957]

Abbreviations

HPO: Human Phenotype Ontology
MEDLINE: Medical Literature Analysis and Retrieval System Online
MeSH: Medical Subject Header
NLM: National Library of Medicine
OVERT-MED: Ontology-Driven Visual Search and Triage Interface for MEDLINE

XSL•FO

RenderX

XSL•FO
RenderX

Original Paper

Does Telehealth Monitoring Identify Exacerbations of Chronic Obstructive Pulmonary Disease and Reduce Hospitalisations? An Analysis of System Data

Melissa Kargiannakis[1*], MHIS; Deborah A Fitzsimmons[1,2,3*], PhD; Claire L Bentley[3], BSc; Gail A Mountain[3,4], PhD

[1]Faculty of Health Sciences, Western University, London, ON, Canada

[2]School of Nursing and Allied Health, Liverpool John Moores University, Liverpool, United Kingdom

[3]School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom

[4]Faculty of Health Studies, University of Bradford, Bradford, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Deborah A Fitzsimmons, PhD
School of Nursing and Allied Health
Liverpool John Moores University
Tithebarn Building
79 Tithebarn Street
Liverpool, L2 2ER
United Kingdom
Phone: 44 151 231 5872
Fax: 44 151 231 5872
Email: D.A.Fitzsimmons@ljmu.ac.uk

Abstract

Background: The increasing prevalence and associated cost of treating chronic obstructive pulmonary disease (COPD) is unsustainable. Health care organizations are focusing on ways to support self-management and prevent hospital admissions, including telehealth-monitoring services capturing physiological and health status data. This paper reports on data captured during a pilot randomized controlled trial of telehealth-supported care within a community-based service for patients discharged from hospital following an exacerbation of their COPD.

Objective: The aim was to undertake the first analysis of system data to determine whether telehealth monitoring can identify an exacerbation of COPD, providing clinicians with an opportunity to intervene with timely treatment and prevent hospital readmission.

Methods: A total of 23 participants received a telehealth-supported intervention. This paper reports on the analysis of data from a telehealth monitoring system that captured data from two sources: (1) data uploaded both manually and using Bluetooth peripheral devices by the 23 participants and (2) clinical records entered as nursing notes by the clinicians. Rules embedded in the telehealth monitoring system triggered system alerts to be reviewed by remote clinicians who determined whether clinical intervention was required. We also analyzed data on the frequency and length (bed days) of hospital admissions, frequency of hospital Accident and Emergency visits that did not lead to hospital admission, and frequency and type of community health care service contacts—other than the COPD discharge service—for all participants for the duration of the intervention and 6 months postintervention.

Results: Patients generated 512 alerts, 451 of which occurred during the first 42 days that all participants used the equipment. Patients generated fewer alerts over time with typically seven alerts per day within the first 10 days and four alerts per day thereafter. They also had three times more days without alerts than with alerts. Alerts were most commonly triggered by reports of being more tired, having difficulty with self-care, and blood pressure being out of range. During the 8-week intervention, and for 6-month follow-up, eight of the 23 patients were hospitalized. Hospital readmission rates (2/23, 9%) in the first 28 days of service were lower than the 20% UK norm.

Conclusions: It seems that the clinical team can identify exacerbations based on both an increase in alerts and the types of system-generated alerts as evidenced by their efforts to provided treatment interventions. There was some indication that telehealth monitoring potentially delayed hospitalizations until after patients had been discharged from the service. We suggest that

XSL•FO
RenderX

telehealth-supported care can fulfill an important role in enabling patients with COPD to better manage their condition and remain out of hospital, but adequate resourcing and timely response to alerts is a critical factor in supporting patients to remain at home.

Introduction

Chronic obstructive pulmonary disease (COPD) is the fifth-highest cause of mortality and second-highest cause of emergency admissions to hospital in the United Kingdom [1]. It costs the National Health Service (NHS) more than £800 million per annum [2]. For hospital patients, COPD accounts for £587 million of the total £1.08 billion spent on admissions for lung disease by the NHS [3,4]. Patients discharged from hospital following COPD exacerbations have a high readmission rate [5]. The forecasted increase in COPD prevalence makes current models of care delivery unsustainable. There is a global need for care delivery models that encourage prevention, self-management [6], and home-based management approaches designed to avoid hospital admission and reduce health care costs [7].

Definitions

Chronic obstructive pulmonary disease is characterized by progressive worsening of lung capacity. Patients with advanced COPD typically experience impaired physical, emotional, and social functioning, which results in poor quality of life [8]. The NHS describes COPD as progressive airflow obstruction that is not fully reversible and does not change markedly over several months [9].

Exacerbations of COPD are described as "a sustained worsening of the patient's symptoms from their usual stable state, which is beyond normal day-to-day variations, and is acute in onset" [8]. Key symptoms indicative of an exacerbation include increased dyspnea; sputum purulence; sputum volume; cough, wheeze, or fatigue; chest tightness; reduced exercise tolerance; fluid retention; or acute confusion [9-14]. Segrelles et al [15] identify one addendum "...that leads to a change in medication" and note that patients with more acute exacerbations of COPD (AECOPDs) have a worse prognosis. Toy et al [16] have identified that patients with COPD are likely to experience exacerbations that are unreported. The severity of AECOPD is closely related to health care delivery costs [17]. Fernández-Granero et al [11] were able to detect AECOPDs an average of 4.8 days before onset with 80.5% accuracy using a questionnaire analyzed by a probabilistic neural network, but this approach is not part of the standard care pathway and adds an incremental step. If telehealth monitoring embedded within a clinical support service is able to provide early and accurate detection of AECOPDs, as suggested by Fernández-Granero et al's results [11], it could offer an opportunity for early intervention to alleviate symptoms and reduce care costs.

Local Context

The region chosen for this pilot study has a high prevalence of COPD linked to the predominant mining industry [16]. The Index of Multiple Deprivation rates this region as one of the most deprived due to poor diet and other adverse lifestyle factors, including a relatively high level of smoking [18]. Between April 2006 and March 2007, COPD-related admissions billed to the local health service cost £2.2 million [14].

Telehealth Intervention

The telehealth-monitoring intervention was introduced with the goals of decreasing hospitalizations, improving the quality of life for patients, and reducing resource use, while significantly increasing capacity of the service. It was believed that the data collected through the telehealth system would enable clinicians to provide a more patient-centered service by identifying whether patients required additional supportive home visits to address any fluctuations in their condition. Using these data, it was also hoped that unnecessary visits could be eliminated, thereby freeing resources that could be used to support additional patients.

The selected telehealth system (Doc@Home) provided both monitoring and self-management support functionality. Using a small hand-held device, patients were required to answer tailored questions about their health status by reading questions on the screen of the device and pressing the appropriate response button. Patients also used a blood pressure monitor and oximeter peripherals to measure their blood oxygen levels each day. The peripherals were connected by Bluetooth to the hand-held device, and all readings were transmitted to a secure Web-based server by telephone line, ready for access by the clinicians. Patients were able to observe their readings each day; this was a core educational element of the service. If reported signs and symptoms fell outside clinician-generated thresholds, or if the patient failed to undertake the monitoring activity, the system generated a color-coded alert visible to the remote clinician when reviewing the data submitted by patients for that day. Further details of the intervention are available online [19].

Installation of the telehealth equipment involved the installer instructing patients (and also their carer, if appropriate) on how to use the equipment, including the peripherals. The installer also informed them of when they should take readings as well as how and when they might request help if required. The installer provided the patient with a customized instruction manual, which included service information and key contact details should they require assistance.

Outcome Measures

If telehealth monitoring can identify AECOPDs, it could provide an opportunity for clinicians to intervene and prevent more invasive and more costly interactions, such as hospital admissions. Consequently, the primary outcome measure of interest was the proportion of participants readmitted to hospital with COPD during the 8-week intervention and 6-month follow-up, determined using patient-level data on hospital readmissions obtained from the Secondary Uses Service (SUS), the single, comprehensive repository for health care data in England [19]. The secondary outcome measure of interest was the proportion of patients requiring unscheduled health care support for the 8-week intervention period and 6-month follow-up, determined through analysis of SUS data [19].

The aim of this paper is to show whether telehealth monitoring can identify AECOPDs for patients with COPD that require follow-up and, potentially, a clinical intervention.

Methods

This paper reports on the analysis of data captured by the telehealth monitoring system during a pilot study [20], completed in preparation for a randomized controlled trial (RCT) of telehealth monitoring for patients with early-stage COPD [19]. The pilot study was conducted over a period of 14 months. Full details of the intervention and the study are available in the published initial findings [20].

Recruitment

Full details of the eligibility criteria, recruitment, and consent procedures used for the pilot study are detailed in the initial published findings of the study [19]. A total of 23 participants were recruited to the experimental group receiving the telehealth-supported intervention.

Data Collection

The telehealth monitoring system captured data from two sources: (1) data uploaded by the 23 participants providing telehealth-monitoring data as part of an 8-week early supported discharge nursing intervention and (2) clinical records entered as nursing notes by the clinicians.

We also requested SUS data from the local Primary Care Trust Commissioner on the frequency and length (bed days) of hospital admissions, frequency of hospital Accident and Emergency (A&E) visits that did not lead to hospital admission, and frequency and type of community health care service contacts other than the COPD discharge service for all participants (who completed the 8-week intervention) for the duration of the intervention and 6-month follow-up.

Statistical Analysis

The patient input data used by the clinicians to determine whether patients may be experiencing an AECOPD were collected as nominal dichotomous data with a "yes/no" alert-triggering system: the patient's inputs did or did not trigger an alert. We calculated the frequency with which each patient generated an alert for each prompt each day, if any.

Hospital admission data were tracked for all 23 participants for 6 months following the end of the intervention. For each patient, we identified the system-generated alerts, any resultant clinical activity, and the need for additional supportive services including visits to a hospital A&E or hospital admissions.

Ethics and Governance

The pilot study received ethical approval from the South Yorkshire Research Ethics Committee (reference 10/H130/48) and research and development approval from the local NHS hospital Trust in the United Kingdom. Approval was subsequently obtained from Western University in London, ON, Canada, for the analysis of the quantitative telehealth system data.

Results

This paper reports on the 23 participants undertaking an 8-week early supported discharge nursing intervention. If the patient was still considered to be too unwell to be discharged from the service at the end of the 8 weeks, the clinicians allowed the participant to retain the equipment until they were admitted into a community nursing program to provide additional support to patients with advanced COPD. Removal of the telehealth monitoring system and discharge of the patients from the service was subject to the availability of the patient, a member of the clinical team, and a technician. Consequently, the number of potential data entry days varied for each patient. We have reported on the 42 days after system installation that all patients provided data (referred to as period 1) and the additional data captured by the telehealth monitoring system after day 42 until removal of the equipment (referred to as period 2).

Use of the Telehealth Monitoring System

The entire cohort provided data for a minimum of 42 days following installation of the equipment, with a mean duration of data provision of 51.1 (SD 7.4) days. As shown in Table 1, two patients were subject to delayed discharge from the service, undertaking telehealth monitoring for an additional 2 weeks. Data were provided by patients for 92.43% (1086/1175) of the time the system was installed, and any explanations for missing data are included in Table 1.

XSL·FO
RenderX

Table 1. Total alerts generated by patients with identification of clinical interventions.

Patient ID[a]	Total alerts generated (n=512), n	Days on the telehealth system (n=1175), n	Missed data input (n=89)		Days ≥1 alert (n=243), n (%)	Days without an alert (n=932), n (%)
			Days, n	Reason		
P4[b,c]	88	48	10	In hospital	25 (52.1)	23 (47.9)
P23[b,c]	84	51	0		32 (62.7)	19 (37.3)
P3[b,c]	72	56	10	In hospital	26 (46.4)	30 (53.6)
P13[b,c]	36	69	15	10 in hospital	13 (18.8)	56 (81.2)
P16	32	47	0		22 (46.8)	25 (53.2)
P19[b,c]	29	54	3	Faulty system	13 (24.1)	41 (75.9)
P2[b]	23	71	7	5 on holiday	15 (21.1)	56 (78.9)
P14[b]	21	59	9		13 (22.0)	46 (78.0)
P10[b]	16	50	0		12 (24.0)	38 (76.0)
P11[b]	13	42	5	4 in hospital	6 (14.3)	36 (85.7)
P6[b]	13	47	0		13 (27.7)	34 (72.3)
P15[b,c]	12	47	0		7 (14.9)	40 (85.1)
P5	12	46	0		6 (13.0)	40 (87.0)
P17	11	52	0		6 (11.5)	46 (88.5)
P9[b,c]	11	49	0		7 (14.3)	42 (85.7)
P21[b,c]	10	49	5		5 (10.2)	44 (89.8)
P22[b]	9	52	1	Forgot	9 (17.3)	43 (82.7)
P7	6	42	0		5 (11.9)	37 (88.1)
P20	6	45	4		2 (4.4)	43 (95.6)
P1	5	51	12		4 (7.8)	47 (92.2)
P12	2	43	0		1 (2.3)	42 (97.7)
P18	1	50	8	Holiday	1 (2.0)	49 (98.0)
P8	0	55	0		0 (0.0)	55 (100.0)

[a] Patient numbers are anonymized.

[b] Patients who received treatment intervention.

[c] Patient who were hospitalized.

Volume and Frequency of Alerts

The 12 telehealth monitoring system prompts and four Bluetooth readings provided by participants—with responses that triggered a system alert—were mapped to the NHS' AECOPD key symptoms and are shown in Table 2.

XSL·FO
RenderX

Table 2. NHS AECOPD symptoms paired with Doc@Home prompts.

Doc@Home prompts	Response-triggering alert	Related NHS AECOPD key symptoms
1. Have you been feeling more tired than usual over the last 24 hours?	More than usual	Increased fatigue / reduced exercise tolerance
2. How is your breathlessness today?	More than usual	Increased dyspnea, wheeze, and/or cough
2a. Is that on normal activities/ exertion or rest?	Doing less than usual	Increased dyspnea, wheeze, and/or cough
3. Has your sleep been affected by coughing or shortness of breath?	More than usual	Increased dyspnea, wheeze, and/or cough
4. Have you produced sputum in the last 24 hours?	More than usual	Increased sputum purulence and increased sputum volume
4a. What color is your sputum?	Any if quantity is more than usual	Increased sputum purulence and increased sputum volume
5. Are your ankles or feet swollen this morning?	More than usual	Fluid retention
6. How able are you to do your self-care activities (dressing/ bathing)?	Not at all	Increased fatigue / reduced exercise tolerance
7. Have you had to use your relieving medication in the last 24 hours?	More than usual	None
8. How anxious have you been over the last 24 hours?	Much more than usual	None
8a. How have you coped with your anxiety?	Any if anxiety level much more than usual	None
9. How has your general health been in the last 24 hours? (1=normal for me; 10=extremely poor)	No response captured in the system	None
10. Have you had any problems in walking about in the past week? (yes/some/no)	No response set to trigger alert	Increased fatigue / reduced exercise tolerance
11. Have you had problems doing your usual activities in the past week?	No response set to trigger alert	Increased fatigue / reduced exercise tolerance
12. Have you had to contact the following within the last 24 hours? (no-one, COPD service, GP, hospital)	No response set to trigger alert	None
13. Blood pressure-Bluetooth systolic blood pressure (mmHg)	Personally adjusted parameter	None
14. Blood pressure-Bluetooth diastolic blood pressure (mmHg)	Personally adjusted parameter	None
15. SpO$_2$ reading-Bluetooth O2 (%)	Personally adjusted parameter	None
16. SpO$_2$ reading-Bluetooth pulse rate (bpm)	Personally adjusted parameter	None

As shown in Figure 1, most patient alerts were triggered during the first 10 days following installation of the telehealth-monitoring equipment. The mean number of patients triggering alerts each day declined over time, with a slope of equation $y=-0.1011x + 7.0354$. A mean of 30% (6.9/23) of the patient cohort generated alerts each day during the first 10 days following installation of the equipment. This reduced to a mean of 20% (4.5/23) from days 11 to 42, and decreased further to a mean of 14% (1.0/7.2) of the patient cohort triggering alerts each day for days 43 to 71.

As depicted in Table 1, patients experienced, on average, four times as many days without generating alerts (932/1175, 79.32%) than with generating alerts (243/1175, 20.68%).

Figure 2 suggests that there is a seasonal effect to the system-generated alerts. When controlling for the differing number of patients receiving the service each month, Figure 3 shows that there was a significant decrease in the number of alerts triggered per patient service day during the spring months of March to May. The number of system-generated alerts per patient service day in December were much lower than for the months of November and January.

Figure 1. Number of patients entering data and number of patients triggering alerts by day of using the equipment (N=23).
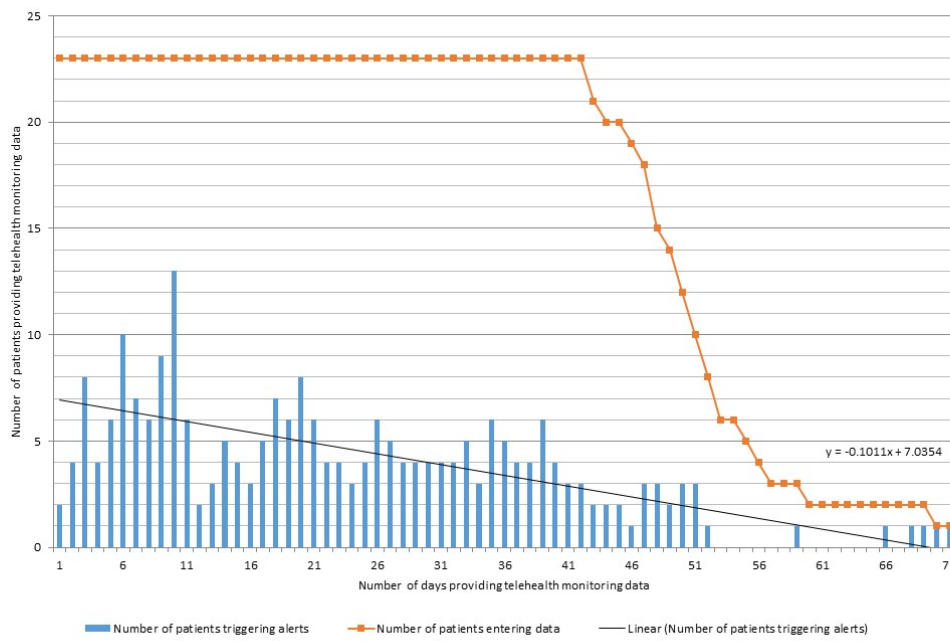
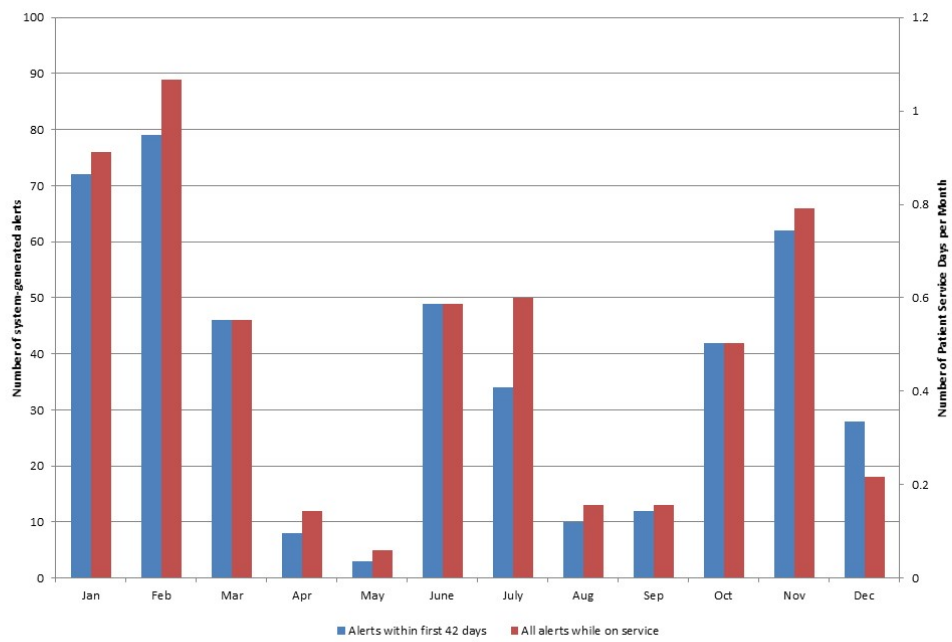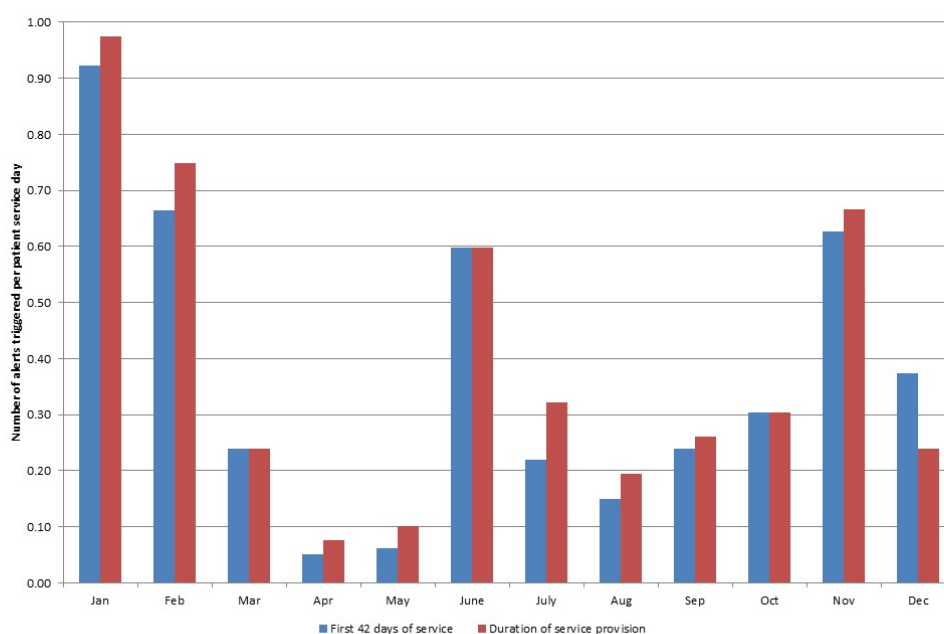

Figure 2. Number of system-generated alerts per month.

Figure 3. Number of system-generated alerts triggered per patient service day.

Alert Triggers

For the 512 alerts triggered during period 1 and period 2, almost half were triggered by just two metrics: feeling more tired than usual (23.4%, 120/512) and taking relieving medication more frequently (23.4%, 120/512), as shown in Table 3. Additional self-reporting (changes in sputum volume or color, increased breathlessness, increased anxiety, disturbed sleep caused by coughing or breathlessness, and swollen feet) accounted for 19.3% (99/512) of triggers. Changes in physiological conditions (blood pressure, oxygen levels, and heart rate) accounted for the remaining 25.8% of alerts (132/512).

Table 3. Patient alerts grouped by trigger (n=512).

Trigger	Patient alerts, n (%)
Feeling more tired than usual	120 (23.4)
Taking relieving medication more often	120 (23.4)
Blood pressure (systolic and/or diastolic)	74 (14.5)
SpO_2 (%) and heart rate	58 (11.3)
Sputum (color or volume)	50 (9.8)
Difficulty with self-care	41 (8.0)
Increased breathlessness	18 (3.5)
Increased anxiety	14 (2.7)
Sleep disturbed by coughing/shortness of breath	9 (1.8)
Swollen feet	8 (1.6)

Clinical Response to Alerts

The nursing notes identified 360 interventions that were undertaken during the study. Clinician interventions undertaken during the study, including responses to system-generated alerts, are identified in Table 4.

Table 4. Alert subcategories, frequency, definitions, and examples of interventions (n=360).

Interventions	Total	Definition/inclusion criteria	Examples
Patient metric review	104	Nursing notes indicated patient within appropriate parameters, or input data "satisfactory"/"stable" so no intervention required	"Trends/Parameters reviewed and all appear satisfactory" [Nurse2]
Patient contact	81	Telephone contact with patient	"Patient contacted by phone regarding his blood pressure, he states that as soon as he is required to check his BP he becomes anxious. In view of this advised him to refrain from checking this and that we will visit next week and check it manually. He is well otherwise and does not require a visit sooner" [Nurse 3]
Orange[a]	52	Nursing notes comment alert changed to orange	"Alert generated [date] passed to Nurse X [date]. Alert changed to orange" [Nurse 7]
Treatment	39	Clinicians provided lifestyle recommendations, medication instructions, or referred patients to seek medical attention from other "clinics"	"Telephoned patient who stated he was more breathless than normal he thought it could be the weather. Advised to increase salbutamol and to contact team if any further problems. Visit by clinician avoided" [Nurse 2]
Note	31	No action taken but relevant note input about the patient	"Nurse practitioner at GP practice informed that [patient's] BP is usually low so will continue to observe" [Nurse 1]
Attempted contact	26	Tried to reach the patient but unable to do so	"Telephoned patient no answer message left to contact if any problems. Visit by Clinician Avoided" [Nurse 2]
Home visit	20	Date of a home visit or scheduled home visit in addition to those on the care pathway	"Patient visited today no complaints, given self-management advice and advised to contact if needed" [Nurse 2]
Hospital	7	Nursing notes indicated patient admitted to hospital	"Admitted to hospital [date] N1;" "Patient contacted by telephone and spoke with wife. She states that [patient] was admitted into hospital this morning following consultation with GP. Therefore Doc@home will not be completed through the next few days" [Nurse 3]

[a] Orange "intervention" auto-generated by the system based on patient parameters. Nursing notes reflected the generation of orange level alert.

Alerts were categorized as "explained" and the rationale documented by nurses within the file, "conditional" as the alert was related to COPD, or an "error" in data input by the patient, as shown in Table 5. A review of patient metrics rarely occurred after a system-generated alert, showing that the clinicians were examining and monitoring patient parameters even when alerts were not being generated. Patient contact was almost always logged after an explained alert (n=61). Treatment interventions consisted of both lifestyle advice and medication adjustments. Only 20 home visit interventions took place during this pilot, and the nursing notes identified 27 instances in which a home visit was avoided through use of the telehealth-monitoring equipment.

Table 5. Clinical intervention subcategories, frequency, definitions, and examples (n=512).

Alerts	Total, n	Definition/inclusion criteria	Examples
Explained	61	Nursing notes indicated a reason for the alert	"Telephoned patient as alerted on being more tired than yesterday. Stated he felt he had just done too much on his allotment. Advised re: pacing, etc; no need to visit" [Nurse 2]
Conditional	447	Nursing notes reveal a connection to COPD causing the alert or unable to determine if the alert was explained so assumed to be triggered by the condition	"Telephone consultation [date]. Patient admits to not taking his inhalers as prescribed resulting n breathlessness. Advised to take his Atrovent and Salamol as instructed to improve condition" [Nurse not identified]
Error	4	Patient input error	"Telephoned patient who stated he had inputted without his oxygen on advised to input with oxygen on stated he was ok" [Nurse not identified]

Table 6 identifies the longest response times between a system-generated alert and a clinical intervention. Of the 260 interventions reported in the system, 152 were preceded by system-generated alerts. In all, 83 of these 152 interventions (54.6%) were delivered within 24 hours of the alert, with the remaining 69 of 152 interventions (45.4%) delivered more than a day after the alert was triggered.

Table 6. Top 10 longest response times following an alert.

Patient number	Intervention date (mm/dd/yy)	Intervention type	Preceding alerts (if any), n	Date of first preceding alert (mm/dd/yy)	Date of most recent preceding alert (mm/dd/yy)	Duration between first alert and intervention (days), n[a]
P14	07/15/11	Patient metric review	3	06/23/11	07/01/11	22
P15	07/15/11	Patient metric review	3	06/24/11	06/26/11	20
P9	03/18/11	Patient metric review	1	03/05/11		13
P16	07/12/11	Patient contact	3	07/03/11	07/11/11	9
P17	07/12/11	Patient contact	6	07/04/11	07/10/11	8
P6	03/11/11	Treatment	3	03/05/11	03/11/11	6
P16	07/01/11	Patient contact	6	06/26/11	06/30/11	5
P22	10/27/11	Patient contact	1	01/22/11		5
P13	06/01/11	Patient contact	6	06/27/11	07/01/11	4
P14	06/27/11	Patient metric review	1	07/23/11		4

[a] Equals number of days between date of first preceding alert and intervention date.

If response times were greater than 1 week, a review of the preceding alerts identified that patients triggered a system alert for being more tired (n=2 alerts), requiring more relieving medication (n=6 alerts), or more tired and requiring more relieving medication (n=4 alerts). After generating system alerts for requiring more relieving medication on July 4 and 9, when patient 17 subsequently generated alerts for their blood pressure, sputum volume, and color on Sunday, July 10, contact was made with the patient within 48 hours.

Patient 6 generated system alerts for requiring more relieving medication than usual on March 5, 6, and 11. On March 4, the patient had been contacted by the service who requested that the patient's general practitioner (GP) either visit in person or provide a prescription of antibiotics. The patient was completing the course of antibiotics in the 6 days that the alerts were generated. On Friday, March 11, the patient telephoned the service stating that they were "struggling with their breathing" and had completed their steroids and antibiotics. Although not at levels to trigger system alerts, the clinician noted that the patient's physiological readings were deteriorating and recommended an out-of-hours GP visit, and arranged for a visit the following day by a community matron (experienced, community-based nurses who coordinate all the health and social care needs for patients with long-term or complicated health conditions).

Health Service Usage

Eight of 23 patients were hospitalized for their COPD, three during the 8-week intervention and five during the 6 months following the intervention. In all, 14 patients (60%) triggered alerts but were able to remain at home. Of the 16 hospital admissions, five were for one patient. Only one patient had an A&E visit without being admitted. Five patients were taken to the hospital by ambulance and were admitted; their admission duration of zero days indicates that they were discharged that same day.

In this study, one patient (P3) was readmitted to hospital on day 5 of the intervention (their first day using the telehealth technology) and another (P13) on day 22 on the intervention after using the telehealth technology for 16 days. The third patient readmitted to hospital while on the intervention (P4) was readmitted on day 31 of the intervention.

All three patients who were hospitalized during the 8-week intervention had alerts prior to their hospitalization and received treatment interventions. One patient (P13) had alerts for 3 days consecutively prior to their hospitalization. On the first day, they generated alerts for two of eight NHS AECOPD key symptoms—being tired and having difficulty with self-care—in addition to alerts related to their blood pressure, SpO_2 (%), and heart rate readings. On the second day, alerts were generated for their blood pressure, SpO_2 (%), and heart rate readings. A clinician telephoned the patient and noted the patient said they "might not be inputting correctly." The clinician noted that a visit was not required at this point, and that they would review the case again the following day to determine whether a home visit was necessary. On the third day, the patient again triggered alerts for their blood pressure, SpO_2 (%), and heart rate readings, and the system generated an "orange" alert status for the patient. On the following day, before the service had an opportunity to arrange a home visit, the patient was admitted to the hospital and remained there for a week.

A second patient (P3) also generated system alerts prior to their hospitalization, which occurred on their first day of being on the intervention. They triggered an alert due to requiring more relieving medication. Later that day they were taken by ambulance to A&E and were subsequently admitted overnight. On discharge from the hospital, on their second day of using

the system, they generated an alert because they were much more anxious than usual. On their third day, they again generated an alert due to their anxiety in addition to feeling more tired than usual and later that day their spouse contacted the service to say that the patient had again been admitted to hospital, although this admission was not recorded in the local hospital SUS data. This patient continued to require significant assistance with their condition and was admitted to hospital a further four times in the 6 months following discharge from the community telehealth-enabled service.

The third patient (P4) generated multiple system alerts almost every day they were on the intervention. The clinicians contacted the patient regularly and provided dietary advice to alleviate symptoms and recommended medication changes. In the week prior to their admission to hospital, this patient triggered multiple alerts every day except one (four days before their admission). The clinicians contacted the patient on four of the days alerts were generated. They also noted that the patient's GP visited on the day before their hospital admission, and on the subsequent day when they recommended that the patient should go to the hospital. Once again, this admission was not recorded in the local hospital SUS data.

Seven of 39 (18%) treatment interventions administered to all patients throughout the study were provided to just one patient. Although this patient generated no more than one system alert per day, and always for the same prompt ("Have you had to use your relieving medication in the last 24 hours?"), there were ongoing medication adjustments and recommendations. The nursing notes also showed short periods of stasis after an alert in which the patient seemed to stabilize and reported feeling well. Despite the significant number of alerts, clinical assistance, and self-management advice, this patient was never hospitalized.

Discussion

Our findings show that the volume, frequency, and type of alerts generated by a telehealth monitoring system is a good indicator of need for clinical contact and/or intervention for patients with early-stage COPD when first discharged from hospital following an exacerbation.

Use of the Telehealth Monitoring System

Allowing for days in hospital, on holiday, or when there was a system fault when the patient could not reasonably be expected to use the system, patients failed to provide data on only 29 of 1175 days (2.47%) they had the system installed in their homes. Although little is currently understood around adherence to health technology interventions, or the required level of adherence for an effect on outcomes [21], this adherence level is higher than others seen in the telehealth field [22]. As a side note, holidays and hospitalizations were not taken into account in this project.

Only four alerts were triggered by patient input error. The ease of use of the equipment was later confirmed in interviews with the patients following completion of the intervention and discharge from the clinical service [23].

Volume and Frequency of Alerts

As shown in Figure 1, the mean number of patients triggering alerts each day declined over the course of the intervention. This could imply that patients are more able to manage their disease, possibly as a result of the self-management advice provided and documented in the nursing notes, or that they experienced fewer symptoms.

It is currently unclear whether use of telehealth promotes or inhibits self-management behavior in people with COPD [24]. Telehealth can empower patients to self-manage their condition through facilitating increased knowledge of the condition and its symptoms [25], which supports the argument that participants in this study gained knowledge, became better at managing their condition, and triggered fewer alerts. However, it is also argued that telehealth may increase dependence on health care services, with reassurance provided through being "watched over" by health care professionals and the knowledge that they will intervene if something goes wrong [24]. To summarize, it is unclear why alerts reduced with time spent with the equipment, although it could be as simple as the fact that participants were recovering from a hospitalization when they first began using the equipment [20].

Although only 22 of 75 patient service days were delivered during December, one patient was responsible for all 28 alerts triggered that month. This patient generated a high number of alerts (n=77) throughout their time on the service, and was hospitalized twice while receiving the intervention, with one of these admissions occurring in mid-December.

Alert Triggers

The most fundamental question upon which this entire analysis relies is whether an alert in the Doc@Home system means that the patient is experiencing an AECOPD. Although 12 system prompts and the four Bluetooth readings provided by the patients could trigger a system alert, there were three prompts—(1) "Have you had any problems in walking about in the past week?" (2) "Have you had problems doing your usual activities in the past week?" and (3) "Have you had to contact the following (health services) within the last 24 hours?"—that did not trigger system alerts regardless of the response entered by the patient. Similarly, answers to the question "How has your general health been in the last 24 hours?" were either not captured or were not reported in the data downloaded from the system. Because questions created on the telehealth system to be posed to participants are determined by clinicians, it would seem logical that all metrics should be set to trigger an alert if they indicate a significant decline in patient health status that would merit follow-up.

Two system prompts that did trigger system alerts—"Have you had to use your relieving medication in the last 24 hours?" and "How anxious have you been over the last 24 hours?"—did not appear to be tied to the NHS AECOPD list of key symptoms. Given that they both triggered a number of system prompts, we question whether they should be included in the key symptoms of an AECOPD.

There were no system prompts that could reasonably be linked to NHS AECOPDs symptoms of chest tightness or acute

confusion. Although the latter may be difficult to incorporate into a remote monitoring system completed by a patient, the former could be easily included in the prompts and could alert clinicians to the onset of an AECOPD.

It is important to note that an explained alert does not mean that the patient's COPD condition had no effect on triggering that alert. It means that, because of their COPD, as shown in Table 5, other variables—be they physical activity or the weather—exacerbated their condition, which then triggered a system alert. Simply having COPD was not the cause of the alert, but rather a confluence of factors. Jehn et al [26] explored the effects of heat stress (days warmer than 25°C) on AECOPDs, finding "heat stress negatively impacts clinical and functional status in patients with COPD and makes patients more vulnerable for disease-related morbidity." In addition, the amount or type of activities of daily living undertaken may positively or negatively influence COPD symptoms, such as dyspnea and fatigue [27].

The care pathway indicated that patient-specific parameters for the four Bluetooth readings (blood pressure, heart rate, and SpO_2) should be reviewed and amended following 10 days of data entry to eliminate unnecessary subsequent system alerts. Although it appears that some patients did have their parameters adjusted, this did not appear to occur for all patients. Patient 13 triggered system alerts for blood pressure, SpO_2 (%), and heart rate for 17 of their 20 alerts.

Clinical Response to Alerts

An AECOPD is a worsening of symptoms that, according to some definitions [12,15,28], can include a change in medication. There were 39 instances in which either a medication change or a lifestyle adjustment as a form of treatment was recommended to improve the patient's health. Although the patient's health status may have otherwise warranted a home visit, the telehealth monitoring system reduced unscheduled visits. This was confirmed in the nursing notes where the need for a visit had been averted in 26 of 360 clinician interventions; a search of the telehealth system data identified a "visit by clinician avoided" (n=4) and a "visit not required" (n=22). However, current evidence demonstrates that a reduction in health care utilization does not necessarily translate into cost effectiveness when looking at the overall cost of providing telehealth in comparison with usual care [29]. This study did not demonstrate cost savings for telehealth, and overall use of health care services increased [20].

The reason for the increased response time for patients who triggered more system alerts is unclear. The additional alerts may indicate a more complex patient requiring more time to develop a suitable care plan. Alternatively, the clinicians were acclimatized to the patients generating alerts more frequently. For example, for one patient the nursing notes state, "still struggles with ADLs [activities of daily living], but this is not new for [patient]." This particular patient generated the most alerts and experienced two hospitalizations during the course of their intervention and the 6-month follow-up period. Because difficulty with self-care is listed as one of the NHS AECOPD key symptoms, we question whether a more active response to

recurring alerts may have halted the patient's decrease in health status and ultimate hospitalizations.

A "treatment" was recommended only 11% of the time, with recommendations including referrals to pulmonary rehabilitation programs (PRPs). PRPs are designed to "help people with chronic lung problems" through "exercise and education...by a multidisciplinary team, which includes physiotherapist, respiratory nurse specialists, and dieticians" [30]. A search of the system data identified that the PRP option was only discussed with three patients on four separate occasions. We suggest that PRPs could be used more frequently as a treatment option to prevent or defer hospitalization. After days with treatment recommendations, such as days 10 and 11, the following days 12 to 15 showed a return to stability. When issues arose on day 18 through day 32, a variety of different treatment methods were administered by the nursing team and the patient's GP. Subsequently, the patient's health status stabilized and remained satisfactory for until day 52 when they were discharged from the service. We suspect that without this intervention, the options available to the patient would have been more limited or slower to access. The telehealth monitoring system seemed to help this patient overcome exacerbations of their COPD.

Health Service Usage

As noted by Bentley et al [20], patients receiving the telehealth intervention were more than twice as likely to be admitted to hospital and for six times longer than the control group. This would seem to suggest, as in other studies [17,31,32], that telehealth monitoring fails to achieve the objective of keeping patients out of hospital or reduce health service usage. However, our analysis suggests that there may be a way to identify AECOPDs with enough time to intervene and thus prevent hospital readmission. Table 6 identifies an increase in alerts just before hospital admission for the three patients hospitalized during the intervention. Paired with the treatment intervention attempts, this suggests that the clinicians are recognizing the onset of an exacerbation and are attempting to intervene through treatment that includes medication changes, referrals to PRPs, and/or lifestyle advice.

The UK norm is 20% of patients are readmitted to hospital within 28 days of an AECOPD [30,34]. In this study, one of the three readmissions was on the first day of the intervention, so there was no opportunity for the service to have an impact. We are cautious of extrapolating from our relatively small sample size but, ignoring this admission, the hospital readmission rate of just 9% (2/23) suggests that the ability of this service to reduce the frequency of hospital readmissions merits further investigation. Although two of the patients saw an increase in alerts for the captured metrics before hospitalization, the third patient did not, so it is not possible to say that the metrics used can definitely predict an AECOPD. Two patients were recommended to go to hospital by their GPs, suggesting that the nursing team and/or patient engaged other health care providers in an effort to exhaust other options before returning to hospital. It appears as though increased effort, clinician attention, medication, and system funding were expended on these patients and yet a readmission was not

avoided. The service was designed to support early intervention, which is critical for preventing worsening of an AECOPD [33], but NHS restructuring, the loss of a key champion for the trial staff, and attrition resulted in an eventual total loss of 60% of staff capacity within the frontline clinical team. With this significant decrease in resources, the clinicians clearly used the system to identify patients who demonstrated greatest need of support, and did their utmost to contact these patients and to alleviate their symptoms.

Interrogation of System Data

Although telehealth monitoring systems are designed for analysis of data by clinicians at the time of care delivery, retrospectively downloaded data has proved to be highly time consuming and complex to analyze. With only 23 participants in this study, we were able to undertake this task; for a larger study, this could prove to be insurmountable unless the data could be downloaded in a format better suited to analysis. We question whether this is the reason system data analysis has not been undertaken or published to date. These data provide critical insights into service provision and patient profiles; therefore, technology providers need to consider not only frontline care providers, but also staff conducting clinical audits, service improvement studies, or research. The uptake of a new intervention requires analysis of all evidence available to demonstrate effectiveness. It seems rather ironic that an intervention designed to capture and report remotely provided data is unable to better support retrospective analysis, and we strongly encourage technology providers to ensure that this current deficiency is addressed.

Study Limitations

The SUS information taken from the PCT NHS records showed some discrepancies when compared to the telehealth system nursing notes that would have affected analysis if they had not been identified. For example, a hospital admission was not shown in the SUS data; however, the telehealth system nursing notes included specific details about a hospital admission of at least 3-days duration. It is possible that the patient was hospitalized outside of the PCT for the region in which the study was conducted and consequently would not be captured within their SUS dataset.

When the clinician viewed the telehealth system at the time of care delivery, the display included color-coded alerts (red, orange, or yellow) [15]. When the same data are retrospectively downloaded from the server for research purposes, with the exception of the orange alert (see Table 4), these color codes were not identified. Alerts were only identified as dichotomous "yes" or "no" nominal data. Consequently, there is no way to determine what specific parameters caused alerts to be yellow, orange, or red. It is also not possible to tell whether an alert had previously been displayed at a yellow level and a change to orange indicated an increase in priority or conversely if the previous alert was red and a change to orange indicated a decrease in urgency. Consequently, our analysis has been restricted to the dichotomous data available to us.

Conclusions

The pilot RCT did not identify a reduction in health care usage; in fact, it had a higher rate of usage among the telehealth group relative to the control [20]. However, from our system data analysis we suggest that this telehealth monitoring-supported service could fulfill an important role: enabling patients with COPD to better manage their condition. Although we question whether the addition of further prompts to assess all the key NHS AECOPD symptoms would further improve the system, the prompts used during this study did seem able to identify when a patient may be experiencing an exacerbation of their COPD and may require clinical intervention. Identification of an AECOPD is insufficient and, to alleviate symptoms and enable a patient to remain at home, timely intervention is required. Where service resourcing limits service capacity, a delay in responding to system-generated alerts can result in patients being hospitalized, thereby negating the value of the intervention and the service. It is clear—predominantly through remote interaction with the patient—that during the intervention, the service successfully enabled a number of patients to remain at home despite exacerbations in their COPD.

Some of the patients were clearly unstable following their discharge from hospital, triggering a high number of system alerts. With a larger study to eliminate the impact of a small number of more complex patients, and more consistent clinical resourcing [20] to enable timely intervention, we question whether the outcomes of the pilot study would have been somewhat different.

We recommend running a larger study using telehealth monitoring with prompts to assess all 10 of the NHS AECOPD symptoms in addition to the two non-NHS AECOPD-related prompts that generated alerts in this study to determine which prompts may assist in the accurate identification of an AECOPD. To test whether removal of the system does defer hospitalizations, we further suggest conducting the study such that following a standard "intervention" period, in which one study group has the service removed, while another continues to use the service for a comparable 6-month period to determine whether there is any difference between group hospitalization rates.

XSL•FO

RenderX

Authors' Contributions

MK undertook data analysis and contributed to the first draft of this manuscript. DAF was involved in the design of the project, developed the pilot study protocol, obtained NHS ethics and governance approval, was responsible for management of the project, conducted data analysis, and wrote the manuscript. CLB contributed to management of the project and carried out revisions to the manuscript. GAM was Principal Investigator and was involved in the conception and design of the study, and carried out revisions to the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Department of Health. An Outcomes Strategy for Chronic Obstructive Pulmonary Disease (COPD) and Asthma in England. 2011. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216139/dh_128428.pdf [accessed 2016-12-23] [WebCite Cache ID 6mvCHOJDw]

2. Department of Health. A Strategic Approach to Prevention and Early Identification of COPD. 2011. URL: http://www.respiratoryresources.org.uk/COPD.html [accessed 2016-12-23] [WebCite Cache ID 6iF1NMQ7X]

3. Meldrum R, Rawbone A, Curran D, Fishwick D. The role of occupation in the development of chronic obstructive pulmonary disease (COPD). Occup Environ Med 2005 Apr;62(4):212-214 [FREE Full text] [doi: 10.1136/oem.2004.015511] [Medline: 15778251]

4. British Lung Foundation. Invisible Lives: Chronic Obstructive Pulmonary Disease (COPD) - Finding the Missing Millions. 2007. URL: http://spirohub.com/wp-content/uploads/2016/01/BLF-Invisible-Lives-report.pdf [accessed 2016-12-23] [WebCite Cache ID 6iF3lP2gT]

5. Prieto-Centurion V, Gussin HA, Rolle AJ, Krishnan JA. Chronic obstructive pulmonary disease readmissions at minority-serving institutions. Ann Am Thorac Soc 2013 Dec;10(6):680-684 [FREE Full text] [doi: 10.1513/AnnalsATS.201307-223OT] [Medline: 24364772]

6. Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. Lancet 2007 Sep 1;370(9589):765-773. [doi: 10.1016/S0140-6736(07)61380-4] [Medline: 17765526]

7. Ram FSF, Wedzicha JA, Wright J, Greenstone M. Hospital at home for patients with acute exacerbations of chronic obstructive pulmonary disease: systematic review of evidence. BMJ 2004 Aug 7;329(7461):315 [FREE Full text] [doi: 10.1136/bmj.38159.650347.55] [Medline: 15242868]

8. Gruffydd-Jones K, Langley-Johnson C, Dyer C, Badlan K, Ward S. What are the needs of patients following discharge from hospital after an acute exacerbation of chronic obstructive pulmonary disease (COPD)? Prim Care Respir J 2007 Dec;16(6):363-368 [FREE Full text] [doi: 10.3132/pcrj.2007.00075] [Medline: 18038104]

9. NHS Royal Cornwall Hospitals. Clinical Guidelines for the Management of Acute Exacerbations of COPD. URL: [WebCite Cache ID 6iF24ZaqN]

10. Saleh S, Larsen JP, Bergsåker-Aspøy J, Grundt H. Re-admissions to hospital and patient satisfaction among patients with chronic obstructive pulmonary disease after telemedicine video consultation - a retrospective pilot study. Multidiscip Respir Med 2014;9(1):6 [FREE Full text] [doi: 10.1186/2049-6958-9-6] [Medline: 24475774]

11. Fernández-Granero MA, Sánchez-Morillo D, León-Jiménez A, Crespo LF. Automatic prediction of chronic obstructive pulmonary disease exacerbations through home telemonitoring of symptoms. Biomed Mater Eng 2014;24(6):3825-3832. [doi: 10.3233/BME-141212] [Medline: 25227099]

12. Wedzicha JA, Seemungal TA. COPD exacerbations: defining their cause and prevention. Lancet 2007 Sep 1;370(9589):786-796. [doi: 10.1016/S0140-6736(07)61382-8] [Medline: 17765528]

13. Garcia-Aymerich J, Farrero E, Félez MA, Izquierdo J, Marrades RM, Antó JM, Estudi del Factors de Risc d'Agudització de la MPOC investigators. Risk factors of readmission to hospital for a COPD exacerbation: a prospective study. Thorax 2003 Feb;58(2):100-105 [FREE Full text] [Medline: 12554887]

14. Burge S, Wedzicha JA. COPD exacerbations: definitions and classifications. Eur Respir J Suppl 2003 Jun;41:46s-53s [FREE Full text] [Medline: 12795331]

15. Segrelles CG, Gómez-Suárez C, Soriano JB, Zamora E, Gónzalez-Gamarra A, González-Béjar M, et al. A home telehealth program for patients with severe COPD: the PROMETE study. Respir Med 2014 Mar;108(3):453-462 [FREE Full text] [doi: 10.1016/j.rmed.2013.12.003] [Medline: 24433744]

16. Toy EL, Gallagher KF, Stanley EL, Swensen AR, Duh MS. The economic impact of exacerbations of chronic obstructive pulmonary disease and exacerbation definition: a review. COPD 2010 Jun;7(3):214-228. [doi: 10.3109/15412555.2010.481697] [Medline: 20486821]

17. Berkhof FF, van den Berg JW, Uil SM, Kerstjens HA. Telemedicine, the effect of nurse-initiated telephone follow up, on health status and health-care utilization in COPD patients: a randomized trial. Respirology 2015 Feb;20(2):279-285 [FREE Full text] [doi: 10.1111/resp.12437] [Medline: 25400242]

XSL·FO

RenderX

18.   Whitten P, Mickus M. Home telecare for COPD/CHF patients: outcomes and perceptions. J Telemed Telecare 2007;13(2):69-73. [doi: 10.1258/135763307780096249] [Medline: 17359569]

19.   Fitzsimmons DA, Thompson J, Hawley M, Mountain GA. Preventative tele-health supported services for early stage chronic obstructive pulmonary disease: a protocol for a pragmatic randomized controlled trial pilot. Trials 2011;12:6 [FREE Full text] [doi: 10.1186/1745-6215-12-6] [Medline: 21214895]

20.   Bentley CL, Mountain GA, Thompson J, Fitzsimmons DA, Lowrie K, Parker SG, et al. A pilot randomised controlled trial of a Telehealth intervention in patients with chronic obstructive pulmonary disease: challenges of clinician-led data collection. Trials 2014;15:313 [FREE Full text] [doi: 10.1186/1745-6215-15-313] [Medline: 25100550]

21.   Cudd PA, Otesile O, Bentley CL. Patient adherence: what does it mean, implications for digital health systems. In: Global Telemedicine and eHealth Updates: Knowledge Resources. 2016 Presented at: Med-e-Tel 2016; Apr 6-8, 2016; Luxembourg p. 28-33 URL: https://www.medetel.eu/?rub=proceedings&page=info

22.   Guzman-Clark JR, van Servellen G, Chang B, Mentes J, Hahn TJ. Predictors and outcomes of early adherence to the use of a home telehealth device by older veterans with heart failure. Telemed J E Health 2013 Mar;19(3):217-223. [doi: 10.1089/tmj.2012.0096] [Medline: 23268695]

23.   Fitzsimmons DA, Thompson J, Bentley CL, Mountain GA. Comparison of patient perceptions of Telehealth-supported and specialist nursing interventions for early stage COPD: a qualitative study. BMC Health Serv Res 2016 Aug 22;16(1):420 [FREE Full text] [doi: 10.1186/s12913-016-1623-z] [Medline: 27549751]

24.   Brunton L, Bower P, Sanders C. The contradictions of telehealth user experience in chronic obstructive pulmonary disease (COPD): a qualitative meta-synthesis. PLoS One 2015;10(10):e0139561 [FREE Full text] [doi: 10.1371/journal.pone.0139561] [Medline: 26465333]

25.   Gorst SL, Coates E, Armitage CJ. "It's sort of a lifeline": Chronic obstructive pulmonary disease patients' experiences of home telehealth. Health Psychol 2016 Jan;35(1):60-68. [doi: 10.1037/hea0000246] [Medline: 26148189]

26.   Jehn M, Donaldson G, Kiran B, Liebers U, Mueller K, Scherer D, et al. Tele-monitoring reduces exacerbation of COPD in the context of climate change-a randomized controlled trial. Environ Health 2013;12:99 [FREE Full text] [doi: 10.1186/1476-069X-12-99] [Medline: 24261700]

27.   Stridsman C, Lindberg A, Skär L. Fatigue in chronic obstructive pulmonary disease: a qualitative study of people's experiences. Scand J Caring Sci 2014 Mar;28(1):130-138. [doi: 10.1111/scs.12033] [Medline: 23517049]

28.   Global Initiative for Chronic Obstructive Lung Disease. 2016. Global strategy for the diagnosis, management and prevention of COPD URL: http://goldcopd.org/ [accessed 2017-02-23] [WebCite Cache ID 6oV4OGWes]

29.   Henderson C, Knapp M, Fernández J, Beecham J, Hirani SP, Cartwright M, Whole System Demonstrator evaluation team. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. BMJ 2013;346:f1035 [FREE Full text] [Medline: 23520339]

30.   National Health Service. Chronic obstructive pulmonary disease: treating COPD URL: http://www.nhs.uk/Conditions/ Chronic-obstructive-pulmonary-disease/Pages/Treatment.aspx [accessed 2016-12-23] [WebCite Cache ID 6iF2VR3AF]

31.   Suh E, Mandal S, Hart N. Admission prevention in COPD: non-pharmacological management. BMC Med 2013;11:247 [FREE Full text] [doi: 10.1186/1741-7015-11-247] [Medline: 24252219]

32.   Royal College of Physicians, British Thoracic Society, British Lung Foundation. Report of The National Chronic Obstructive Pulmonary Disease Audit: Clinical Audit of COPD Exacerbations Admitted to Acute NHS Units Across the UK. 2008. URL: https://www.rcplondon.ac.uk/file/1037/download?token=Gsv5H38I [accessed 2016-12-23] [WebCite Cache ID 6mv9ovGOk]

33.   Wilkinson TM, Donaldson GC, Hurst JR, Seemungal TA, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004 Jun 15;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]

34.   Anstey K, Lowe D, Roberts C, Hosker H. Report of the 2003 National COPD Audit. 2004. URL: https://www.rcplondon.ac.uk/ file/1045/download?token=iLNh-YZp [accessed 2016-12-23] [WebCite Cache ID 6iF2zVua3]

Abbreviations

A&E: Accident and Emergency
AECOPD: acute exacerbation of chronic obstructive pulmonary disease
COPD: chronic obstructive pulmonary disease
GP: general practitioner
NHS: National Health Service
PCT: Primary Care Trust
PRP: pulmonary rehabilitation program
RCT: randomized controlled trial
SUS: Secondary Uses Service

XSL•FO

RenderX

Original Paper

Patient-Specific Predictive Modeling Using Random Forests: An Observational Study for the Critically Ill

Joon Lee[1], PhD

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

Corresponding Author:
Joon Lee, PhD
Health Data Science Lab
School of Public Health and Health Systems
University of Waterloo
200 University Ave W
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567 ext 31567
Fax: 1 519 746 6776
Email: joon.lee@uwaterloo.ca

Abstract

Background: With a large-scale electronic health record repository, it is feasible to build a customized patient outcome prediction model specifically for a given patient. This approach involves identifying past patients who are similar to the present patient and using their data to train a personalized predictive model. Our previous work investigated a cosine-similarity patient similarity metric (PSM) for such patient-specific predictive modeling.

Objective: The objective of the study is to investigate the random forest (RF) proximity measure as a PSM in the context of personalized mortality prediction for intensive care unit (ICU) patients.

Methods: A total of 17,152 ICU admissions were extracted from the Multiparameter Intelligent Monitoring in Intensive Care II database. A number of predictor variables were extracted from the first 24 hours in the ICU. Outcome to be predicted was 30-day mortality. A patient-specific predictive model was trained for each ICU admission using an RF PSM inspired by the RF proximity measure. Death counting, logistic regression, decision tree, and RF models were studied with a hard threshold applied to RF PSM values to only include the M most similar patients in model training, where M was varied. In addition, case-specific random forests (CSRFs), which uses RF proximity for weighted bootstrapping, were trained.

Results: Compared to our previous study that investigated a cosine similarity PSM, the RF PSM resulted in superior or comparable predictive performance. RF and CSRF exhibited the best performances (in terms of mean area under the receiver operating characteristic curve [95% confidence interval], RF: 0.839 [0.835-0.844]; CSRF: 0.832 [0.821-0.843]). RF and CSRF did not benefit from personalization via the use of the RF PSM, while the other models did.

Conclusions: The RF PSM led to good mortality prediction performance for several predictive models, although it failed to induce improved performance in RF and CSRF. The distinction between predictor and similarity variables is an important issue arising from the present study. RFs present a promising method for patient-specific outcome prediction.

Introduction

Harnessing the information contained in health data from various sources toward personalized medicine has been a research topic of interest and discussed by a number of highly regarded health researchers recently [1-5]. In particular, patient outcome prediction is an important topic in health care since accurate prognostic information can inform treatment planning and resource allocation. While prognostic scoring systems have traditionally been developed based on large population studies, the current rapid transition to electronic health records (EHRs) has led to an increased interest in data-driven, patient-specific outcome prediction models. Large-scale EHR data enable

personalized predictive models where the degree of similarity between an index patient (for whom a prediction is to be made) and a past patient (the clinical data of whom can be found in an EHR repository) is taken into consideration. Such a personalized approach ensures that the predictive model is optimized for the index patient rather than the average patient in the population by building a customized predictive model just for the index patient.

A key to personalized patient outcome prediction is how patient similarity is defined. Various similarity measures have been investigated in the context of EHR-based outcome prediction, including distance-based [6,7] and cluster-based [8] methods. In addition, propensity score matching [9] employs a similar approach by identifying patients with similar likelihoods of receiving the treatment under investigation. A defined similarity measure is called a patient similarity metric (PSM), which can be calculated between patients. Subsequently, PSM values can be used to either discard the EHR patient data below a certain threshold or weight EHR patients' contributions to predictive modeling proportionately to the PSM magnitude.

Outside of health, several domains have employed similarity approaches in machine learning and predictive analytics. One prime example is product recommendation in e-commerce where purchase histories of similar consumers are leveraged to recommend products to a given customer [10]. Furthermore, a variety of subspace clustering [11,12] and mixture models [13] have been developed and applied to identify and use similar cases across different application domains.

Our previous studies in this line of research investigated a cosine similarity PSM for personalized mortality prediction in intensive care unit (ICU) patients with hard thresholding [14] and bagging [15]. The results were promising and showed that using data from only similar patients, rather than the entire available data set, leads to better predictive performance. In order to study the effects of the particular choice of PSM on the results, however, it is worth investigating other PSMs.

One interesting way to define a PSM is to use the random forest (RF) proximity measure, which represents the likelihood of 2 cases falling in the same terminal node in the trees of an RF [16]. Xu and colleagues have developed the case-specific random forest (CSRF) using the RF proximity measure as a bootstrap sampling weight [17]. While this approach used RF proximity in personalizing the standard RF, the RF proximity can also be used as an independent PSM for predictive models other than the RF. Being a stochastic similarity measure, RF proximity is certainly distinct from the cosine similarity PSM we previously studied and as a result may capture patient similarity from a different perspective. Also, other similar approaches reviewed above (propensity score matching, subspace clustering, etc) tend to rely on well-known distance measures (such as the Euclidean or Mahalanobis) or clustering algorithms (such as K-means) to identify similar cases. To the best of our knowledge, an RF-inspired PSM has never been investigated for the purpose of patient outcome prediction in the ICU.

The objective of this study was to evaluate the effectiveness of RF patient similarity on improving mortality prediction performance in the critically ill.

Methods

Patient data were extracted from the public ICU database Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) version 2.6 [18,19]. Because MIMIC-II is a deidentified, publicly available database, the need for a research ethics review was waived for this study. In order to make a direct comparison with our previous results [14,15], the same data set of 17,152 ICU admissions was analyzed. Table 1 lists the predictor variables extracted from MIMIC-II. The outcome variable was 30-day mortality. ICU admissions with missing data were excluded, and each included ICU admission was treated as an independent "patient," as was done before.

Table 1. List of predictor variables.

Category	Predictor variables
Demographics	Age, gender
Administrative information	Admission type (elective, urgent, emergency), ICU[a] service type (MICU[b], SICU[c], CCU[d], CSRU[e])
Vital signs (min. and max. every 6 hours during the first 24 hours in the ICU)	Heart rate, mean blood pressure, systolic blood pressure, SpO$_2$, spontaneous respiratory rate, body temperature
Labs (min. and max. from the first 24 hours in the ICU	Hematocrit, white blood cell count, serum glucose, serum HCO$_3$, serum potassium, serum sodium, blood urea nitrogen, and serum creatinine
Intervention (yes/no during the first 24 hours in the ICU)	Vasopressor therapy, mechanical ventilation, or continuous positive airway pressure
Others (from the first 24 hours in the ICU)	Worst Glasgow Coma Scale score, total urinary output every 6 hours

[a]ICU: intensive care unit.

[b]MICU: medical intensive care unit.

[c]SICU: surgical intensive care unit.

[d]CCU: coronary care unit.

[e]CSRU: cardiac surgery recovery unit.

XSL•FO

RenderX

In MIMIC-II, there are 1.24 ICU admissions per patient on average which indicates that most patients have only one ICU admission. Other admissions from the same patient are likely to contain useful and relevant information since they represent the most similar "patients" (determined by the PSM) for the index patient. This is akin to incorporating past patient history in most prognostic scoring systems. Most importantly, if the data from other admissions from the same patient are available at the time of prediction, it would be a waste of data if they are not taken into account.

Out of 29,149 adult ICU admissions in MIMIC-II, 17,152 (58.84%) had complete data and were included in the present study.

In the data set, the overall 30-day mortality rate was 15.10% (4,401/29,149), while 56.70% (16,527/29,149) were male. The average age was 64.5 years with a standard deviation of 17.0. The percentages of elective, urgent, and emergency admissions were 18.00% (5,247/29,149), 3.70% (1,078/29,149), and 78.30% (22,824/29,149), respectively. More detailed descriptive statistics of the data set can be found elsewhere [14]. All data were extracted from MIMIC-II using SQL Developer version 3.2.09 (Oracle Corp).

Following how bootstrap sampling weights were calculated in CSRF [17], the RF PSM in the present study was calculated first by growing an RF in unsupervised mode using all data with an *mtry* value equal to the total number of predictor variables, a *nodesize* of 5, and 500 trees. When used in an unsupervised manner, RFs are capable of developing a dissimilarity measure among unlabeled data, which has been deployed for differentiating between observed and synthetic data [16,20]. The proximity values from this RF were normalized as displayed in Figure 1 to quantify the similarity between an index and another patient.

Figure 1. Random forest patient similarity metric formula.

$$RFPSM_i(j) = \frac{Prox_{i,j}}{\sum_{k=1}^{N} Prox_{i,k}}, i = 1, \ldots, N, j = 1, \ldots, N, i \neq j, i \neq k$$

where i and j refer to the index and the jth patient in the data, respectively, and $Prox_{i,j}$ is the number of trees in the grown RF that have both the index and jth patient in the same terminal node. This RF PSM was calculated for every pair of patients in the data.

For every index patient, the M most similar patients in the training data (ie, hard thresholding on RF PSM) were used to train a customized predictive model. A total of 4 different models were evaluated with hard thresholding: death counting (DC, predicted mortality risk is equal to the empirical probability of death among the M most similar patients), logistic regression (LR), decision tree (DT), and RF.

The range of M values varied depending on the predictive model. For DC, M ranged from 10 to 15,000 with a step size of 10, whereas for LR and RF the range was from 4000 to 15,000 with a step size of 1000. The M range for DT was from 5000 to 15,000 with a step size of 1000. This variation in M range accounted for computational burden and lack of variability in categorical variables (either predictor or outcome) among

the M most similar patients when M was sufficiently small (only 1 category remained in some categorical variables when the M patients were too homogenous). Moreover, for LR, DT, and RF, training data size had to be sufficiently large (at least 1000), given that there were 75 predictor variables (Table 1). Since DC was associated with the least computational burden and was not subject to the issue of insufficient variability in categorical variables, it was evaluated with the widest M range with the smallest step size. The lower ends of the ranges for LR, DT, and RF were determined by trial and error. The step size of 1000 resulted in sufficient resolution allowing identification of predictive performance patterns, as will be evident in the Results section. In addition to the M ranges specified above, all training data were also used to represent traditional predictive modeling except for DC where using all training data implies using the overall mortality rate in the entire training data as the predicted mortality risk for all patients. In any case, 15,000 was very close to using all training data which included approximately 15,500 patients.

Note that we did not attempt to select the optimal M value for each patient, since our objective was to investigate the effects of the RF PSM on prediction performance as a function of M, as was done in our previous work [14].

As a fifth predictive model, CSRF was investigated. The entire data set was used for training CSRF models since the RF PSM was used as bootstrap sampling weights instead. Since this is a soft thresholding method, applying a range of M values was not applicable to CSRF.

Note that while DC, LR, and DT were investigated previously [14], RF and CSRF were not. RF and CSRF were included in this study in the spirit of conducting a comprehensive RF investigation in personalized predictive analytics.

Predictive performance was evaluated using the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC). A 10-fold cross-validation was conducted for all models to avoid overfitting. In each iteration of the cross-validation, the RF PSM between each patient in the test data and each patient in the training data was computed. Then, for each patient in the test data, the predictive models described above were trained using the M most similar cases in the training data (except CSRF), where M was varied as explained above. Once all patients in the test data were predicted, the AUROC and AUPRC for that fold were computed. Hence, the 10 iterations of the cross-validation yielded 10 AUROCs and 10 AUPRCs.

All computation was conducted in R version 3.3.1 (R Foundation). In particular, the *randomForest* package [21] was used to build RFs (with the default parameter values except for the unsupervised RF for proximity measure calculation, for which the parameter values were described above). CSRF models were constructed using the R code supplied by Xu et al [17], with a *nodesize* of 1 and 500 trees, while *mtry* was set to the same default value as the *randomForest* package (ie, floor of the square root of the number of predictors). LR and DT models were created using the *stats* and *rpart* packages, respectively, with the default parameter settings.

Results

Figures 2 and 3 show the predictive performances of DC, LR, DT, and RF with hard thresholding, in terms of AUROC and AUPRC, respectively, as a function of the number of similar patients included in model training (ie, M). The pattern of exhibiting suboptimal performances when M is too small (due to small sample size) or too large (due to dissimilar patients being included in the training data) and a peak performance somewhere in the middle was most prominent in DC and mildly visible in LR in terms of both AUROC and AUPRC. Interestingly, DT showed 2 local maxima in Figure 2 and weakly showed 1 peak in Figure 3. RF did not seem to benefit from hard thresholding on RF PSM as its performance was relatively independent of M.

DC and DT showed statistically significant improvement (via 2-sided *t* tests) between the best performance (in terms of mean

AUROC or AUPRC) and when the maximum number of patients were used as training data, with respect to both AUROC (DC: *P*<.001; DT: *P*<.001) and AUPRC (DC: *P*<.001; DT: *P*<.001). LR and RF showed statistically significant performance improvement for neither AUROC (LR: *P*=.07; RF: *P*=.75) nor AUPRC (LR: *P*=.36; RF: *P*=.85).

Table 2 tabulates the best predictive performance of each model with respect to mean AUROC and AUPRC. The numbers of similar patients at which the best performance occurred as well as the performance of CSRF are also reported in Table 2. Figures 4 and 5 (AUROC and AUPRC) are boxplots that correspond to the best performances shown in Table 2 and enable a quick visual comparison among all 5 models. Note that these best performances simply correspond to the maximum mean AUROC or AUPRC and do not represent statistically significant peak performances, as evident in Figures 2 and 3. Overall, RF and CSRF resulted in the best performances, followed by LR, DC, and DT in decreasing order of performance.

Table 2. Best predictive performance from each random forest patient similarity metric (PSM) model in terms of mean area under the receiver operating characteristic curve and area under the precision-recall curve in comparison with cosine PSM and traditional models with no PSM. All cosine PSM results are from Lee et al [14].

	Number of similar patients at best predictive performance				Best predictive performance, mean (95% CI)					
	AUROC[a]		AUPRC[b]		AUROC			AUPRC		
	RF[c] PSM[d]	Cosine PSM	RF PSM	Cosine PSM	RF PSM	Cosine PSM	No PSM	RF PSM	Cosine PSM	No PSM
DC[e]	260	100	230	60	0.801 (0.792-0.811)	0.797 (0.791-0.803)	0.693 (0.679-0.707)	0.429 (0.409-0.449)	0.393 (0.378-0.407)	0.280 (0.263-0.297)
LR[f]	5000	6000	9000	6000	0.824 (0.815-0.832)	0.830 (0.825-0.836)	0.811 (0.799-0.821)	0.460 (0.437-0.484)	0.474 (0.460-0.488)	0.449 (0.430-0.468)
DT[g]	5000	2000	7000	4000	0.779 (0.775-0.784)	0.753 (0.742-0.764)	0.721 (0.705-0.738)	0.352 (0.337-0.367)	0.347 (0.335-0.358)	0.339 (0.324-0.353)
RF	15000	—	4000	—	0.839 (0.835-0.844)	—	0.839 (0.835-0.844)	0.507 (0.527-0.486)	—	0.505 (0.487-0.523)
CSRF[h]	—	—	—	—	0.832 (0.821-0.843)	—	—	0.496 (0.520-0.471)	—	—

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

[c]RF: random forest.

[d]PSM: patient similarity metric.

[e]DC: death counting.

[f]LR: logistic regression.

[g]DT: decision tree.

[h]CSRF: case-specific random forest.

XSL•FO

RenderX

Figure 2. Mortality prediction performance measured in area under the receiver operating characteristic curve as a function of the number of similar patients. Mean and 95% confidence interval from 10-fold cross-validation are shown.
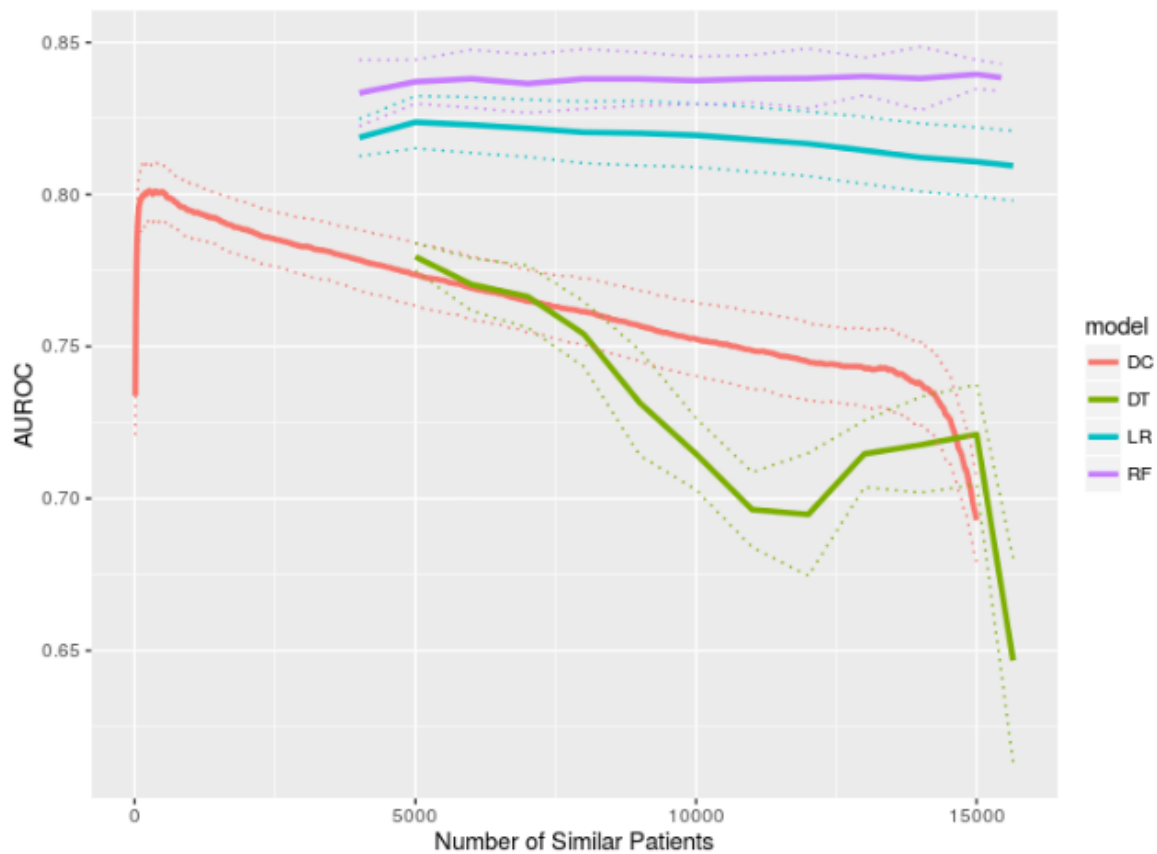

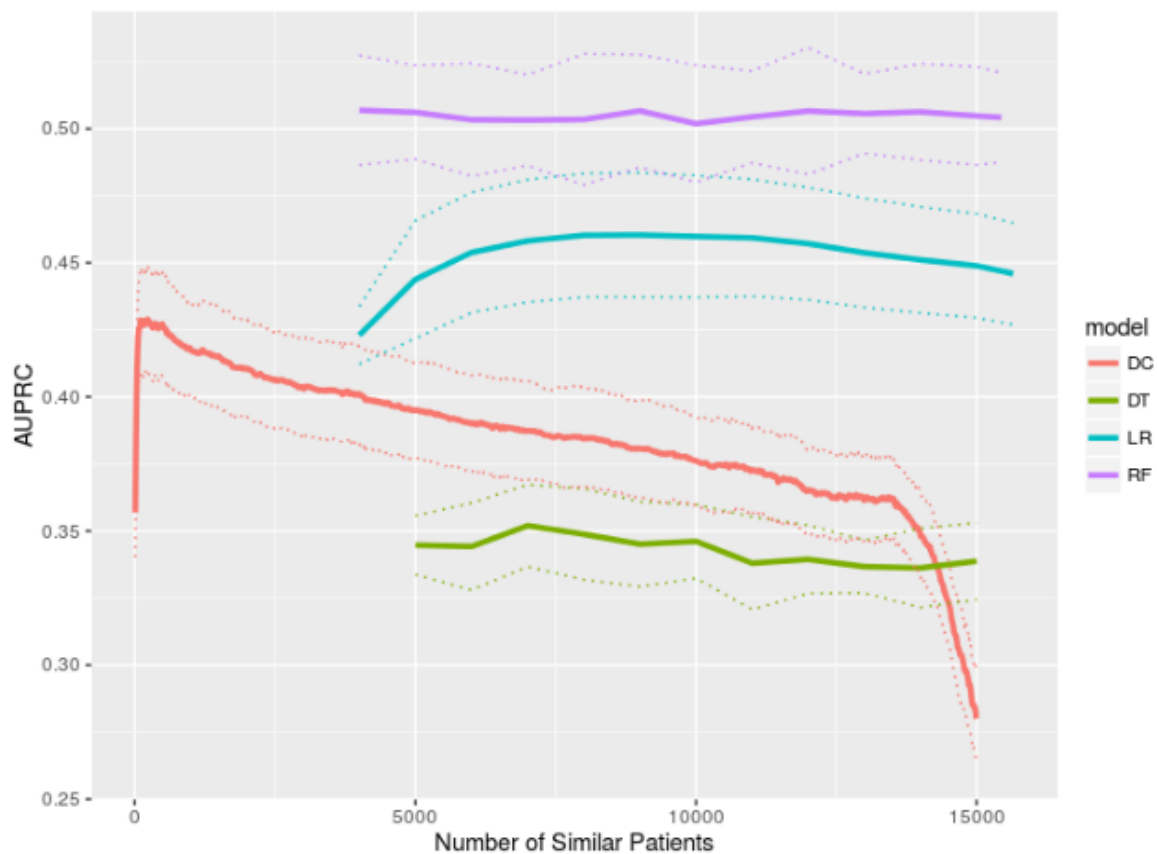**Figure 3.** Mortality prediction performance measured in area under the precision-recall curve as a function of the number of similar patients. Mean and 95% confidence interval from 10-fold cross-validation are shown.

Figure 4. Box plot comparing area under the receiver operating characteristic curves (AUROCs) from all 5 models. For death counting, logistic regression, decision tree, and random forest, the performance from the number of similar patients corresponding to the maximum mean AUROC is shown.
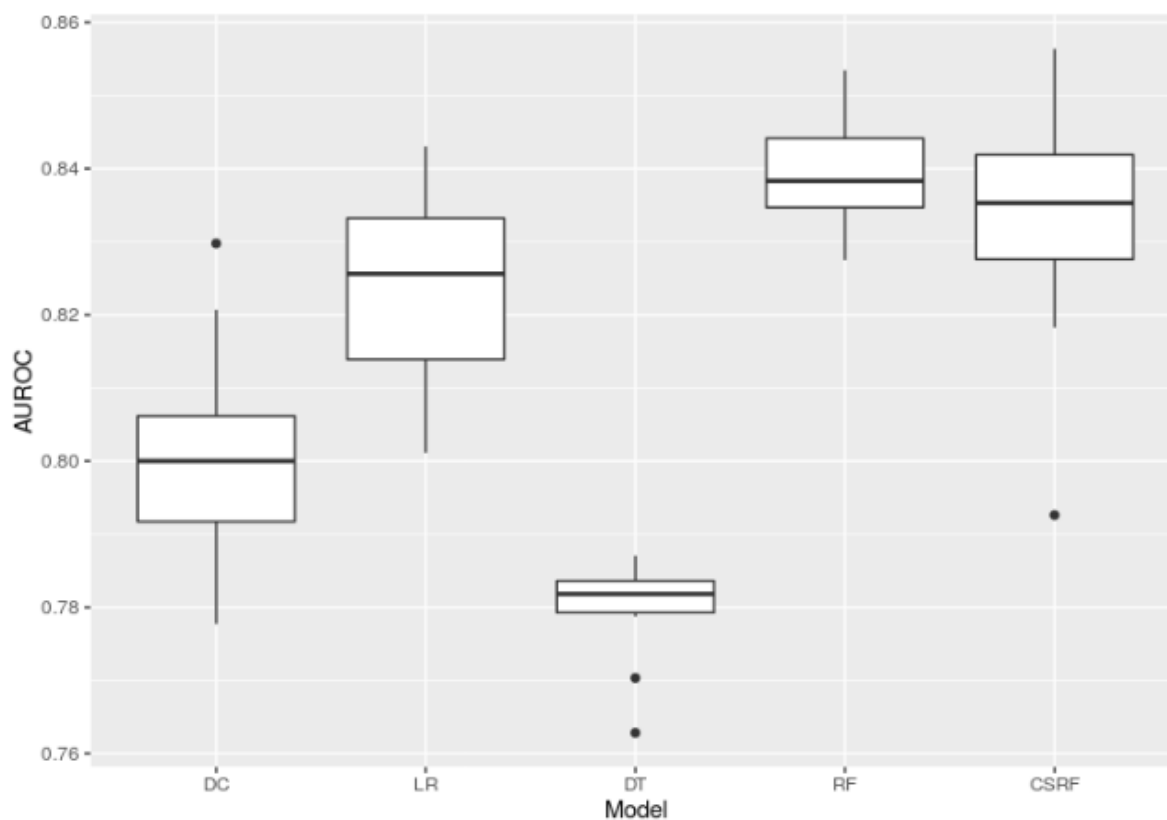
Figure 5. Box plot comparing area under the precision-recall curves (AUPRCs) from all 5 models. For death counting, logistic regression, decision tree, and random forest, the performance from the number of similar patients corresponding to the maximum mean AUPRC is shown.
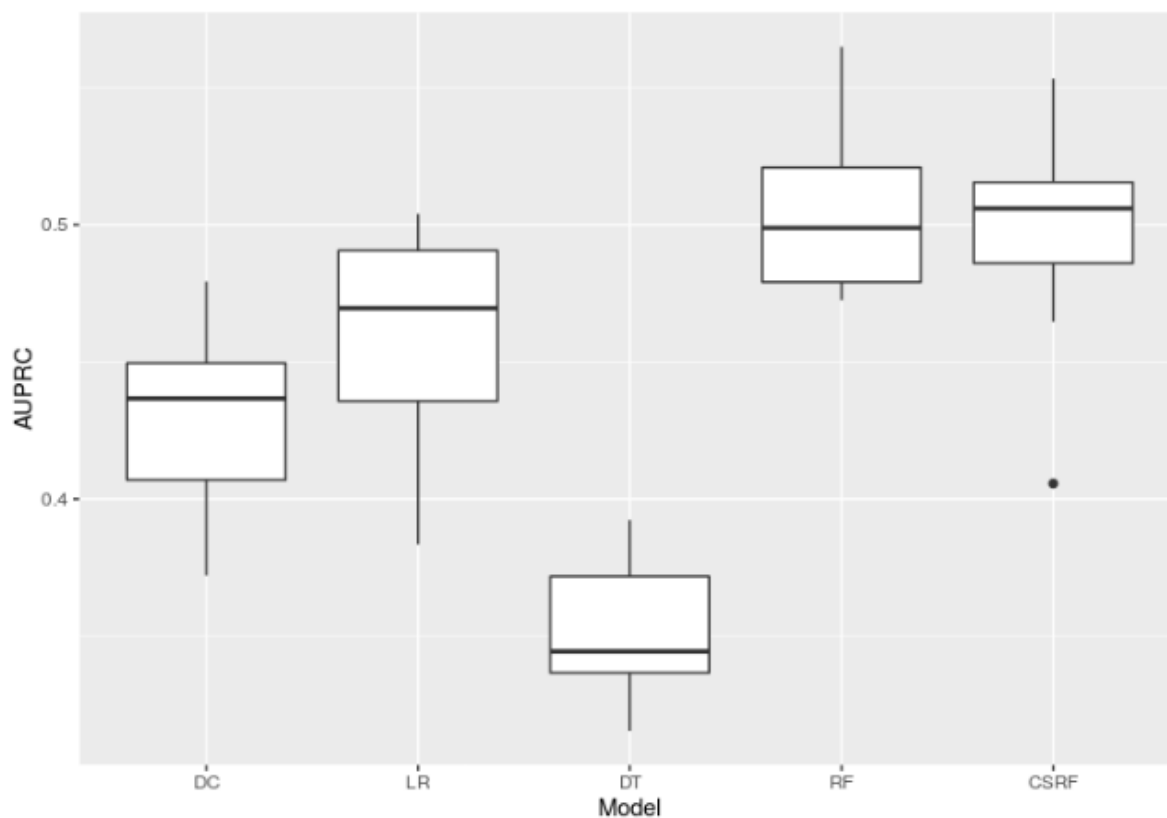
Discussion

Principal Findings

The conventional doctrine in machine learning is that it is always beneficial to collect more training data. This is true if collected data represent the same underlying phenomenon, but it is often difficult to make this assumption in medicine largely due to enormous variability in patient/clinical characteristics, as well as our limited understanding of complex human health and disease pathways. In the era of big data, we can now afford to be more selective regarding which cases should be included in predictive modeling. Data-driven patient similarity matching, via a PSM, leads to objective training data selection and uses hidden patterns in multidimensional data that are difficult for human clinicians to identify.

In comparison with conventional, one-size-fits-all predictive models such as those from the Framingham Heart Study [22], patient-specific predictive models improve predictive performance at the cost of increased computational burden associated with computation of all pairwise PSM values and training of a unique custom model for each patient. This is a reasonable trade-off today given that powerful computing is available at ever falling prices. With big data analytics leveraging parallel computing, it is feasible to train and use patient-specific models in real time at the point of care.

In this study, RF and CSRF, which were not studied in our previous work, outperformed DC, LR, and DT, while the difference between RF and CSRF was statistically insignificant. The comparison between RF with CSRF is interesting because the difference is essentially hard thresholding versus weighted bootstrapping. Also, CSRF did not improve upon the conventional RF that used all data as training data. The comparable performances from RF and CSRF, and the negligible effects of the number of similar patients on RF performance (Figures 2 and 3), imply that not all predictive models benefit from the use of a PSM. This could be a characteristic of ensemble models but needs further research to clarify.

An appropriate method to circumvent the lack of variability in categorical variables when M is too small should be investigated in future work. One simple solution is to exclude such problematic categorical variables in model training after "using them up" for similarity matching. For example, gender may be used for PSM calculation but if the training data subsequently only include 1 gender (same as that of the index patient), then gender can be dropped from model training to avoid the computational issue. This solution will not work for the outcome variable, however. This issue is closely related to the important topic of how best to use the variables in a given data set in personalized predictive analytics: should they be used as predictor or similarity variables, or both? This topic requires further research involving both real life and simulated data.

Although this study only investigated prediction performance as a function of M without attempting to select optimal M values for individual patients, optimal M selection will become important when the results of this study are taken to practice. However, in the context of patient similarity, selecting the optimal M for a given patient is not trivial because it may vary across different types of patient. This implies that optimal M values should not be selected based on Figures 2 and 3 because they represent the cohort as a whole and individual patients may not follow the global patterns shown in Figures 2 and 3. For this reason, a subset of similar patients would have to be compiled first (assuming that they are sufficiently similar to yield similar optimal M values), and then this subset would have to be partitioned into training, validation, and test data so that the validation data can be used for M selection. The main challenge with this approach is that insufficient sample size is likely to occur, especially given that there were 75 predictor variables in this study. Furthermore, it is difficult to determine how many similar patients should be included in the subset and how similar they need to be to the patient under consideration.

Instead of selecting the M most similar patients from training data, it is also feasible to threshold RF PSM values so that all patients with an RF PSM above this threshold are included in model training. One advantage of this approach is that the quality of similarity in the training data can be ensured and controlled, whereas a fixed M would force M patients to be used for model training regardless of how similar they are to the index patient in absolute sense. However, a challenge with this approach is that a good understanding of the magnitude of the RF PSM is required, which could be the subject of another research study. Moreover, the threshold is likely to vary across patients.

In addition to the future work mentioned above, future PSM research directions should involve the following: other PSMs especially those that can capture temporal patterns, other health data sets, patient outcomes other than mortality, other predictor variables (such as diagnosis, medications other than vasopressors, information from free-text clinician notes, etc), and other predictive models.

Comparison With Prior Work

In comparison with the DC, LR, and DT results from our previous work that studied a cosine similarity PSM [14], the predictive performance patterns as a function of the number of similar patients were similar in this study. In terms of the best performance of each model, 2 performances were better than the corresponding performances in our previous work (ie, no overlap between the 95% confidence intervals): DC and DT in terms of AUPRC and AUROC, respectively. The rest of the reported best performances did not show any significant difference between the 2 studies. Despite the modest performance improvement, these results indicate that the RF PSM outperformed the cosine similarity PSM.

However, the number of similar patients associated with the best performance was different between the 2 studies. The RF PSM tended to require more similar patients to be included in the training data to achieve peak performance; the cosine similarity PSM results peaked with respect to AUROC and AUPRC at 100 and 60 for DC, 6000 and 6000 for LR, and 2000 and 4000 for DT. This finding suggests that the RF and cosine similarity PSMs quantified patient similarity somewhat differently.

Limitations

This study has limitations. First, MIMIC-II is a single-center database and hence the results may not be generalizable to other ICU data. Second, no extensive investigation of various parameter values was conducted for the predictive models in order to keep the number of models that had to be trained at a reasonable level. Although the default parameter settings from the widely used R packages are reasonable choices, the effects of the parameter values on predictive performance could be investigated further in future work.

Conclusions

PSM-driven predictive analytics is an exciting topic, as accurate, tailor-made patient outcome prediction can greatly inform risk stratification and resource allocation. The results from this study that stem from the use of an RF PSM corroborate the utility of PSMs in enhancing predictive performance at the patient level. The superior predictive performances of RF and CSRF, as well as the fact that the RF PSM outperformed the cosine similarity PSM in some models, indicate that RFs are well suited for patient-specific predictive modeling.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Darcy A, Louie A, Roberts L. Machine learning and the profession of medicine. JAMA 2016 Feb 09;315(6):551-552. [doi: 10.1001/jama.2015.18421] [Medline: 26864406]
2. Bates D, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data In health care: using analytics to identify and manage high-risk and high-cost patients. Health Affairs 2014 Jul 08;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041]
3. Andreu-Perez J, Poon C, Merrifield R, Wong S, Yang G. Big data for health. IEEE J Biomed Heal informatics 2015;19(4):1208.
4. Joyner M, Paneth N. Seven questions for personalized medicine. JAMA 2015 Sep 08;314(10):999-1000. [doi: 10.1001/jama.2015.7725] [Medline: 26098474]
5. Celi L, Marshall J, Lai Y, Stone D. Disrupting electronic health records systems: the next generation. JMIR Med Inform 2015 Oct 23;3(4):e34 [FREE Full text] [doi: 10.2196/medinform.4192] [Medline: 26500106]
6. Park YJ, Kim BC, Chun SH. New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis. Expert Systems 2006 Feb;23(1):2-20. [doi: 10.1111/j.1468-0394.2006.00321.x]
7. Ng K, Sun J, Hu J, Wang F. Personalized predictive modeling and risk factor identification using patient similarity. 2015 Presented at: AMIA Summits Transl Sci Proc; 2015; San Francisco.
8. Panahiazar M, Taslimitehrani V, Pereira NL, Pathak J. Using EHRs for heart failure therapy recommendation using multidimensional patient similarity analytics. Stud Health Technol Inform 2015;210:369-373 [FREE Full text] [Medline: 25991168]
9. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol 2006 Jun 15;163(12):1149-1156 [FREE Full text] [doi: 10.1093/aje/kwj149] [Medline: 16624967]
10. Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput 2003 Jan;7(1):76-80. [doi: 10.1109/MIC.2003.1167344]
11. Liu F, Du P, Weng F, Qu J. Use clustering to improve neural network in financial time series prediction. 2007 Presented at: Proc Third Int Conf Natural Computation; 2007; Xian.
12. Trivedi S, Pardos Z, Heffernan N. The utility of clustering in prediction tasks. arXiv Prepr 2015.
13. McLachlan G, Peel D, Peel D. Finite mixture models. New York: Wiley; 2000.
14. Lee J, Maslove DM, Dubin JA. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. PLoS One 2015;10(5):e0127428 [FREE Full text] [doi: 10.1371/journal.pone.0127428] [Medline: 25978419]
15. Lee J. Personalized mortality prediction for the critically ill using a patient similarity metric and bagging. 2016 Presented at: Proc IEEE-EMBS Int Conf Biomed Heal Informatics; 2016; Las Vegas p. 332-335.
16. Breiman L. Random forests. Mach Learn 2001;45:32. [doi: 10.1023/A:1010933404324]
17. Xu R, Nettleton D, Nordman D. Case-Specific Random Forests. J Comput Graph Stat 2016 Mar 09;25(1):49-65. [doi: 10.1080/10618600.2014.983641]
18. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L, Moody G, et al. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. Crit Care Med 2011 May;39(5):952-960 [FREE Full text] [doi: 10.1097/CCM.0b013e31820a92c6] [Medline: 21283005]

19.   Lee J, Scott D, Villarroel M, Clifford G, Saeed M, Mark R. Open-access MIMIC-II database for intensive care research. 2011 Presented at: Proc Annu Int Conf IEEE Eng Med Biol Soc; 2011; Milan p. 8315-8318.

20.   Shi T, Horvath S. Unsupervised learning With random forest predictors. J Comput Graph Stat 2006 Mar;15(1):118-138. [doi: 10.1198/106186006X94072]

21.   Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2(3):18-22.

22.   Mahmood S, Levy D, Vasan R, Wang T. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014 Mar 15;383(9921):999-1008 [FREE Full text] [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]

Abbreviations

AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
CCU: coronary care unit
CSRF: case-specific random forest
CSRU: cardiac surgery recovery unit
DC: death counting
DT: decision tree
EHR: electronic health record
ICU: intensive care unit
LR: logistic regression
MICU: medical intensive care unit
MIMIC II: Multiparameter Intelligent Monitoring in Intensive Care II
PSM: patient similarity metric
RF: random forest
SICU: surgical intensive care unit

XSL•FO
RenderX