JMIR Medical Informatics

Impact Factor (2023): 3.1
Volume 10 (2022), Issue 1 ISSN 2291-9694 Editor in Chief: Christian Lovis, MD, MPH, FACMI

Contents

Reviews

Viewpoints

Original Papers

Corrigenda and Addenda

Review

Machine Learning for Cardiovascular Outcomes From Wearable Data: Systematic Review From a Technology Readiness Level Point of View

Arman Naseri Jahfari[1,2], MSc; David Tax[1], PhD; Marcel Reinders[1], PhD; Ivo van der Bilt[2], MD, PhD

[1]Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, Netherlands

[2]Department of Cardiology, Haga Teaching Hospital, The Hague, Netherlands

Corresponding Author:
Arman Naseri Jahfari, MSc
Pattern Recognition and Bioinformatics
Delft University of Technology
van Mourik Broekmanweg 6
Delft, 2628 XE
Netherlands
Phone: 31 152786052
Email: a.naserijahfari@tudelft.nl

Abstract

Background: Wearable technology has the potential to improve cardiovascular health monitoring by using machine learning. Such technology enables remote health monitoring and allows for the diagnosis and prevention of cardiovascular diseases. In addition to the detection of cardiovascular disease, it can exclude this diagnosis in symptomatic patients, thereby preventing unnecessary hospital visits. In addition, early warning systems can aid cardiologists in timely treatment and prevention.

Objective: This study aims to systematically assess the literature on detecting and predicting outcomes of patients with cardiovascular diseases by using machine learning with data obtained from wearables to gain insights into the current state, challenges, and limitations of this technology.

Methods: We searched PubMed, Scopus, and IEEE Xplore on September 26, 2020, with no restrictions on the publication date and by using keywords such as “wearables,” “machine learning,” and “cardiovascular disease.” Methodologies were categorized and analyzed according to machine learning–based technology readiness levels (TRLs), which score studies on their potential to be deployed in an operational setting from 1 to 9 (most ready).

Results: After the removal of duplicates, application of exclusion criteria, and full-text screening, 55 eligible studies were included in the analysis, covering a variety of cardiovascular diseases. We assessed the quality of the included studies and found that none of the studies were integrated into a health care system (TRL<6), prospective phase 2 and phase 3 trials were absent (TRL<7 and 8), and group cross-validation was rarely used. These issues limited these studies’ ability to demonstrate the effectiveness of their methodologies. Furthermore, there seemed to be no agreement on the sample size needed to train these studies’ models, the size of the observation window used to make predictions, how long participants should be observed, and the type of machine learning model that is suitable for predicting cardiovascular outcomes.

Conclusions: Although current studies show the potential of wearables to monitor cardiovascular events, their deployment as a diagnostic or prognostic cardiovascular clinical tool is hampered by the lack of a realistic data set and proper systematic and prospective evaluation.

(JMIR Med Inform 2022;10(1):e29434) doi:10.2196/29434

KEYWORDS

mHealth; wearable; machine learning; cardiovascular disease; digital health; review; mobile phone

Introduction

Background

The use of diagnostic modalities in cardiovascular disease is often limited to hospital visits. As a result, the clinical value may be limited by the short observation period. This is especially problematic for cardiovascular problems that do not manifest constantly, such as paroxysmal arrhythmias, heart failure, or even chest discomfort that may not be present during the hospital visit. Advancements in eHealth, especially in wearable technology, such as electrocardiograms (ECGs) [1] and photoplethysmograms (PPGs) [2], and subsequent signal processing by machine learning have enabled new opportunities for remote monitoring in the outpatient setting.

Continuous monitoring over long periods has shown to be effective [3,4]. For example, remote monitoring of patients with cardiac diseases, using pacemakers or implantable cardioverter defibrillators and patients with heart failure have improved patient care [5]. However, current sensors used in health care, such as Holter devices, are limited to a maximum of 14 days (but typically endure 24 hours) of continuous monitoring, limiting the use of these devices. Overcoming this could enable early warning systems for acute events such as cardiac arrest and could capture subtle cardiovascular exacerbation or rehabilitation that manifests over a much longer time because of, for example, changes in lifestyle or intervention.

Although widely used, currently 24-hour ECG or blood pressure monitoring devices are cumbersome to wear and impose a burden on patients in a longitudinal setting. Rechargeable, easy-to-wear sensors, such as smartwatches, are becoming an interesting alternative as they contain sensors with a potentially unlimited observation period with minimal burden to the patient for a fraction of the costs. However, the signals that these wearables measure, such as the PPG-derived heart rate, activity, and skin temperature, are not clinically informative enough for clinical decision-making by a cardiologist. With current developments in artificial intelligence (AI), a powerful solution is expected from machine learning algorithms that can learn the relationship between the wearable sensor signals and a cardiovascular outcome in a (fully) data-driven manner.

Another great benefit of automatic cardiovascular diagnostics and prognostics by machine learning is minimizing inter- and intraobserver variability, which is a major problem in the subjective interpretation of clinical and diagnostic information by human cardiologists. Interobserver disagreement [6,7] because of, for example, differences in experience or specialization and intraobserver disagreement because of stress or fatigue [8], can be minimized. Variations in clinical practice may lead to medical errors, whereas automatic systems are not (or less) susceptible to such factors. Another possibility is to exclude patients who experience symptoms such as chest pain, which are not caused by cardiovascular disease. Automatic exclusion of these patients can reduce unnecessary visits to a cardiologist; relieving the cardiologist, thereby increasing the capacity of cardiovascular care; and directing patients to the proper specialist quicker.

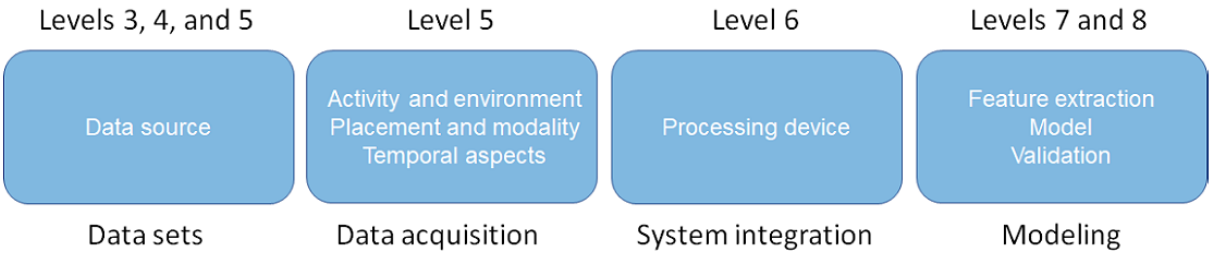
Because of these promises, the field of research on diagnosing cardiovascular events from wearable data is very active and many machine learning solutions are being presented to automatically detect cardiovascular events. Various reviews have been presented to categorize the developed machine learning tools. A study by Krittanawong et al [9] shows that a plethora of wearable devices are researched for a variety of cardiovascular outcomes and discusses a paradigm for remote cardiovascular monitoring consisting of sensors, machine learning diagnosis, data infrastructure, and ethics. They conclude that especially the latter two aspects have several unaddressed challenges. An overview of wearable devices on the market is provided by Bayoumy et al [10]. The study reports their frequency of use in (cardiovascular) trials and Food and Drug Administration status. As reported by Giebel and Gissel [11], most mobile health devices for atrial fibrillation detection are not Food and Drug Administration approved and therefore cannot be used in cardiovascular monitoring systems.

Objectives

Although many machine learning tools have been proposed and studies have shown good performance, they do not seem to have been implemented in operational and functional health care systems. Therefore, we decided to systematically review the machine learning tools to detect cardiovascular events from wearable data from the perspective of their technology readiness level (TRL), that is, how far these proposed tools are in realizing an operational system and what factor is impeding them to get there. The TRL paradigm originates from the National Aeronautics and Space Administration and is a way to assess the maturity level of a particular technology used in space travel by giving solutions a score from 1 to 9 in increasing order of readiness, from basic technology research (score 1) to launch operations (score 9) [12].

Interestingly, 2 studies tailor the TRL framework for medical machine learning. A study by Komorowski [13] proposes a TRL for supervised, unsupervised, and reinforcement learning problems and describes criteria to reach TRLs 3, 4, 6, and 7. A description of the 9 TRLs for medical machine learning in intensive care medicine, including examples, is proposed by Fleuren et al [14]. We review the wearable-based cardiovascular machine learning solutions following the framework by Fleuren et al [14] adjusted for remote medicine. We identify aspects in the studies and systematically assign these to TRLs and group some of the TRLs together in a taxonomy to help interpret their relevance (Figure 1). We address the overuse of benchmark data sets, considerations on data acquisition related to the environment and type of sensor, integration in a health care system, construction of the machine learning model, and subsequent model validations.

Figure 1. Taxonomy of the eligible studies. TRLs are based on the proposed descriptions for machine learning for medical devices proposed by Fleuren et al [14]. The studies were categorized according to the relevance of their content to these descriptions (aspects within boxes) and were grouped and assigned to the different TRLs (below and above boxes). TRL: technology readiness level.



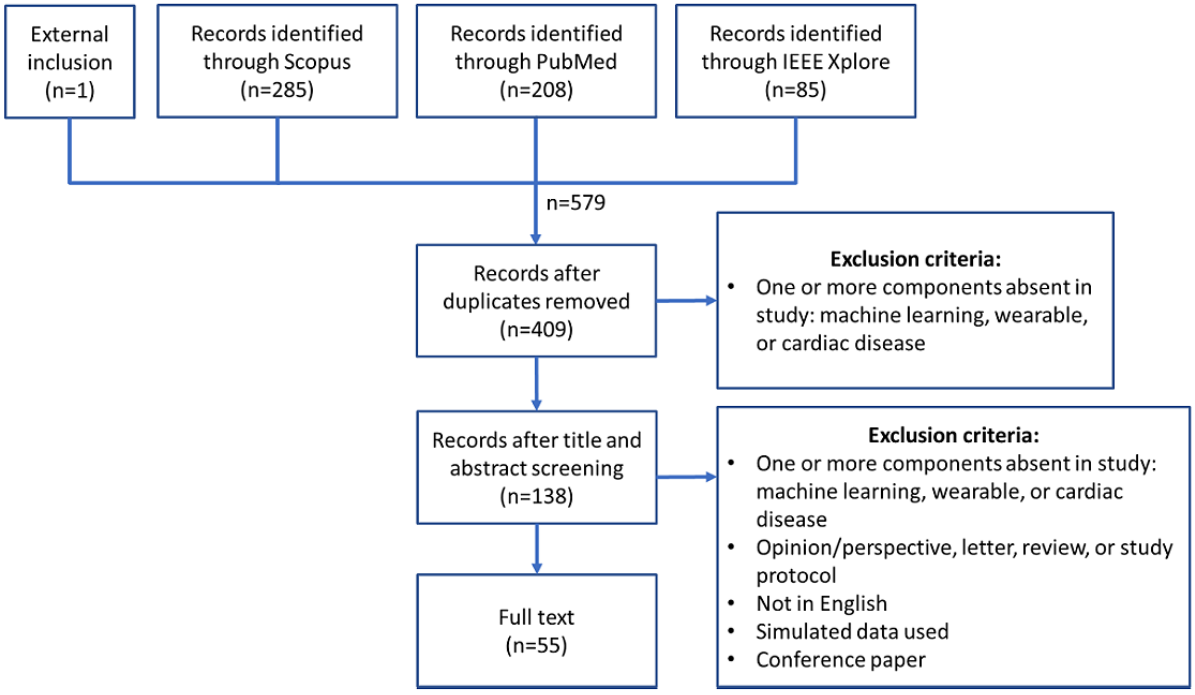
By assessing current methods by their technological readiness, we show that the current methodologies are promising but that deployment is severely hampered by the lack of realistic data sets and proper systematic and prospective evaluation. To arrive at a readiness that is operational at the health care system level, these bottlenecks need to be resolved.

Methods

Screening

The systematic review was performed by following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15], as shown in Figure 2. We followed the patient or population, intervention, comparison, and outcomes framework for our research question, which was as follows: "In patients with cardiovascular disease, using machine learning with data from wearables, what methods and accompanying limitations are used, to deploy this technology to detect and predict cardiovascular disease in standard healthcare?"

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the systematic review.



Study Inclusion

Search queries were performed on September 26, 2020, in the electronic databases Scopus, PubMed, and IEEE Xplore. Only peer-reviewed journals were considered. Studies were eligible for inclusion if data were acquired from wearables, a machine learning method was used, and had the goal to detect or predict cardiovascular disease (see Multimedia Appendix 1 for used queries). The following exclusion criteria were used: opinion or perspective, letter, review, study protocol, or conference paper; studies not in English; and studies in which only simulated data were used. The eligibility assessment was performed by the first author, ANJ. First, the title and abstract of each study were assessed for relevance based on the inclusion and exclusion criteria. The full texts of the remaining studies were then read and again subjected to the selection criteria. The

second author, DT, verified this by reading a subsample of the selection.

TRL and Taxonomy

From the eligible studies through discussions with all authors, the first author, ANJ, identified some general overarching evaluation aspects that the studies had in common and assigned these studies to a taxonomy (Multimedia Appendix 2 [16-70]). These aspects were related to one or more TRLs, as defined by Fleuren et al [14]. Accordingly, the eligible studies were assigned to the taxonomy and different TRLs (Figure 1). The TRL framework states that studies that use only a benchmark data set as a data source do not progress further than level 3. Furthermore, the framework originally grouped levels 3 and 4 together. We split these levels and assigned studies using their own acquired data without an external validation set from a different study level 4. Next, we assigned studies that use an external validation set from a different study to level 5; although, according to Fleuren et al [14], level 5 further requires that the acquired data set is realistic. However, we interpreted the independently acquired data representative of data recorded during the deployment of the machine learning system as realistic. Therefore, we differentiated levels 3, 4, and 5 mostly on the data sets being used for model deployment and related these levels to the data sets taxonomy. As level 5 mainly focuses on realistic data sets we also assigned practical aspects of the wearables to this TRL. Here, we differentiated the following three aspects: (1) which modality is being measured by the wearable and where on the body it is placed; (2) under which conditions data are measured, such as in the wild or in controlled environments; and (3) for how long data are recorded, that is, the temporal aspect of the acquired data. Level 6 required integrating the machine learning model into a health care system. Therefore, the device in which the model was integrated into was assigned to this level. Finally, levels 7 and 8 required demonstrating the model as a cardiovascular tool. Therefore, the model effectiveness and validation aspects were assigned to these levels. Levels 1, 2, and 9 were disregarded here because none of the papers fit into these categories.

Results

Article Identification

A total of 578 records were retrieved from electronic databases. After the removal of duplicates, 70.8% (409/578) of records remained. One was externally included as it fulfilled the inclusion criteria but was missed by the search query because it did not explicitly mention the term machine learning. As shown in Figure 1, these were further narrowed down during title or abstract screening, resulting in 23.9% (138/578) of records. Finally, after full-text reading, 9.5% (55/578) of records remained to be covered in this study.

We related each of the studies to different TRLs for machine learning methods (*Methods*) according to an identified taxonomy of different evaluation criteria that relate to these TRLs (Figure 1; *Methods*). The TRL framework states that studies that use only a benchmark data set do not progress further than level 3.

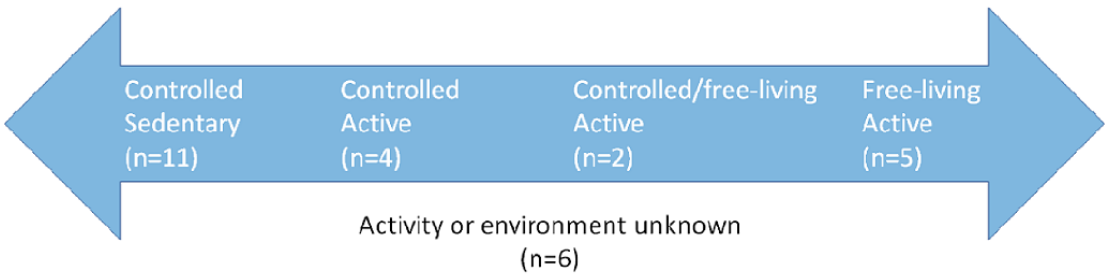
Study Characteristics

The key characteristics of the eligible studies are summarized in Multimedia Appendix 2. Notably, of the 55 studies, 27 (49%) exclusively used benchmark data sets, which were all defined as benchmark studies. Furthermore, of the 55 included studies, 6 (11%) were published before 2018 and the remaining 49 (89%) were published thereafter. In the following sections, the study characteristics are described more closely based on the taxonomy.

Activity and Environment (Level 5)

For studies that did not use benchmark data sets, they reported the data acquired either in a controlled environment (hospital or research laboratory) or in a free-living environment, where participants were remotely observed performing their natural daily routines. The latter is also known as *in-the-wild*. Furthermore, the activities of the participants can be divided into sedentary or active during data acquisition. To capture these two related aspects, we assigned studies on an axis representing a controlled environment and sedentary activity on one side and in-the-wild measurement of active participants on the other side of the axis (Figure 3). Interestingly, only 5 [16-20] studies mapped to the active, free-living situation that complied with the requirement of realistic data acquisition for these aspects that map to TRL5. Thus, only one-tenth of the studies used the potential of wearables to be used for remote, longitudinal monitoring.

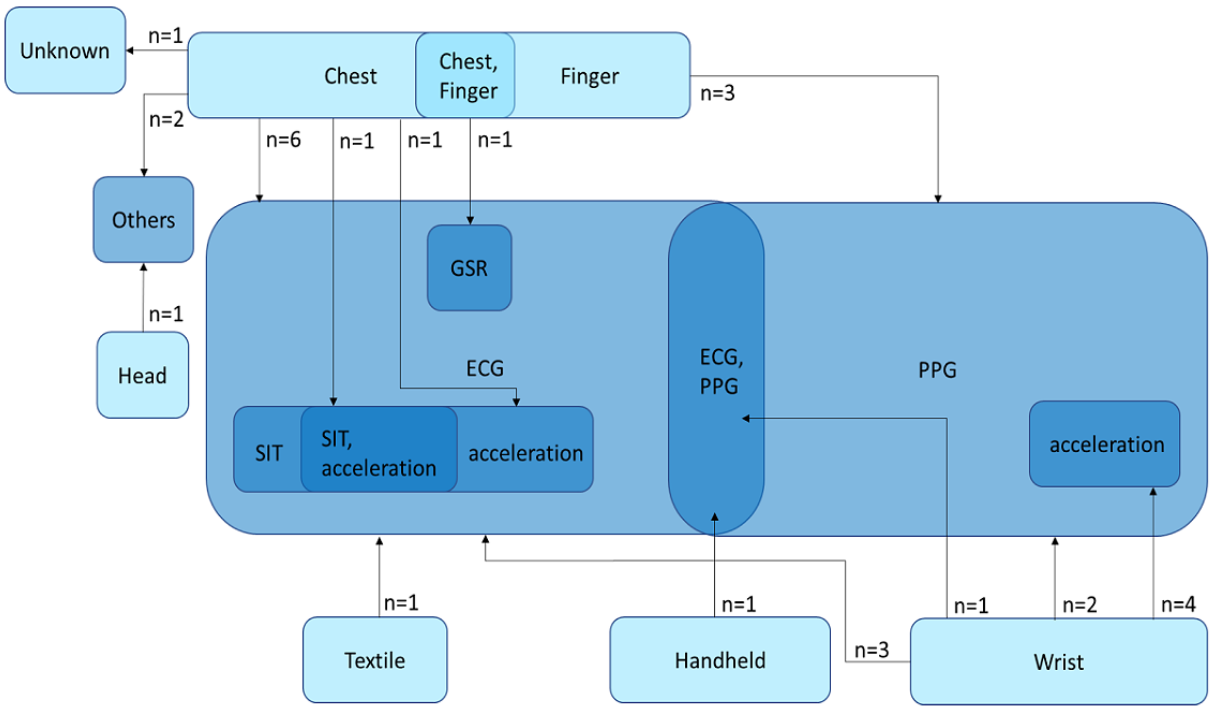
Figure 3. Studies ordered based on participant activity and acquisition environment. The leftmost scenario indicates highly controlled acquisition with sedentary participants. The opposite is described by the rightmost scenario where participants are monitored in an active, free-living situation. Controlled environment includes hospitals or laboratories. Free-living participants are monitored during their daily routines.



Placement and Modality (Level 5)

Realistic data acquisition requires continuous monitoring. Practically, the wearable should therefore not burden the participant when wearing. This burden depended mostly on the placement of the sensor on the body. In addition, the placement also restricted the type of biometric signals that could be measured, which was referred to as the modality. We categorized studies based on the placement and modality for the nonbenchmark studies jointly (Figure 4). The sensor placements for cardiovascular monitoring that results in the least burden for the patient, and thus would be the best candidates to acquire a realistic data set, were the wrist and finger. Less than half (N=13) of the studies were reported with such placements, of which 8 (62%) studies acquired one modality: 3 (23%) studies acquired wrist-based ECGs [18,21,22], 2 (15%) studies acquired wrist-based PPGs [17,23], and 3 (23%) studies acquired finger-based PPGs [24,30,37]. Of the 13 studies, the remaining 5 (39%) studies acquired wrist-based multimodal data: 4 (31%) studies acquired PPGs and accelerometer data [19,20,29,47] and 1 (8%) study acquired both ECGs and PPGs [25]. Thus, the wrist and finger severely limited the additional modalities that were measured (usually only acceleration), although wearables were shown to be able to measure increasing number of modalities [10].

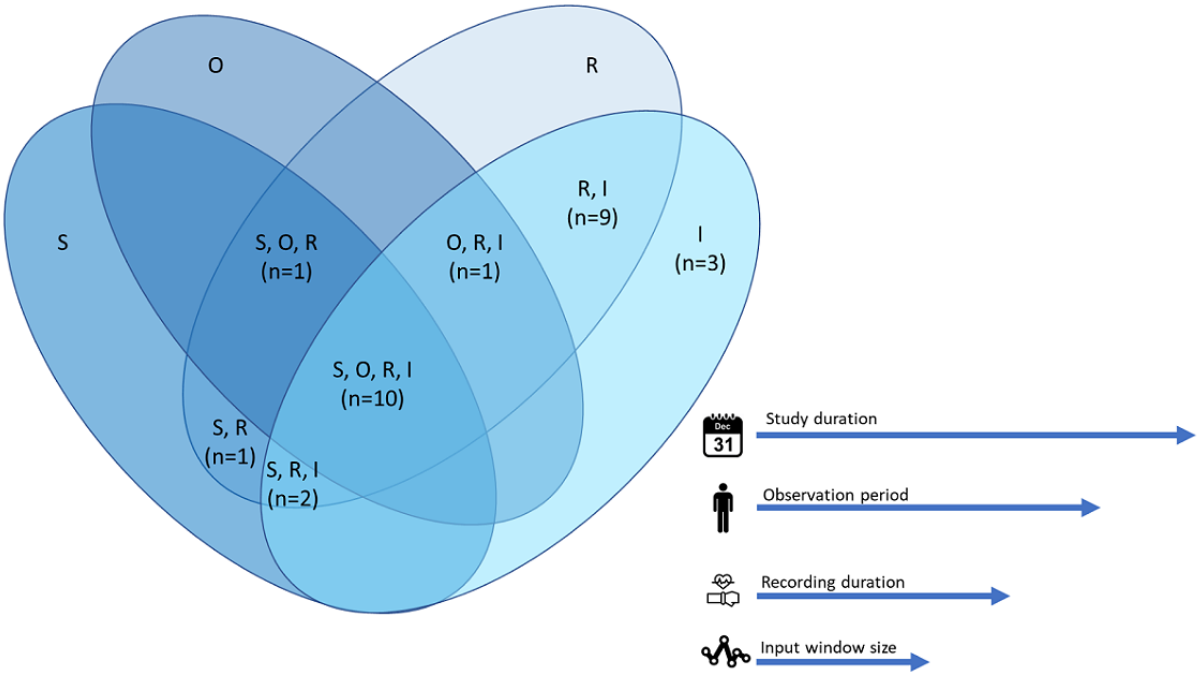
Figure 4. Placement and modalities of wearable sensors: light blue, placement of sensors; blue, modalities used. Others: head, near-infrared spectroscopy; chest, seismocardiography or gyrocardiography. Overlapping blocks represent multiple placements or modalities used. ECG: electrocardiogram; GSR: galvanic skin response; PPG: photoplethysmogram; SIT: skin impedance and temperature.



Temporal Aspects (Levels 5, 7, and 8)

Besides the requirement of a realistic data set in level 5, levels 7 and 8 required phase 2 and phase 3 studies, respectively. In the context of drug testing, this requires an investigation of the effective, but safe, drug dosage. Analogously, for wearable machine learning, this translated to the time a participant must be exposed to a machine learning model before a cardiovascular outcome could be accurately detected or predicted. Therefore, a realistic deployment setting is dependent on the length of time participants are observed. As it is further essential to characterize the data for reproducibility and the description under which circumstances a model is valid, we decided to outline the temporal aspect of the acquired wearable data in more detail. We recognized the following four levels of time aspects: (1) study duration, (2) observation period, (3) recording duration, and (4) input window size (Figure 5). Within the study duration, patients were included and observed for a certain period—the observation period. The lengths of these periods had an impact on the realistic deployment of a system. For example, Quer et al [71] used wrist-worn Fitbit devices to show that resting heart rate within individuals had a significant seasonal trend in longitudinal data. Therefore, a model constructed using data from a certain period might not be valid for another period. It was therefore important to consider how long the participants were observed to ensure this seasonal effect was incorporated in the model. Within the observation period, the wearable recorded a time series. Theoretically, this could be as long as the observation period itself. However, patients could interrupt the measurements for several reasons (eg, to charge the device and low compliance rate). We denoted the duration of a continuously measured part of the time series as the recording duration. Finally, the records were further segmented into windows, from which features were generated or which were used as raw inputs to a machine learning model. We referred to the duration of these windows as the input window size (I).

Figure 5. Venn diagram of reported temporal aspects described in the studies. The S, O, R, and I are represented in the legend. I: input window size; O: observation period; R: recording duration; S: study duration.

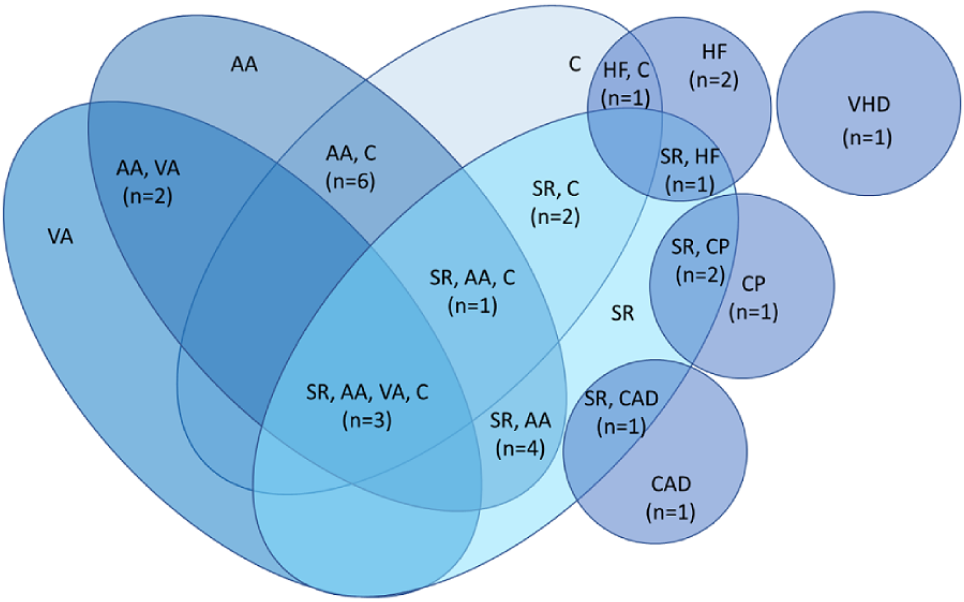


We assessed the temporal aspects of all the nonbenchmark studies (Figure 5). One study did not report any aspects [26] and was omitted from the Figure 5. Another study used multiple fixed input window sizes to incorporate different timescales of the data [19]. Overall, most studies did not report all the aspects and were thus not comprehensive about their data characteristics. In almost all studies, the recording rate and input window size were reported, whereas the study and observation periods were mentioned in about half of the studies. For a realistic data set, required for level 5 and progression to level 7 or 8, the observation period and recording duration were specifically important, as we found in 12 studies. Three studies used an observation period of 24 hours [23,32,64]; one for a week [17], one for 2 weeks [27], and one for 90 days [16]. Overall, 2 studies implied an observation period of months but did not explicitly report it [19,20]. One considered recordings of at least eight hours [19] and one reported an average recording duration of 11.3 hours [20]. Finally, only one [27] fully used the potential of wearables and reported a (near-) continuous recording duration.

Cardiovascular Outcomes (All Levels)

Although the required observation period and recording duration to detect or predict a cardiovascular outcome is still an open and active research topic, these periods will be different for different outcomes. Therefore, we inventoried which (combinations of) cardiovascular outcomes were considered in which studies (Figure 6). Interestingly, the control group was defined differently in each study. Only half of the nonbenchmark studies included a (normal) sinus rhythm class as control and could therefore exclude the presence of cardiovascular disease in participants. From these, 8 studies [17,21-23,28-31] used data from healthy individuals to represent normal sinus rhythm. The remaining 6 studies [32-37] derived normal sinus rhythm data from patients with arrhythmia (such as paroxysmal atrial fibrillation) or were unclear about the control group. Three studies had cardiovascular (disease) prevention as the target. One of these described this as a cardiovascular risk assessment where the predicted classes were healthy, precaution, and critical status [28]. Another study predicted vascular age and 10-year cardiovascular disease risk [34]. The third assigned a cardiorespiratory fitness score [27]. Notably, only the first 2 studies constructed a prognostic model. Two other prognostic models forecast cardiac arrest and heart failure exacerbation by forecasting rehospitalization after heart failure admission [16,21].

Figure 6. Studies categorized according to the type of cardiovascular outcomes predicted by the models. AA: atrial arrhythmia; C: control; CAD: coronary artery disease; CP: cardiovascular prevention; HF: heart failure; SR: sinus rhythm; VA: ventricular arrhythmia; VHD: valvular heart disease.



Bottleneck TRL5

Although many cardiovascular outcomes were investigated with wearables, the promising studies that have reached level 5 were all focused on atrial arrhythmia using wrist-based PPGs. However, their temporal properties were often inconclusive, as they were not reported. Moreover, to progress to level 6, a model should be functional within a health care system (even if it was merely used observationally). None of the studies progressed to this level. An overview of the level 5 models, including the modalities that they are based on, is given in Table 1. Although none of the methodologies progressed to level 6, we decided to prospectively evaluate the studies to investigate the progression of the current state.

Table 1. Studies fulfilling requirements for technology readiness level 5.

Study	Outcomes	Modality	O[a]	R[b]	I[c]
Torres-Soto and Ashley [17]	Sinus rhythm, atrial arrhythmia	PPG[d]	1 week	NR[e]	25 seconds
Bashar et al [18]	Atrial arrhythmia, ventricular arrhythmia	ECG[f]	NR	NR	2 minutes
Tison et al [19]	Atrial arrhythmia, control	PPG, accelerometer[g]	NR	>8 hours a day	5 seconds, 30 seconds, 5 minutes, and 30 minutes
Wasserlauf et al [20]	Atrial arrhythmia, control	PPG, accelerometer	NR	11.3 hours a day	1 hour

[a]O: observation period.

[b]R: recording duration.

[c]I: input window size.

[d]PPG: photoplethysmogram.

[e]NR: not reported.

[f]ECG: electrocardiogram.

[g]Sensor-provided heart rate and step counter data.

Processing Device (Level 6)

Integration in a health care system could be carried out on different devices. These studies demonstrated their models on either a computer (eg, a server), smartphone, or embedded device (Table 2). Only the latter two enabled real-time cardiovascular monitoring locally on the patient side, required for real-time detection and prevention of acute cardiovascular disease, as real-time information exchange to an external system would require high battery consumption and was therefore not feasible. Smartphones were used in both benchmark [38-40] and nonbenchmark [21,30,31,35] studies. Embedded devices, however, had only been demonstrated in benchmark studies [41-44].

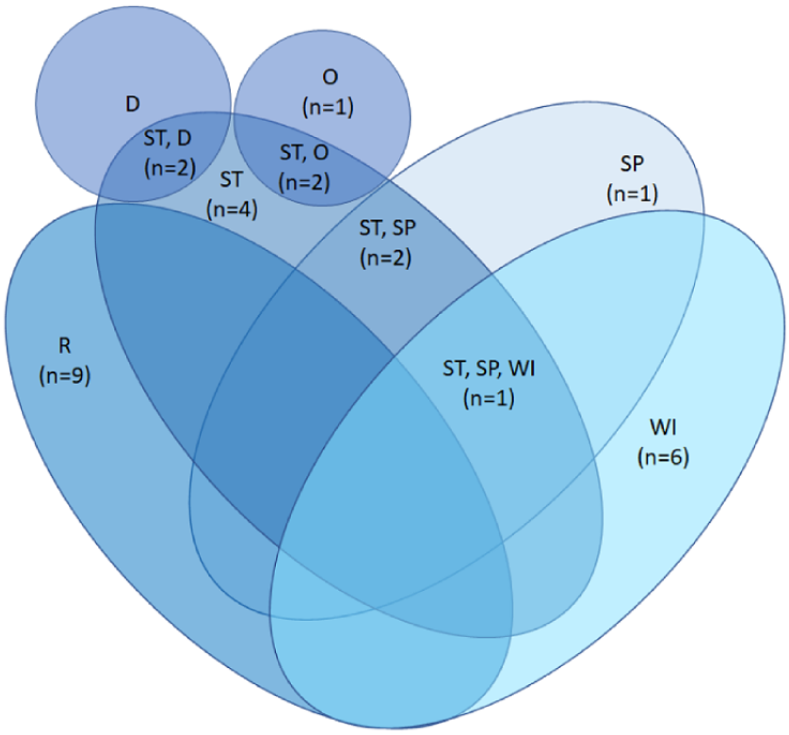
Table 2. Processing device of trained models used in studies.

Processing device	Benchmarks included, n	Benchmarks excluded, n
Computer	44	24
Smartphone	7	4
Embedded device	4	0

Feature Extraction Methods (Levels 7 and 8)

Levels 7 and 8 of the TRL assessed the model effectiveness through phases 2 and 3 clinical trials. We translated that to what features from the observed modalities were being used to construct the models. A significant number of studies used ECG as a modality and used different information from fiducial points [72] to extract features (Figure 7). In many studies, samples were selected before and after the R-peak. For example, the RR interval is the time interval between 2 adjacent R-peaks. Some studies also used techniques to locate other fiducial points and used the time interval between them as features [45]. Together, we denoted these types of features as waveform information features. Next to the specific ECG features, more general features could be derived, such as statistical features (eg, heart rate [variability] derived from 10 RR intervals) or spectral features obtained through techniques such as the Fourier transform. Raw data could also be used as features upon which a neural network can be used to automatically learn informative features [46]. Next to the features based on the sensed signal, demographic information could be used to provide more context [28,47]. Benchmark studies mostly use raw features (using the same data set) and were, therefore, excluded from this study. However, it is noteworthy that 2 of these used more advanced methods, namely, compressed learning [48] combined with dynamic time warping [49].

Figure 7. Features used in the studies. D: demographic; O: others; R: raw; SP: spectral; ST: statistical; WI: waveform information.



The most commonly used features were raw features (studies: 9/28, 32.1%). This was followed by waveform information and statistical features. In all, 2 studies also included demographic metadata from participants [28,47]. One study used hemoglobin parameters [26], which we represented in the *others* group in Figure 7. Interestingly, 1 study included timestamps [19]. From the 11 studies that used multimodal data (Figure 4), 6 (55%) studies extract features for each modality were separately extracted. Of the 11 studies, the remaining 5 (46%) studies exploited the covariance among the modalities in feature extraction, although 1 (9%) study did not elaborate on the exact method [16]. For example, of the 15 studies, 1 (9%) study computed the time between an R-peak in the ECG and the closest following peak in the PPG [34]. Of the 5 studies, 2 (40%) studies concatenated windows of the different modalities and then extracted the features [20,50] and 1 (20%) study concatenated windows whereafter a convolutional layer in a neural network is used to automatically extract features from the concatenated data [19].

Model Construction Methods (Levels 7 and 8)

Another aspect that defines the model effectiveness relates to the type of models being constructed, which we categorized across both the benchmark and nonbenchmark studies (Table 3). Most of the studies used a neural network, and most of them were nonsequential (eg, convolutional and multilayer perceptron). A noteworthy type is the spiking neural network [51,52], which is designed to be energy efficient and suitable for real-time cardiovascular monitoring in an embedded device. Although sequential models were specifically designed for sequence or time series, these types of models were used much less. Some studies had combined sequential and nonsequential neural network architectures [17,19,32,42,46,53]. After the neural networks, most of the models were classical machine learning methods, including linear models: support vector machines; decision trees; and similarity-based models, such as k-nearest neighbor classifiers. Furthermore, ensemble methods had been used that combined multiple simpler models to construct a more complex model [22,28,44,50,54-56]. Finally, 2 studies used models that explicitly exploit the hierarchical structure of medical time series data: a hierarchical Bayesian model [27] and a Multiple-Instance Learning via Embedded instance Selection model [23].

Table 3. Types of machine learning models used in the studies.

Model type	Number of times used
Nonsequential	30
Classical	20
Ensemble	9
Sequential neural network	6
Nonsequential + sequential neural network	5
Hierarchical	2

Validation (Levels 7 and 8)

The effectiveness of a model was heavily influenced by the number of samples with which the model had been trained. In phase 2 and phase 3 studies, a priori power analyses were performed to estimate the required sample size per group or class to observe an effect. It was empirically shown by Quintana [73] that for heart rate variability studies, an effect size of 0.25, 0.5, and 0.9 corresponded to a low, medium, and high effect, respectively. The corresponding sample sizes were 233, 61, and 21 for 80% statistical power and 312, 82, and 28 for a 90% statistical power. We considered nonbenchmark studies with a sufficient sample size per group or class, from which 9 studies remained. From the remaining 9 studies, a power of 90% was reached with small [19,20,24] and large [16,30,37,47] effect sizes, and 2 studies [29,32] achieved 80% power with a large effect size.

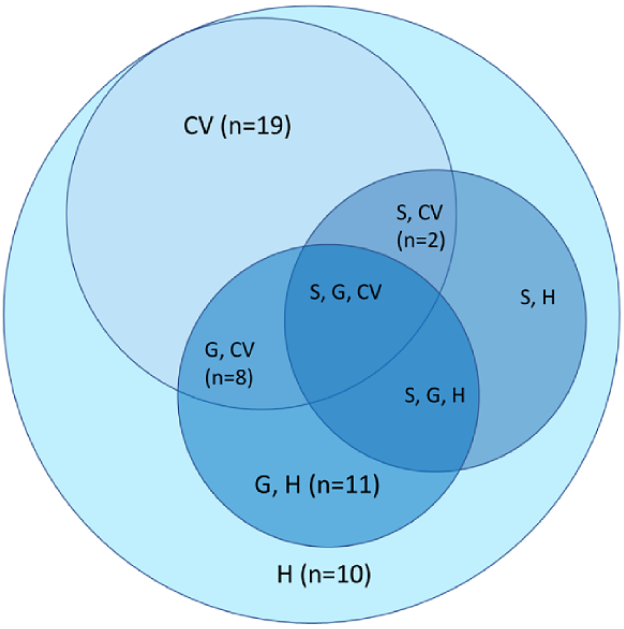
This showed that studies generally choose a train sample size (per group or class) that is too small to find a significant effect based on a priori power analysis.

In contrast to a priori power analysis, the purpose of model validation is to retrospectively analyze the performance of the model on data it has not seen before, that is, to assess the generalization error of the model. The included studies chose from 2 validation schemes: cross-validation and holdout [74] (Figure 8), although 5 studies [16,20,28,64,65] did not report the validation method. When splitting data into training and testing, one needed to ensure nonoverlapping grouping and stratification of the data (Figure 8). With nonoverlapping grouping [75], one ensured that the same group of data did not appear in both the training and test sets, for example, avoiding that data from the same participant was in both the training and test set, albeit the samples might be from different periods. With stratification, one ensured that both the training samples and the test samples exhibit a similar proportion of samples for an arbitrary variable. For example, it was important to keep the proportion of men and women consistent or to ensure that the proportion of sensor samples representing normal rhythm and arrhythmia is equal. For progressing to TRL 7, 4 studies used leave-one-subject-out group cross-validation [18,23,27,45] and 4 other types of group cross-validation [29,30,37,44]. Ideally, a stratified group cross-validation is used, but none of the studies used this. In addition to validation strategies, it is important to use replication data, that is, completely independently acquired data, which was only done in 11 [17,18,21,24,25,31,33,35,36,40,70] studies.

It is important to realize that data sets could suffer from highly imbalanced classes. An example is when there are proportionally more samples representing sinus rhythm than atrial fibrillation. In this case, the model may be biased to focus more on correctly classifying sinus rhythm, as this contributed more to higher overall classification performance. However, this led to poor characterization of cardiovascular disease, as the corresponding samples would be misclassified more often than sinus rhythm. In all, 6 studies [32,41,59-62] mitigated this by (randomly) up-sampling the minority class. A total of 4 studies [22,29,48,52] used the synthetic minority oversampling technique [76].

Finally, it is noteworthy that some studies [41-43,45,49,51,63] constructed a semi–patient-specific model. This could be beneficial, as there were large differences in heart rate data among individuals [71]. This was done by training only a small number of samples from the target patient together with data from other patients. The test set consisted of the remainder of the target patient's samples, which caused overlapping grouping between the training and test sets.

Figure 8. Venn diagram of validation methods used in the studies. CV: cross-validation; G: grouping; H: holdout; S: stratification.



Discussion

Principal Findings

We have shown that machine learning–based technologies that detect cardiovascular outcomes using wearables, bottleneck at TRL5, most dominantly on the requirement of proper realistic data acquisition. To progress to the next level of technology readiness, models need to become operational (either interventional or observational) in a health care system. A study by Komorowski [13] supports these observations and defines the lack of testing or deployment in clinical practice as an *information bottleneck*, which often occurs in medical machine learning. Moreover, half of the eligible studies used a benchmark data set (27/55, 49%), and the most common data set [77] was used 18 times. We argue that overusing a data set can introduce bias and overfitting, effectively making such a data set useless, thereby increasing the need for realistic data sets even more.

The usefulness of wearable cardiovascular diagnostics lies in free-living and active situations because the low burden for wearing them and the 24/7 monitoring abilities. Placement of the sensor on the wrist does fit these criteria best. Moreover, commercial-grade smartwatches can measure multimodal data with low battery consumption. This makes these types of sensors promising to use wearable technology for cardiovascular diagnostics. However, most studies do not fully demonstrate this potential. Moreover, very few prognostic models have been proposed so that cardiovascular disease prevention using wearable machine learning is, in fact, not (yet) well researched.

Although most studies include detailed baseline characteristics of the study population, it is worrisome that the data were not described with a similar level of consistency, structure, and detail. For example, some studies (explicitly or implicitly) have reported acquiring continuous wearable data, but participants do need to take off the device for charging or otherwise have a low compliance rate. These studies then fail to report these details; thus, it is unknown how *continuous* the data, that is, the length of the recording duration, actually is. We believe that, analogous to the baseline characteristics, data characteristics should be reported in detail to predict how a model will generalize when deployed in a particular setting and environment.

The segmentation of the time series data in the windows was performed with a fixed window size in all studies. None of the studies have considered a variable-length or adaptive window size. Furthermore, no previous physiological knowledge has been used to determine informative timescales. For example, the exercise-recovery curve (usually obtained from an exercise tolerance test) is often used to quantify cardiovascular characteristics during activity. This describes a participant's ability to adaptively increase the heart rate during exercise and recover it back to a resting level after exercise. Studies that had access to accelerometer data did not look at similar timescale events. To this end, we believe that identifying informative timescales within the time series and incorporating this in the model can be valuable to detect cardiovascular diseases.

Remarkably, studies primarily prefer nonsequential neural networks over sequential ones, although the latter is designed for time series data. Similarly, the hierarchical structure of the data has rarely been exploited in the published models. We advocate that much more emphasis should be on the exploration of these models, although this also requires larger data sets as these methods are data hungry.

Although some studies make use of a healthy control group, most do not include a group with *no arrhythmia*, *sinus rhythm*, or a similar group, although diagnosing a participant having no arrhythmia at all is just as, or even more powerful, than detecting a specific heart problem. From a machine learning point of view,

this can be seen as a one-class classification (outlier detection) problem: instead of predicting a diverse set of clinical outcomes, the focus of these models lies in modeling the *normal class* as good as possible and consider deviating data as abnormal. Thus, this would be an interesting avenue to explore. In general, it is important to have clearly defined data annotations. For example, some studies have annotated sinus rhythm events in patients with arrhythmia. One might question whether this is similar to annotated sinus rhythm events for nonarrhythmic individuals and whether a machine learning–based approach might fail by mixing these annotations.

We have shown that studies use a training sample size that is too small according to a priori power analysis. Sample size determination in machine learning [78] is focused on posthoc methods, such as learning curves [79]. Prehoc methods, such as power analysis, are difficult in machine learning, as there are many factors that influence the effect size of the model. Furthermore, we have discussed different validation schemes that can be used. An important observation is that a significant number of studies do not validate their model using a nonoverlapping grouping strategy. We believe that validation based on nonoverlapping grouping is crucial for cardiovascular machine learning and any medical machine learning validation in general. Without, experiments will simply suggest performances that are too optimistic.

We have shown that only a few papers used multimodal data and even less considered features across modalities. In our view, this is a missed opportunity; there is valuable information to be extracted when combining features from different modalities. An example is the correlation between heart rate and activity. When the heart rate changes abruptly without activity, this can indicate an interesting segment for a model to detect heart problems. As another example, 1 study used timestamps as features that can provide information about seasonality in longitudinal data. This could be used to inspect (change in) circadian rhythm as a biomarker for cardiovascular disease. Interestingly, ECG morphology is well researched and used as a feature. However, no analogous decomposition of PPG signals is used in the studies. Therefore, we advocate a similar exploration of the PPG signals.

Finally, we argue that in addition to the technical shortcomings discussed, societal factors (under the umbrella term ethical or socially responsible AI) must also be addressed [80]. From the patients' point of view, there are concerns regarding reliability, privacy, and especially fairness and AI bias of the system [81]. Our findings of the lack of realistic data and the imbalance in data link to the latter, as it introduces sampling bias [82], for example. A study by Parikh et al [83] refers to this as a statistical bias and argues that, especially in the medical field, there can also be social biases that are caused by inequity of patients' access to health care (technology) or a combination of both, for example, missing data in certain subgroups. Efforts should be made to remove bias in data (before exposing to an AI model) [80] and in the model itself. This referred to as *debiasing* [80,82,84].

From the physicians' point of view, the performance of machine learning models is potentially reaching that of health care professionals' point of view [85,86], which brings techno-dystopic fear of rivalry between AI and human experts. The study by Di Ieva [87] offers an alternative view by stating that this fear can be overcome by considering the success of multidisciplinary teams in modern medicine and that in line with that paradigm, AI is an assisting expert in that team, rather than a competitor.

As a final note, we would like to emphasize that we did not fully perform a quality assessment of the risk of bias in the clinical data acquisition of the studies. Instead, we used the TRL to capture these risks from a machine learning perspective and describe these limitations throughout. To this end, studies with low methodological quality did not achieve a higher TRL. In addition, we did not consider conference papers as journal papers are more comprehensive and elaborate in general. However, in the field of machine learning, conferences are used to publish completed research (not limited to an abstract as in other fields). Therefore, we might have missed new developments from conference papers that have been described in detail, yet not fully scrutinized as in journal papers.

Conclusions

TRL has enabled us to perform a structured assessment of the (required) progression of machine learning–based wearable technology for deployment in an operational setting. We discussed that the promise is mainly achieved by acquiring longitudinal data from participants in a free-living environment, which is made possible because of low–energy-consuming sensors that are easy to wear. However, we have also observed that none of the studies detect or predict cardiovascular outcomes on realistic data, which limits TRL of this technology. In addition, we identified many other aspects that hamper deployment progression, which need to be addressed before the promise of using wearable technology for cardiovascular disease detection and prevention becomes reality. On the other hand, of the 55 included studies, 6 (11%) were published before 2018 and the remaining 49 (89%) after. Therefore, we expect a large increase in research popularity in the coming years.

Authors' Contributions

The literature search and study inclusion, formal analysis, conceptualization, and the writing of original draft and preparation of the figures were carried out by ANJ. Supervision, conceptualization, and writing—review and editing—were carried out by DT, MR, and IVDB.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search queries performed in the three electronic databases.
[DOCX File , 17 KB - medinform_v10i1e29434_app1.docx]

Multimedia Appendix 2
Tables of study characteristics.
[DOCX File , 250 KB - medinform_v10i1e29434_app2.docx]

References

1. Rizwan A, Zoha A, Mabrouk IB, Sabbour HM, Al-Sumaiti AS, Alomainy A, et al. A review on the state of the art in atrial fibrillation detection enabled by machine learning. IEEE Rev Biomed Eng 2021;14:219-239. [doi: 10.1109/rbme.2020.2976507]
2. Pereira T, Tran N, Gadhoumi K, Pelter MM, Do DH, Lee RJ, et al. Photoplethysmography based atrial fibrillation detection: a review. NPJ Digit Med 2020;3:3 [FREE Full text] [doi: 10.1038/s41746-019-0207-9] [Medline: 31934647]
3. Chua S, Chen L, Lien L, Lo H, Liao Z, Chao S, et al. Comparison of arrhythmia detection by 24-hour holter and 14-day continuous electrocardiography patch monitoring. Acta Cardiol Sin 2020;36(3):251-259. [doi: 10.6515/ACS.202005_36(3).20190903A]
4. Pradhan S, Robinson JA, Shivapour JK, Snyder CS. Ambulatory arrhythmia detection with ZIO® XT Patch in pediatric patients: a comparison of devices. Pediatr Cardiol 2019 Jun 1;40(5):921-924. [doi: 10.1007/s00246-019-02089-0] [Medline: 30937502]
5. Brahmbhatt DH, Cowie MR. Remote management of heart failure: an overview of telemonitoring technologies. Card Fail Rev 2019 May 24;5(2):86-92 [FREE Full text] [doi: 10.15420/cfr.2019.5.3] [Medline: 31179018]
6. van Stipdonk A, Vanbelle S, Horst TI, Luermans J, Meine M, Maass A, et al. Large variability in clinical judgement and definitions of left bundle branch block to identify candidates for cardiac resynchronisation therapy. Int J Cardiol 2019 Jul 01;286:61-65 [FREE Full text] [doi: 10.1016/j.ijcard.2019.01.051] [Medline: 30661850]
7. Slomka PJ, Dey D, Sitek A, Motwani M, Berman DS, Germano G. Cardiac imaging: working towards fully-automated machine analysis and interpretation. Expert Rev Med Devices 2017 Mar 13;14(3):197-212 [FREE Full text] [doi: 10.1080/17434440.2017.1300057] [Medline: 28277804]
8. Trockel MT, Menon NK, Rowe SG, Stewart MT, Smith R, Lu M, et al. Assessment of physician sleep and wellness, burnout, and clinically significant medical errors. JAMA Netw Open 2020 Dec 01;3(12):e2028111 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.28111] [Medline: 33284339]
9. Krittanawong C, Rogers AJ, Johnson KW, Wang Z, Turakhia MP, Halperin JL, et al. Integration of novel monitoring devices with machine learning technology for scalable cardiovascular management. Nat Rev Cardiol 2021 Feb;18(2):75-91 [FREE Full text] [doi: 10.1038/s41569-020-00445-9] [Medline: 33037325]
10. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021 Aug 04;18(8):581-599 [FREE Full text] [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
11. Giebel GD, Gissel C. Accuracy of mHealth devices for atrial fibrillation screening: systematic review. JMIR Mhealth Uhealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
12. Héder M. From NASA to EU: the evolution of the TRL Scale in public sector innovation. Innov J 2017;22(2):1-23 [FREE Full text]
13. Komorowski M. Artificial intelligence in intensive care: are we there yet? Intens Care Med 2019 Sep;45(9):1298-1300. [doi: 10.1007/s00134-019-05662-6] [Medline: 31236638]
14. Fleuren LM, Thoral P, Shillan D, Ercole A, Elbers PW, Right Data Right Now Collaborators. Machine learning in intensive care medicine: ready for take-off? Intens Care Med 2020 Jul;46(7):1486-1488. [doi: 10.1007/s00134-020-06045-y] [Medline: 32399747]
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Br Med J 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
16. Stehlik J, Schmalfuss C, Bozkurt B, Nativi-Nicolau J, Wohlfahrt P, Wegerich S, et al. Continuous wearable monitoring analytics predict heart failure hospitalization: the LINK-HF multicenter study. Circ Heart Fail 2020 Mar;13(3):e006513. [doi: 10.1161/CIRCHEARTFAILURE.119.006513] [Medline: 32093506]
17. Torres-Soto J, Ashley EA. Multi-task deep learning for cardiac rhythm detection in wearable devices. NPJ Digit Med 2020 Sep 09;3(1):116 [FREE Full text] [doi: 10.1038/s41746-020-00320-4] [Medline: 34497341]
18. Bashar SK, Han D, Zieneddin F, Ding E, Fitzgibbons TP, Walkey AJ, et al. Novel density poincaré plot based machine learning method to detect atrial fibrillation from premature atrial/ventricular contractions. IEEE Trans Biomed Eng 2021 Feb;68(2):448-460. [doi: 10.1109/TBME.2020.3004310] [Medline: 32746035]

19. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol 2018 May 01;3(5):409-416. [doi: 10.1001/jamacardio.2018.0136] [Medline: 29562087]
20. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. Circ Arrhythm Electrophysiol 2019 Jun;12(6):e006834 [FREE Full text] [doi: 10.1161/CIRCEP.118.006834] [Medline: 31113234]
21. Majumder AJ, ElSaadany YA, Young R, Ucci DR. An energy efficient wearable smart IoT system to predict cardiac arrest. Adv Hum-Comput Interact 2019 Feb 12;2019:1-21. [doi: 10.1155/2019/1507465]
22. Lee K, Kim S, Choi HO, Lee J, Nam Y. Analyzing electrocardiogram signals obtained from a nymi band to detect atrial fibrillation. Multimed Tools Appl 2018 Dec 17;79(23-24):15985-15999. [doi: 10.1007/s11042-018-7075-1]
23. Green EM, van Mourik R, Wolfus C, Heitner SB, Dur O, Semigran MJ. Machine learning detection of obstructive hypertrophic cardiomyopathy using a wearable biosensor. NPJ Digit Med 2019 Jun 24;2(1):57 [FREE Full text] [doi: 10.1038/s41746-019-0130-0] [Medline: 31304403]
24. Pereira T, Ding C, Gadhoumi K, Tran N, Colorado RA, Meisel K, et al. Deep learning approaches for plethysmography signal quality assessment in the presence of atrial fibrillation. Physiol Meas 2019 Dec 27;40(12):125002 [FREE Full text] [doi: 10.1088/1361-6579/ab5b84] [Medline: 31766037]
25. Chen E, Jiang J, Su R, Gao M, Zhu S, Zhou J, et al. A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. Heart Rhythm 2020 May;17(5 Pt B):847-853 [FREE Full text] [doi: 10.1016/j.hrthm.2020.01.034] [Medline: 32354449]
26. Chou W, Wu P, Fang C, Yen Y, Lin B. Design of smart brain oxygenation monitoring system for estimating cardiovascular disease severity. IEEE Access 2020;8:98422-98429. [doi: 10.1109/access.2020.2997865]
27. Altini M, Casale P, Penders J, Amft O. Cardiorespiratory fitness estimation in free-living using wearable sensors. Artif Intell Med 2016 Mar;68:37-46. [doi: 10.1016/j.artmed.2016.02.002] [Medline: 26948954]
28. Akbulut FP, Akan A. A smart wearable system for short-term cardiovascular risk assessment with emotional dynamics. Measurement 2018 Nov;128:237-246. [doi: 10.1016/j.measurement.2018.06.050]
29. Corino VD, Laureanti R, Ferranti L, Scarpini G, Lombardi F, Mainardi LT. Detection of atrial fibrillation episodes using a wristband device. Physiol Meas 2017 May;38(5):787-799. [doi: 10.1088/1361-6579/aa5dd7] [Medline: 28151434]
30. Kwon S, Hong J, Choi E, Lee B, Baik C, Lee E, et al. Detection of atrial fibrillation using a ring-type wearable device (CardioTracker) and deep learning analysis of photoplethysmography signals: prospective observational proof-of-concept study. J Med Internet Res 2020 May 21;22(5):e16443 [FREE Full text] [doi: 10.2196/16443] [Medline: 32348254]
31. Mena LJ, Félix VG, Ochoa A, Ostos R, González E, Aspuru J, et al. Mobile personal health monitoring for automated classification of electrocardiogram signals in elderly. Comput Math Methods Med 2018;2018:9128054 [FREE Full text] [doi: 10.1155/2018/9128054] [Medline: 30002725]
32. Jeon E, Oh K, Kwon S, Son H, Yun Y, Jung E, et al. A lightweight deep learning model for fast electrocardiographic beats classification with a wearable cardiac monitor: development and validation study. JMIR Med Inform 2020 Mar 12;8(3):e17037 [FREE Full text] [doi: 10.2196/17037] [Medline: 32163037]
33. Xia Y, Xie Y. A novel wearable electrocardiogram classification system using convolutional neural networks and active learning. IEEE Access 2019;7:7989-8001. [doi: 10.1109/access.2019.2890865]
34. Miao F, Wang X, Yin L, Li Y. A wearable sensor for arterial stiffness monitoring based on machine learning algorithms. IEEE Sensors J 2019 Feb 15;19(4):1426-1434. [doi: 10.1109/jsen.2018.2880434]
35. Kim Y, Mahmood M, Lee Y, Kim NK, Kwon S, Herbert R, et al. Stretchable hybrid electronics: all ‐ in ‐ one, wireless, stretchable hybrid electronics for smart, connected, and ambulatory physiological monitoring (Adv. Sci. 17/2019). Adv Sci 2019 Sep 04;6(17):1970104. [doi: 10.1002/advs.201970104]
36. Sharma A, Garg N, Patidar S, San Tan R, Acharya UR. Automated pre-screening of arrhythmia using hybrid combination of Fourier-Bessel expansion and LSTM. Comput Biol Med 2020 May;120:103753. [doi: 10.1016/j.compbiomed.2020.103753] [Medline: 32421653]
37. Kwon S, Hong J, Choi E, Lee E, Hostallero DE, Kang WJ, et al. Deep learning approaches to detect atrial fibrillation using photoplethysmographic signals: algorithms development study. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e12770 [FREE Full text] [doi: 10.2196/12770] [Medline: 31199302]
38. Oresko JJ, Duschl H, Cheng AC. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. IEEE Trans Inf Technol Biomed 2010 May;14(3):734-740. [doi: 10.1109/TITB.2010.2047865] [Medline: 20388600]
39. Sadrawi M, Lin C, Lin Y, Hsieh Y, Kuo C, Chien JC, et al. Arrhythmia evaluation in wearable ECG devices. Sensors (Basel) 2017 Oct 25;17(11):2445 [FREE Full text] [doi: 10.3390/s17112445] [Medline: 29068369]
40. Allami R. Premature ventricular contraction analysis for real-time patient monitoring. Biomed Signal Process Control 2019 Jan;47:358-365. [doi: 10.1016/j.bspc.2018.08.040]
41. Zhao Y, Shang Z, Lian Y. A 13.34 µW event-driven patient-specific ANN cardiac arrhythmia classifier for wearable ECG sensors. IEEE Trans Biomed Circuits Syst 2020 Apr;14(2):186-197. [doi: 10.1109/TBCAS.2019.2954479] [Medline: 31794404]

42. Wu J, Li F, Chen Z, Pu Y, Zhan M. A neural network-based ECG classification processor with exploitation of heartbeat similarity. IEEE Access 2019;7:172774-172782. [doi: 10.1109/access.2019.2956179]
43. Saadatnejad S, Oveisi M, Hashemi M. LSTM-based ECG classification for continuous monitoring on personal wearable devices. IEEE J Biomed Health Inform 2019 Apr 15;24(2):515-523. [doi: 10.1109/JBHI.2019.2911367] [Medline: 30990452]
44. Sopic D, Aminifar A, Aminifar A, Atienza D. Real-time event-driven classification technique for early detection and prevention of myocardial infarction on wearable systems. IEEE Trans Biomed Circuits Syst 2018 Oct;12(5):982-992. [doi: 10.1109/tbcas.2018.2848477]
45. Tang X, Ma Z, Hu Q, Tang W. A real-time arrhythmia heartbeats classification algorithm using parallel delta modulations and rotated Linear-Kernel support vector machines. IEEE Trans Biomed Eng 2020 Apr;67(4):978-986. [doi: 10.1109/TBME.2019.2926104] [Medline: 31265382]
46. Feng K, Pi X, Liu H, Sun K. Myocardial infarction classification based on convolutional neural network and recurrent neural network. Appl Sci 2019 May 07;9(9):1879. [doi: 10.3390/app9091879]
47. Shah AJ, Isakadze N, Levantsevych O, Vest A, Clifford G, Nemati S. Detecting heart failure using wearables: a pilot study. Physiol Meas 2020 May 04;41(4):044001. [doi: 10.1088/1361-6579/ab7f93] [Medline: 32163936]
48. Zhang H, Dong Z, Gao J, Lu P, Wang Z. Automatic screening method for atrial fibrillation based on lossy compression of the electrocardiogram signal. Physiol Meas 2020 Aug 21;41(7):075005. [doi: 10.1088/1361-6579/ab979f] [Medline: 32464608]
49. Huang S, Lu H. Classification of temporal data using dynamic time warping and compressed learning. Biomed Signal Process Control 2020 Mar;57:101781. [doi: 10.1016/j.bspc.2019.101781]
50. Yang C, Aranoff ND, Green P, Tavassolian N. Classification of aortic stenosis using time-frequency features from chest cardio-mechanical signals. IEEE Trans Biomed Eng 2020 Jun;67(6):1672-1683. [doi: 10.1109/TBME.2019.2942741] [Medline: 31545706]
51. Amirshahi A, Hashemi M. ECG classification algorithm based on STDP and R-STDP neural networks for real-time monitoring on ultra low-power personal wearable devices. IEEE Trans Biomed Circuits Syst 2019 Dec;13(6):1483-1493. [doi: 10.1109/TBCAS.2019.2948920] [Medline: 31647445]
52. Yan Z, Zhou J, Wong W. Energy efficient ECG classification with spiking neural network. Biomed Signal Process Control 2021 Jan;63:102170. [doi: 10.1016/j.bspc.2020.102170]
53. Lui HW, Chow KL. Multiclass classification of myocardial infarction with convolutional and recurrent neural networks for portable ECG devices. Informatics Med Unlocked 2018;13:26-33. [doi: 10.1016/j.imu.2018.08.002]
54. Shao M, Zhou Z, Bin G, Bai Y, Wu S. A wearable electrocardiogram telemonitoring system for atrial fibrillation detection. Sensors (Basel) 2020 Jan 22;20(3):606 [FREE Full text] [doi: 10.3390/s20030606] [Medline: 31979184]
55. Gilani M, Eklund JM, Makrehchi M. Automated detection of atrial fibrillation episode using novel heart rate variability features. Conf Proc IEEE Eng Med Biol Soc 2016 Dec;2016:3461-3464. [doi: 10.1109/EMBC.2016.7591473] [Medline: 28269045]
56. Mei Z, Gu X, Chen H, Chen W. Automatic atrial fibrillation detection based on heart rate variability and spectral features. IEEE Access 2018;6:53566-53575. [doi: 10.1109/access.2018.2871220]
57. Lown M, Brown M, Brown C, Yue AM, Shah BN, Corbett SJ, et al. Machine learning detection of Atrial Fibrillation using wearable technology. PLoS One 2020;15(1):e0227401 [FREE Full text] [doi: 10.1371/journal.pone.0227401] [Medline: 31978173]
58. Xia Y, Zhang H, Xu L, Gao Z, Zhang H, Liu H, et al. An automatic cardiac arrhythmia classification system with wearable electrocardiogram. IEEE Access 2018;6:16529-16538. [doi: 10.1109/ACCESS.2018.2807700]
59. Ma F, Zhang J, Liang W, Xue J. Automated classification of atrial fibrillation using artificial neural network for wearable devices. Math Probl Eng 2020 Apr 25;2020:1-6. [doi: 10.1155/2020/9159158]
60. Wang N, Zhou J, Dai G, Huang J, Xie Y. Energy-efficient intelligent ECG monitoring for wearable devices. IEEE Trans Biomed Circuits Syst 2019 Oct;13(5):1112-1121. [doi: 10.1109/TBCAS.2019.2930215] [Medline: 31329129]
61. Scirè A, Tropeano F, Anagnostopoulos A, Chatzigiannakis I. Fog-computing-based heartbeat detection and arrhythmia classification using machine learning. Algorithms 2019 Feb 02;12(2):32. [doi: 10.3390/a12020032]
62. Fan X, Yao Q, Cai Y, Miao F, Sun F, Li Y. Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings. IEEE J Biomed Health Inform 2018 Nov;22(6):1744-1753. [doi: 10.1109/JBHI.2018.2858789] [Medline: 30106699]
63. Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng 2016 Mar;63(3):664-675. [doi: 10.1109/TBME.2015.2468589] [Medline: 26285054]
64. Lai D, Bu Y, Su Y, Zhang X, Ma C. A flexible multilayered dry electrode and assembly to single-lead ECG patch to monitor atrial fibrillation in a real-life scenario. IEEE Sensors J 2020 Oct 15;20(20):12295-12306. [doi: 10.1109/JSEN.2020.2999101]
65. Inan OT, Pouyan MB, Javaid AQ, Dowling S, Etemadi M, Dorier A, et al. Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients. Circ Heart Fail 2018 Jan;11(1):e004313 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.117.004313] [Medline: 29330154]
66. Alfaras M, Soriano MC, Ortín S. A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. Front Phys 2019 Jul 18;7:103. [doi: 10.3389/fphy.2019.00103]

67. Rubin J, Parvaneh S, Rahman A, Conroy B, Babaeizadeh S. Densely connected convolutional networks for detection of atrial fibrillation from short single-lead ECG recordings. J Electrocardiol 2018;51(6S):18-21. [doi: 10.1016/j.jelectrocard.2018.08.008] [Medline: 30122456]
68. Xu X, Liu H. ECG heartbeat classification using convolutional neural networks. IEEE Access 2020;8:8614-8619. [doi: 10.1109/ACCESS.2020.2964749]
69. Li Y, Pang Y, Wang J, Li X. Patient-specific ECG classification by deeper CNN from generic to dedicated. Neurocomputing 2018 Nov;314:336-346. [doi: 10.1016/j.neucom.2018.06.068]
70. Mastoi Q, Wah T, Gopal Raj R. Reservoir computing based echo state networks for ventricular heart beat classification. Appl Sci 2019 Feb 18;9(4):702. [doi: 10.3390/app9040702]
71. Quer G, Gouda P, Galarnyk M, Topol E, Steinhubl S. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: retrospective, longitudinal cohort study of 92,457 adults. PLoS One 2020;15(2):e0227709 [FREE Full text] [doi: 10.1371/journal.pone.0227709] [Medline: 32023264]
72. Fritz G. Dale Dubin: Rapid interpretation of EKG's. 6th edition. Cover Publishing Company/2000. Tampa, Fla., USA. 388 pp. Intens Care Med 2001 Oct 30;27(11):1832. [doi: 10.1007/s001340101100]
73. Quintana DS. Statistical considerations for reporting and planning heart rate variability case-control studies. Psychophysiology 2017 Mar;54(3):344-349. [doi: 10.1111/psyp.12798] [Medline: 27914167]
74. Bishop CM. Pattern recoginiton and machine learning. In: Information Science and Statistics. New York: Springer; 2006:32-33.
75. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 2017 Mar 03;40(8):913-929. [doi: 10.1111/ecog.02881]
76. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
77. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag 2001;20(3):45-50. [doi: 10.1109/51.932724] [Medline: 11446209]
78. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. Can Assoc Radiol J 2019 Nov 29;70(4):344-353. [doi: 10.1016/j.carj.2019.06.002] [Medline: 31522841]
79. Richter AN, Khoshgoftaar TM. Sample size determination for biomedical big data with limited labels. Netw Model Anal Health Inform Bioinforma 2020 Jan 11;9(1):1-13. [doi: 10.1007/s13721-020-0218-0]
80. Cheng L, Varshney KR, Liu H. Socially responsible ai algorithms: issues, purposes, and challenges. J Artif Intell Res 2021 Aug 28;71:1137-1181. [doi: 10.1613/jair.1.12814]
81. Tran V, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. NPJ Digit Med 2019 Jun 14;2(1):53 [FREE Full text] [doi: 10.1038/s41746-019-0132-y] [Medline: 31304399]
82. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. Commun Med (London) 2021 Aug 23;1(1):25. [doi: 10.1038/s43856-021-00028-w] [Medline: 34522916]
83. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. J Am Med Assoc 2019 Dec 24;322(24):2377-2378. [doi: 10.1001/jama.2019.18058] [Medline: 31755905]
84. Chen J, Dong H, Wang X, Feng F, Wang M, He X. Bias and debias in recommender system: a survey and future directions. arXiv. 2020. URL: http://arxiv.org/abs/2010.03240 [accessed 2022-01-05]
85. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019 Oct;1(6):271-297. [doi: 10.1016/s2589-7500(19)30123-2]
86. van Doorn WP, Stassen PM, Borggreve HF, Schalkwijk MJ, Stoffers J, Bekers O, et al. A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. PLoS One 2021 Jan 19;16(1):e0245157 [FREE Full text] [doi: 10.1371/journal.pone.0245157] [Medline: 33465096]
87. Di Ieva A. AI-augmented multidisciplinary teams: hype or hope? Lancet 2019 Nov;394(10211):1801. [doi: 10.1016/s0140-6736(19)32626-1]

Abbreviations

AI: artificial intelligence
ECG: electrocardiogram
PPG: photoplethysmogram
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TRL: technology readiness level

Edited by C Lovis; submitted 02.07.21; peer-reviewed by HL Tam, B Merid; comments to author 06.11.21; revised version received 22.11.21; accepted 04.12.21; published 19.01.22.

Please cite as:
Naseri Jahfari A, Tax D, Reinders M, van der Bilt I
Machine Learning for Cardiovascular Outcomes From Wearable Data: Systematic Review From a Technology Readiness Level Point of View
JMIR Med Inform 2022;10(1):e29434
URL: https://medinform.jmir.org/2022/1/e29434
doi:10.2196/29434
PMID:35044316

Review

The Use of Blockchain Technology in the Health Care Sector: Systematic Review

Deepa Elangovan[1], BPharm (Hons); Chiau Soon Long[2], BSc; Faizah Safina Bakrin[1], PhD; Ching Siang Tan[1], PhD; Khang Wen Goh[3], PhD; Siang Fei Yeoh[4], MClinPharm; Mei Jun Loy[5], BEng (Hons); Zahid Hussain[6], PhD; Kah Seng Lee[7], PhD; Azam Che Idris[8], PhD; Long Chiau Ming[9], PhD

[1]School of Pharmacy, KPJ Healthcare University College, Nilai, Malaysia

[2]Faculty of Computing and Engineering, Quest International University, Ipoh, Malaysia

[3]Faculty of Information Technology, INTI International University, Nilai, Malaysia

[4]Department of Pharmacy, National University Hospital, Singapore, Singapore

[5]Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia

[6]Faculty of Health, University of Canberra, Canberra, Australia

[7]Faculty of Pharmacy, University of Cyberjaya, Cyberjaya, Malaysia

[8]Faculty of Integrated Technologies, Universiti Brunei Darussalam, Gadong, Brunei Darussalam

[9]Pengiran Anak Puteri Rashidah Sa'adatul Bolkiah Institute of Health Sciences, Universiti Brunei Darussalam, Gadong, Brunei Darussalam

Corresponding Author:
Khang Wen Goh, PhD
Faculty of Information Technology
INTI International University
Persiaran Perdana BBN Putra Nilai
Nilai, 71800
Malaysia
Phone: 60 6 798 2000
Email: khangwen.goh@newinti.edu.my

Abstract

Background: Blockchain technology is a part of Industry 4.0's new Internet of Things applications: decentralized systems, distributed ledgers, and immutable and cryptographically secure technology. This technology entails a series of transaction lists with identical copies shared and retained by different groups or parties. One field where blockchain technology has tremendous potential is health care, due to the more patient-centric approach to the health care system as well as blockchain's ability to connect disparate systems and increase the accuracy of electronic health records.

Objective: The aim of this study was to systematically review studies on the use of blockchain technology in health care and to analyze the characteristics of the studies that have implemented blockchain technology.

Methods: This study used a systematic review methodology to find literature related to the implementation aspect of blockchain technology in health care. Relevant papers were searched for using PubMed, SpringerLink, IEEE Xplore, Embase, Scopus, and *EBSCOhost*. A quality assessment of literature was performed on the 22 selected papers by assessing their trustworthiness and relevance.

Results: After full screening, 22 papers were included. A table of evidence was constructed, and the results of the selected papers were interpreted. The results of scoring for measuring the quality of the publications were obtained and interpreted. Out of 22 papers, a total of 3 (14%) high-quality papers, 9 (41%) moderate-quality papers, and 10 (45%) low-quality papers were identified.

Conclusions: Blockchain technology was found to be useful in real health care environments, including for the management of electronic medical records, biomedical research and education, remote patient monitoring, pharmaceutical supply chains, health insurance claims, health data analytics, and other potential areas. The main reasons for the implementation of blockchain technology in the health care sector were identified as data integrity, access control, data logging, data versioning, and nonrepudiation. The findings could help the scientific community to understand the implementation aspect of blockchain technology. The results from this study help in recognizing the accessibility and use of blockchain technology in the health care sector.

(JMIR Med Inform 2022;10(1):e17278) doi:10.2196/17278

KEYWORDS

blockchain; health care; hospital information system; data integrity; access control; data logging; health informatics

Introduction

Health informatics (HI) is an extension of medical informatics that concentrates on the clinical sector and implementation of technology in the distribution of health care [1]. Changes in both technology and health care are leading to the evolution of health care informatics. With current technology, HI provides fundamental, indivisible knowledge bases to health professionals and health care organizations to provide patients with a better quality of care services [2]. Acquiring and recording medical and patient information, liaising with health care professionals, choosing an appropriate diagnostic method, elucidating laboratory findings, and gathering clinical research information are known as information processing and communication in the health care sector [3].

Electronic health record (EHR) systems and hospital information systems (HISs) are widely used across the world. However, the current HISs are mainly cloud based, are stored by one particular data contractor, and have several disadvantages, such as a lack of sufficient security measures. This has led to innumerable breaches of data, as well as issues of data validity and data sharing, which have left patients exposed to economic threats and possible social stigma. Centralized data or information is an appealing target for cyberattacks, and issues arise due to establishing a persistent view of the patient data across a network [4].

Taking these issues into consideration, an improved tamperproof and hackproof database management system is much needed to replace the current system that has been used for the past several decades. The new innovative system should have better data security and be able to integrate with other information technology (IT) systems, such as finance and admission systems. When blockchain technology was introduced in 2008, it largely fulfilled all of these criteria, alongside its versatility for applications in banking and finance.

Blockchain is a decentralized database that maintains an uninterrupted, growing list of data records that are established by the nodes involved. The information is recorded in a public ledger that includes data from every completed transaction [5]. Along with this, blockchain is also a sort of dispersed ledger of cryptographically chained blocks where value-exchanged transactions are consecutively aggregated. Blockchain also exhibits properties such as decentralization, security, anonymity, and data integrity with the absence of a mediator to control agreement and inalterability [6]. The information in blockchain is transparent and tamperproof due to the continuous series of blocks, which contain information and data [7].

Blockchain is a decentralized database that is not owned by anybody and is simultaneously owned by everybody, as the contents are available to all parties involved. For example, with Bitcoin, since all transactions are processed by users via a particular pseudonym, the information contained in the blockchain is completely anonymous [8]. This revolutionary system has indeed overcome some of the limitations faced by the existing system. Nevertheless, further exploration in terms of implementation and practicality is much needed and will be discussed in the following sections [9].

This study aimed to systematically review studies on the use of blockchain technology in health care. We also analyzed the characteristics of the studies that have implemented blockchain technology. This study will be impactful in helping the scientific community to understand the use aspect of blockchain technology based on the findings of completed studies. The results from this study will help to identify the accessibility as well as the implementation of blockchain technology in the health care sector.

Methods

Study Design

The methodology used in this research was a systematic review, modeled on a recent systematic review about blockchain reported by Böhme et al [7].

Data Sources and Search Strategy

A search was conducted for scientific papers on the research topics. All papers that were relevant for these topics were gathered by using a search protocol that was developed for each scientific database. Possible keywords were tested, and appropriate terms were chosen for the search string. The Medical Subject Headings (MeSH) database was used to derive keywords and search term combinations. PubMed, SpringerLink, IEEE Xplore, Scopus, Embase, and *EBSCOhost* databases were chosen to search for all the relevant literature. The search strings were constructed in accordance with the research domains and research questions and are listed in Multimedia Appendix 1.

Online digital libraries were used to search for relevant papers from January 2008 to September 2019. The year 2008 was chosen as the beginning of the range for this research study because the first published application of blockchain technology (ie, Bitcoin) was introduced in that year, so no blockchain-related studies were conducted before 2008. In this systematic review, the search query was purposely made broad, in order to identify many papers related to the research question. However, when “Bitcoin” was used as a search term, a large number of papers were identified, but the papers were mainly about economic applications rather than applications in the health care sector.

Because the aim of this research was based on finding and mapping the papers related to blockchain technology in the health care sector, “Bitcoin” was dropped as a search term. By using “blockchain” and “health care” as search terms, the majority of Bitcoin-related papers with a technical perspective on blockchain were still included. A manual search was carried

out for papers that were published at workshops, at conferences, in journals, and at symposiums.

Study Selection

Screening of Relevant Papers

The next step in the process was screening relevant papers, wherein the papers that had been found during the previous step were assessed for actual relevance. The screening process started with all of the publications gathered from online digital libraries. A process inspired by Yli-Huumo et al [10] was used to screen for relevant papers. Applicable quotes from the search were entered into, and sorted with the aid of, EndNote X8.0 (Clarivate), which was used to remove duplicate papers. Duplicate references across databases and references that were not appropriate for the study were eliminated from the literature search reference lists. The remaining duplicates were deleted manually.

The iterative approach for title, abstract, and full-text searches was used, and the results were exported to Microsoft Excel 2013. The titles and abstracts of the searched papers were screened first to determine the relevance or appropriateness for this systematic review. At this stage, studies that were clearly not about the use of blockchain technology in health care were excluded. The titles and abstracts were screened by two reviewers based on the inclusion and exclusion criteria.

Inclusion and Exclusion Criteria

The papers that had passed the previous screening phase were screened based on their abstract. In addition, the following specific inclusion and exclusion criteria were used to screen each paper:

1. Inclusion criteria:
   a. Original research study.
   b. Study in English.
   c. Publication on blockchain technology in the health care sector.
   d. Publication including sufficient explanation of the research findings.
2. Exclusion criteria:
   a. Papers without full-text availability.
   b. Papers for which English was not the main language.
   c. Papers that had some other focus instead of the use of blockchain in the health care sector.
   d. Papers that were duplicates.
   e. Search results that were editorials, prefaces, paper summaries, summaries of tutorials, interviews, news items, correspondences, discussions, comments, readers' letters, workshops, panels, and poster sessions.
   f. Publications indicating ideas, magazine publications, and discussion papers.

Abstract Screening Based on Keywords

Screening based on keywords, as defined by Dyba and Dingsøyr [11], was done in two steps. In the first step, identifiable keywords and concepts from the abstracts were analyzed that reflected the contribution of the papers. Developing a greater level of understanding based on these keywords was the second step in this keywording process.

The keywords were used to cluster and form categories. All the selected papers were read after the categories had been clustered. The categories were updated after reading each paper, or if the paper revealed something new, then a new category was created. This step resulted in clustered categories being formed from all the relevant papers based on this research topic. Papers with poor, misleading, or lost abstracts were excluded due to irrelevant information.

After the title-, abstract-, and keyword-screening process, each remaining paper underwent full-text screening based on the same eligibility criteria. Two reviewers resolved discrepancies through discussion, and no adjudication by a third reviewer was required.

Data Gathering and Data Extraction

A template was designed to collect the information required to address the research question. Basic metadata about the publication were collected, such as author name and country, year of publication, source type, and type of publisher.

To categorize the 22 selected papers, further data were extracted. Each full paper was read to extract the keywords or outcomes related to our research question; these were then sorted into the identified categories, as follows:

1. Use cases of blockchain technology in health care that indicate the specific health care area, such as electronic medical records (EMRs), biomedical research and education, remote patient monitoring, drug or pharmaceutical supply chains, health insurance claims, health data analytics, or other areas.
2. Reasons for using blockchain technology in health care, such as data integrity, access control, logging, data versioning, and nonrepudiation.

Literature Quality Assessment

An assessment of literature quality was performed. All of the final 22 publications were independently reviewed and scored by two reviewers. The assessment tool for blockchain-related studies proposed by Petersen et al [12] was used to critically appraise and summarize evidence in the searched papers.

In accordance with Hölbl et al [13], the quality of the papers was assessed using the criteria defined in Table 1.

This tool was used to assess the trustworthiness, relevance, and results of the published papers. These led to the decision of which papers were believable and useful and could be used for the research. A three-tier scale was used to rank the quality of all four questions. A value of 0 ("barely" or "no") was assigned when the criterion was addressed very poorly or not at all, a value of 1 ("partially") was assigned when a criterion was partially addressed, and a value of 2 ("satisfactorily" or "yes") was assigned when the reviewer felt that the publication had successfully satisfied the criterion.

Two reviewers assessed each query from question 1 (Q1) to question 4 (Q4), which resulted in a minimum of 0 points to a maximum of 4 points per query. The minimum score for the

sum of responses to question 2 (Q2), question 3 (Q3), and Q4 was 0 points, and the maximum score was 12 points. The score of the response to Q1 was converted to 40% of the total value, and the total summed score of responses to Q2, Q3, and Q4 was converted to 60% of the total value, with Q2, Q3, and Q4 each contributing 20% of the total points.

So, the overall score was the sum of responses to Q1 to Q4, which is presented as a percentage to enhance readability and comprehension. An explanation of scoring of responses to queries Q1 to Q4 is given in Table 1.

To find the percentage score of the response to Q1, the following equation applies:

To find the percentage score for the sum of responses to Q2 to Q4, the following equation applies:

The overall score is represented by the following equation: Overall score = Percentage score of response to Q1 (%) + Percentage score of sum of responses to Q2 to Q4 (%) [13].

From the calculation, the publications that have an overall score of 90% and above are high-quality papers. An overall score between 80% and 89% indicates a moderate-quality paper, and low-quality papers are represented by an overall score of 79% or less.

Table 1. Parameters for measuring quality of the publications [13].

Question (Q)	Quality assessment query	Responses (scores)
Q1	Is the publication relevant to blockchain?	“barely” (0), “partially” (1), or “satisfactorily” (2)
Q2	Does the publication include and define research objectives adequately?	“no” (0), “partially” (1), or “yes” (2)
Q3	Are limitations and challenges well defined?	“no” (0), “partially” (1), or “yes” (2)
Q4	Is the proposed contribution well described?	“no” (0), “partially” (1), or “yes” (2)

Data Availability

All data have been reported in this manuscript.

Results

Study Selection

A total of 271 papers were initially retrieved as a result of implementation of the search protocol that was designed for searching the selected scientific databases. Of the 271 papers, 34 (12.5%) were from PubMed, 52 (19.2%) were from SpringerLink, 40 (14.8%) were from IEEE Xplore, 56 (20.7%) were from Embase, 45 (16.6%) were from Scopus, and 44 (16.2%) were from *EBSCOhost*. Then, the first screening was done based on the titles of the retrieved papers. All of the paper titles were examined independently by one reviewer based on the inclusion and exclusion criteria, which led to the selection of 175 papers. In the first screening, a total of 52 papers were excluded because they were not related to the research topic (eg, some excluded papers discussed the business perspective of Bitcoin rather than the use of blockchain technology in the health care sector). Meanwhile, 25 papers related to other scientific areas, such as mathematics and chemistry, were excluded from the first screening, as the term “blockchain” had other meanings apart from the technology used in computer science and IT. Through a manual search and using references from the included papers, an additional 10 papers were collected.

After the selection of 185 papers from the first screening, 87 duplicate papers were removed using Endnote X8. This resulted in 98 papers, which underwent further screening based on abstracts where, in some cases, the introduction and conclusion of the full text were analyzed. The abstracts of all the selected papers were read by two reviewers. Some of the papers were removed because the abstracts indicated no relevance to the research topic. The unclear or grey-area abstracts or papers were moved to the next screening step for more in-depth analysis.

A total of 35 papers were identified for full-paper analysis, which was the last stage of paper selection for this systematic review. Each paper was read in full, independently, and assessed for eligibility using the inclusion and exclusion criteria. This resulted in the selection of 22 primary papers. Of the 35 papers identified for full-text analysis, 3 (9%) were dropped because they focused on the economic perspective of Bitcoin and not the health care setting. Of the 35 papers, 5 (14%) were removed as they only described blockchain and how it works, without discussing any actual blockchain implementation in a real health care environment. Furthermore, 3 out of 35 (9%) papers identified as review papers and 2 (6%) papers identified as proposal papers were removed. Figure 1 shows the results of the search strategy. The list of 22 selected primary papers and the extracted data items are included in Table 2 [14-35]. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist for this research study on the use of blockchain technology in the health care sector is included in Multimedia Appendix 2.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the search strategy.

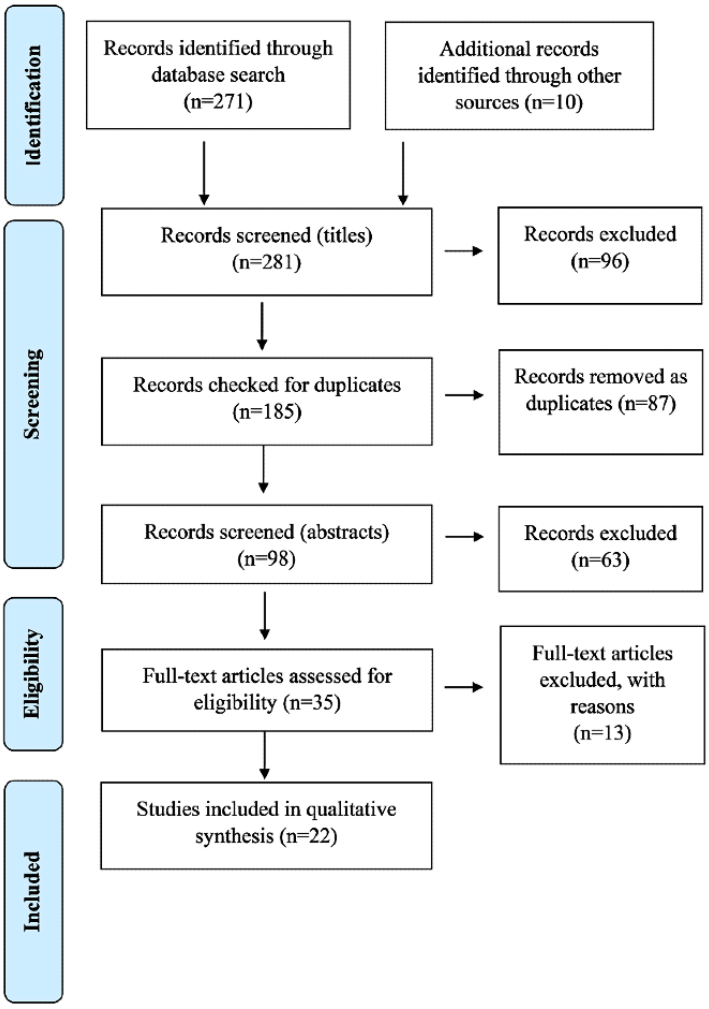


Table 2. Information extracted and collected from the selected papers.

Study first author, year	Location	Use cases and fields		Usability and reasons for using blockchain		Technology
		Application area	Description	Simplified classification	Description	
Maslove, 2018 [14]	Kingston, Canada	Biomedical research and education (ie, clinical trials)	Development of a system that uses a web-based interface to allow users to run trial-related smart contracts on an Ethereum network	Data integrity	Enabling of clinical trials data management; functions allow patients to grant researchers access to their data and allow researchers to submit queries for data that are stored off-chain	Ethereum
Cunningham, 2017 [18]	The Netherlands	Electronic medical record (EMR)	A system that uses smart contract–based Ethereum blockchain technology to operate in a verifiably secure, trustless, and openly auditable environment	Access control	Improvement of the uptake and acceptance of medical informatics platforms where patients directly control medical data in an open and secure manner	Ethereum
Nugent, 2016 [29]	London, the United Kingdom	Biomedical research and education (ie, clinical trials)	A system that uses smart contracts, which enhance the trust in the data and clinical trials; this reduces patient risk and financial strain in health care by allowing better-informed decisions to be made by medical professionals	Data integrity and logging	Improvement of data transparency in clinical trials and immutable records of trial history, which act as trusted administrators; tamper-resistant characteristics of blockchain prevent all forms of manipulation; mainly used for complex clinical trial management	Ethereum
Benchoufi, 2017 [15]	Paris, France	Biomedical research and education (ie, clinical trials)	A system with time-stamping of each patient's consent using blockchain technology in a securely unfalsifiable and transparent way	Nonrepudiation, logging, and data versioning	All consent-related data on the blockchain enhance security, reliability, and transparency and could be a consistent step toward reproducibility	N/A[a]
Ichikawa, 2017 [16]	Tokyo, Japan	Remote patient monitoring (ie, mobile health [mHealth])	Development of a smartphone app with blockchain technology to provide an mHealth system for cognitive behavioral therapy for insomnia	Data integrity	Establishment of accessibility and transparency of data without the third party by incorporating blockchain technology into mHealth; blockchain also serves as a tamperproof system for mHealth	Hyperledger Fabric
Cichosz, 2018 [25]	Aalborg, Denmark	EMR (ie, health care data)	Development of a platform using the New Economy Movement (NEM) multi-signature blockchain contracts to access data management, sharing, and encryption	Access control and data integrity	Improvement in privacy and diabetes data management, where patients have access to control and share their own data	NEM
Omar, 2019 [26]	The United States	EMR	Development of a patient-centric health care data management system using blockchain technology as storage, which enhances privacy	Data integrity	Patients will have overall control over their data; MediBchain increases patients' interest in EMRs or electronic health records and enhances accountability, integrity, pseudonymity, security, and privacy	N/A

Study first author, year	Location	Use cases and fields		Usability and reasons for using blockchain		Technology
		Application area	Description	Simplified classification	Description	
Liang, 2017 [30]	Norfolk, England	Remote patient monitoring	Development of an mHealth care system for personal health data collection, sharing, and collaboration between individuals, health care providers, and insurance companies, and its implementation in a distributed and trustless way	Data integrity, access control, and logging	Improvement of personal health data collection, sharing, validation, protection, and integrity and health care collaboration; this system ensures the scalability and efficiency of the data process by handling a large data set at low latency	Hyperledger Fabric
Kleinaki, 2018 [28]	The Netherlands	Biomedical research and education (ie, database queries)	Presentation and testing of the use of smart digital contracts by a blockchain-based notarization service to seal a biomedical database query using a real blockchain infrastructure	Data integrity and data versioning	Improvement of retrieved data integrity, nonrepudiation, and biomedical evidence data versioning	Ethereum
Bocek, 2017 [32]	Zurich, Switzerland	Pharmaceutical supply chain (ie, ambient temperature)	Implementation of sensor devices using blockchain technology to enhance data immutability and public accessibility of temperature records	Logging	This system can be evaluated automatically, and the stored data are tamperproof with Ethereum, which can be used at a low cost	Ethereum
Mendes, 2018 [22]	Évora, Portugal	EMR	Development of a system with raw blockchain with Hyperledger Fabric by DLA	Data integrity	While consuming low computational power, it enhances tamperproof, fair, and democratic maintenance of the ledger	Hyperledger Fabric
Li, 2018 [20]	Beijing, China	EMR (ie, health record)	Development of a novel blockchain-based data preservation system based on the real-world blockchain-based platform, and its implementation for medical data	Data integrity	Preservation of important data in perpetuity and verification of data originality; illegal operation of the data is detected, and the user is notified on time	Ethereum
Azaria, 2016 [17]	The United States	EMR (ie, health record)	Development of a decentralized record management system using blockchain technology to handle EMRs	Logging and access control	The system becomes more convenient and adaptable in its management of authentication, confidentiality, accountability, and data sharing	Ethereum
Zhou, 2018 [33]	Beijing, China	Health insurance claims	Development of a blockchain-based medical insurance storage system, MIStore; this helps insurance companies obtain patients' medical spending records, which are always confidential	Data integrity and logging	The system provides decentralization and tamper resistance; this gives users high credibility and record-nodes, which help users verify publicly verifiable data	Ethereum
Angeletti, 2017 [27]	Rome, Italy	Biomedical research and education (ie, clinical trials)	Presentation of a digital health application enabling clinical trials recruitment using Internet of Things data; using Ethereum, a proof of concept was implemented, and the application's performance was studied in a real-world evaluation	Data integrity and access control	The clinical research institute can be guaranteed that it is acquiring useful and original data; until an agreement is reached, the individual can keep personal data private	Ethereum

Study first author, year	Location	Use cases and fields		Usability and reasons for using blockchain		Technology
		Application area	Description	Simplified classification	Description	
Saravanan, 2017 [31]	Chennai, India	Remote patient monitoring	Implementation of a new health care paradigm (SMEAD[b]) to aid diabetic patients via development of an end-to-end secured system; implementation of a blockchain-based disruptive technology to facilitate cryptographic security and formalized data access through smart contracts	Access control	The system aids in data storage for millions of patients, and analysis was performed in real time, which promotes an evidence-based medicine system with privacy and security concerns	Ethereum
Zhang, 2018 [24]	The United States	EMR (ie, health record)	Development of a system to support collaborative clinical decision-making via a remote tumor board case study	Access control and data integrity	Improvement of security, trust, and scalable data sharing, which is important for collaborative clinical decision-making; also results in greater data readability	Ethereum
Fan, 2018 [19]	China	EMR (ie, health record)	Development of a blockchain-based information management system, MedBlock, to handle patients' information; this allows for efficient EMR access and retrieval, exhibiting high information security	Access control	Patients can easily access the EMRs of different hospitals; data sharing via blockchain helps the hospital get a full history of patients' medical history before consultations are carried out	N/A
Liu, 2018 [21]	China	EMR	Implementation of blockchain-based privacy-preserving data sharing for EMRs	Access control	The EMRs cannot be modified arbitrarily, which leads to reduced medical data leakage; security analysis shows that this system is a secure and effective way to realize data sharing for EMRs	N/A
Nagasubramanian, 2018 [23]	London, the United Kingdom	EMR	Ensuring secrecy of digital signatures and authentication by using keyless signature infrastructure in the system	Access control	The system ensures data transparency, privacy, confidentiality, and verification of data	N/A
Kotsiuba, 2018 [34]	Ukraine	Health care data analytics	Implementation of a decentralized system with blockchain technology that protects the confidentiality of medical data; patients receive a personal data monitoring tool, allowing them to participate in accelerating medical analytics	Data integrity	Enhancement of medical data safety, extension of the base of clinical data collection, and creation of an effective shared health infrastructure	N/A
Talukder, 2018 [35]	The United States	Others	Implementation of an Ethereum-based Proof of Disease consensus protocol to enhance the accuracy of transactions and eliminate medical errors	Access control and data integrity	Aids in achieving all the complex needs of P6 (participatory, personalized, proactive, preventive, predictive, and precision) medicine and decreases disease burden	Ethereum

[a]N/A: not applicable: the technology was not reported in this paper.

[b]SMEAD: Secured Mobile-Enabled Assisting Device for Diabetics.

Publication Year, Publication Type, and Geographical Distribution

All the selected papers were published since 2016. This indicates that blockchain technology in health care settings is very new. It is noted that from the 22 selected papers, the majority (n=12, 55%) were published in 2018, 7 (32%) papers were published in 2017, 2 (9%) papers were published in 2016, and 1 (5%) paper was published in 2019.

The locations (ie, countries) of the institutions of the authors of the selected primary papers were used to distinguish the geographical distribution of the research community members who were involved in the research. If a paper had authors from different countries, the country of the corresponding author was used. It is noted that the authors, universities, and companies in the United States and China were leading, having 4 (18%) papers each. This was followed by the United Kingdom with 3 (14%) papers and the Netherlands with 2 (9%) papers. The rest of the countries, namely Ukraine, India, Italy, Portugal, Switzerland, Denmark, Japan, France, and Canada, contributed 1 (5%) paper each. This geographical distribution of the 22 selected papers indicates that blockchain technology in the health care sector has gathered research interest around the world.

The channels where the papers were published determined the publication type. The two publication types included in this systematic review were conferences and journals. The majority (n=15, 68%) of the 22 selected primary papers were published in journals, while 7 (32%) were published as conference proceedings.

The 22 selected primary papers were studied, and the data or keywords related to this systematic review's question were extracted. A classification scheme was constructed based on the iterative identification of data, charting keywords extracted from the selected papers. The papers were then sorted into identified categories.

Each of the selected primary papers addressed one or more different aspects of the use cases of blockchain technology in the health care sector. Therefore, the identified use cases were used to further classify the selected papers. Out of 22 selected papers, 10 (45%) addressed the application of blockchain in the management of EMRs, 5 (23%) addressed the use of blockchain technology in biomedical research and education, and 3 (14%) demonstrated the use of blockchain technology in remote patient monitoring. The remaining papers addressed the use of blockchain technology in drug or pharmaceutical supply chains (n=1, 5%), health insurance claims (n=1, 5%), health data analytics (n=1, 5%), and other applications (n=1, 5%).

In the selected primary papers, blockchain was implemented in the real health care environment to address several information security components. The use of blockchain technology or the main reason it was implemented in health care was classified. From this data, it was noted that each of the selected primary papers addressed one or more reasons, out of a total of 34 reasons or benefits, for using blockchain technology in health care. Most papers addressed the application of blockchain in health care for data integrity (14/34, 41%). The next largest purpose of blockchain application was access control, which contributed 11 out of 34 (32%) reasons in the papers. Meanwhile, data logging was addressed 6 (18%) times, data versioning was addressed 2 (6%) times, and nonrepudiation was addressed 1 (3%) time.

The setting of the studies, specifically the type of hospital used for the implementation of blockchain technology, was analyzed among the 22 selected papers. Only 3 out of 22 (14%) papers gave the name of hospital where the study was carried out. Maslove et al [14] implemented a blockchain-based smart contract at Kingston General Hospital, Canada, to study how blockchain technology could be used in clinical trial data management, which could enhance data integrity. Another study on clinical trials data management was conducted by Benchoufi et al [15] at Hospital Hôtel Dieu, Paris, France, which looked at whether the implementation of blockchain could enhance the transparency and traceability of clinical trial consent, thereby benefitting both patients and researchers. Ichikawa et al [16] conducted their study at the Institute of Neuropsychiatry, Seiwa Hospital, Tokyo, Japan, using blockchain technology to implement a tamper-resistant mobile health (mHealth) system, which could enhance both data transparency and accessibility without the involvement of a third party.

Information regarding the blockchain platform that was used was gathered from the 22 selected papers. Ethereum was the most commonly used blockchain platform (n=12, 55%), followed by Hyperledger Fabric (n=3, 14%), while the least used platform was the New Economy Movement (NEM) blockchain platform (n=1, 5%). The rest of the studies did not state which blockchain platform was used.

Literature Quality Analysis

The final and crucial part of this systematic review involved reviewers scoring the 22 papers to evaluate their quality and the relevance of the blockchain usage. The scoring results are shown in Table 3 [14-35]. These show a greater quality of average overall scores among 15 journal papers (mean score 81.0%, SD 5.8%) compared to 7 conference papers (mean score 77.1%, SD 5.9%). Out of the 22 papers, 3 (14%) high-quality papers, 9 (41%) moderate-quality papers, and 10 (45%) low-quality papers were identified. It is noted that no published conference papers from January to September 2019 were found or included in this study.

The 2 (9%) papers published in 2016 each had a relatively high average overall score. In 2017, the average overall score of the papers was 80%, which was slightly lower than that of 2016. This dropped to 78% in 2018 and rose again to 80% in 2019.

Table 3. Summary of scores for measuring quality of the publications.

Study first author, year	Type of publication	Points per question (Q)[a]				Sum of scores for Q2-Q4	Overall score for Q1-Q4 (%), mean (SD)	Quality of paper[b]
		Q1	Q2	Q3	Q4			
Maslove, 2018 [14]	Journal	4	3	3	4	10	90.0 (0.5)	High
Benchoufi, 2017 [15]	Journal	4	3	2	3	8	80.0 (0.7)	Moderate
Ichikawa, 2017 [16]	Conference proceeding	4	4	2	3	9	85.0 (0.8)	Moderate
Azaria, 2016 [17]	Journal	4	3	2	3	8	80.0 (0.7)	Moderate
Cunningham, 2017 [18]	Journal	4	3	2	4	9	85.0 (0.8)	Moderate
Fan, 2018 [19]	Journal	4	3	1	3	7	75.0 (1.1)	Low
Li, 2018 [20]	Conference proceeding	4	3	2	4	9	85.0 (0.8)	Moderate
Liu, 2018 [21]	Journal	4	3	2	2	7	75.0 (0.8)	Low
Mendes, 2018 [22]	Journal	4	3	1	3	7	75.0 (1.1)	Low
Nagasubramanian, 2018 [23]	Journal	4	3	2	3	8	80.0 (0.7)	Moderate
Zhang, 2018 [24]	Journal	4	3	1	3	7	75.0 (1.1)	Low
Cichosz, 2018 [25]	Journal	4	4	2	3	9	85.0 (0.8)	Moderate
Omar, 2019 [26]	Conference proceeding	4	3	1	3	7	75.0 (1.1)	Low
Angeletti, 2017 [27]	Journal	4	4	2	4	10	90.0 (0.9)	High
Kleinaki, 2018 [28]	Journal	4	3	1	3	7	75.0 (1.1)	Low
Nugent, 2016 [29]	Journal	4	4	2	4	10	90.0 (0.9)	High
Liang, 2017 [30]	Conference proceeding	4	3	1	3	7	75.0 (1.1)	Low
Saravanan, 2017 [31]	Journal	4	3	1	3	7	75.0 (1.1)	Low
Bocek, 2017 [32]	Journal	4	4	1	4	9	85.0 (1.3)	Moderate
Zhou, 2018 [33]	Conference proceeding	4	4	1	3	8	80.0 (1.2)	Moderate
Kotsiuba, 2018 [34]	Conference proceeding	4	3	1	2	6	70.0 (1.1)	Low
Talukder, 2018 [35]	Conference proceeding	4	3	0	3	6	70.0 (1.5)	Low

[a]Two reviewers assessed each query from Q1 to Q4, based on a 5-point ordinal scale ranging from 0 to 4, where 0 indicates the lowest level (criterion was addressed very poorly or not at all) and 4 indicates the highest level (criterion was exceptional).

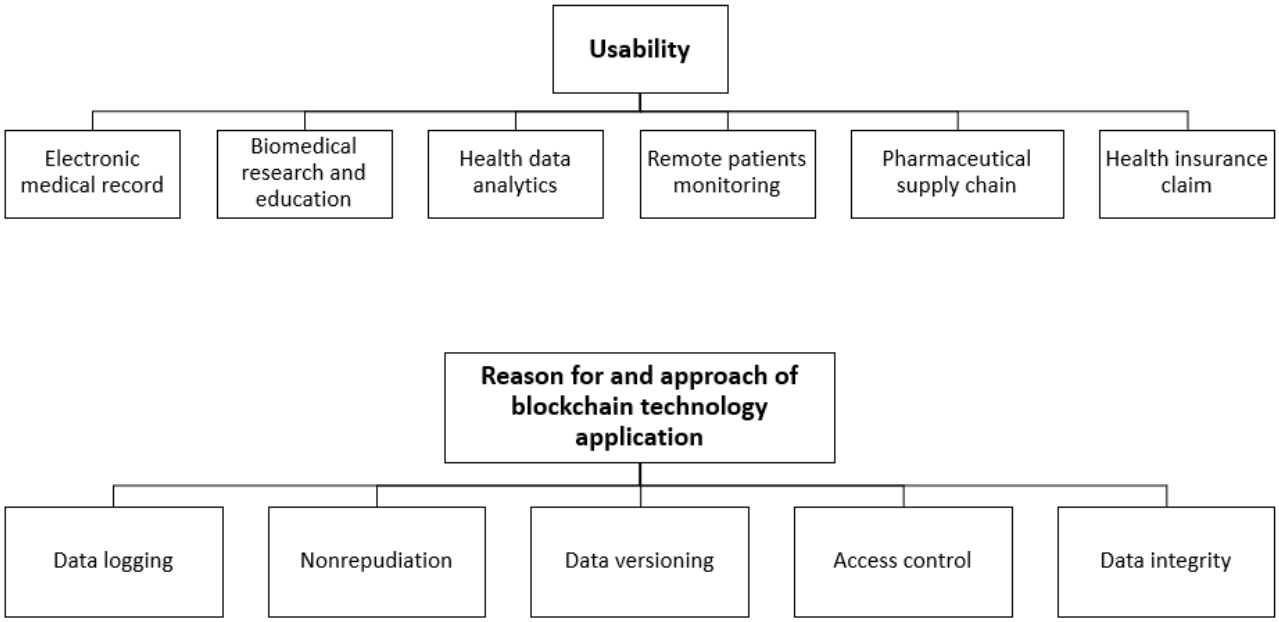
[b]An overall score of ≥90% indicates a high-quality paper; an overall score of 80%-89% indicates a moderate-quality paper; an overall score ≤79% indicates a low-quality paper.

Discussion

Identified Themes

The identified themes are summarized in Figure 2.

Figure 2. The uses of blockchain technology in the health care sector.



The Use of Blockchain Technology in Real Health Care Environments

The results from this systematic review show that the majority of the research regarding blockchain technology in health care environments was focused on the management of EMRs, followed by biomedical research and education, remote patient monitoring, pharmaceutical supply chains, health insurance claims, health data analytics, and other potential areas.

Electronic Medical Records

Out of 22 selected papers, 10 (45%) concentrated on the management of EMRs. EMRs, similar to EHRs or personal health records, involve electronic modeling, storage, and management of patients' personal, medical, or health-related data. Traditionally, different systems have been used to store patients' records separately across different service providers, where the service providers have control over the records, which may limit data sharing with other health care stakeholders.

The application of blockchain in the management of EHRs will make data sharing among health care stakeholders easier, more transparent, and more trustworthy, and patients will have control over their own data. This is because the characteristics of blockchain technology, such as decentralization, immutability, data provenance, reliability, robustness, smart contracts, security, and privacy, make it suitable for the management and storage of patient EHRs [18].

Azaria et al [17] presented MedRec, which is a project from the MIT (Massachusetts Institute of Technology) Media Lab and Beth Israel Deaconess Medical Center that uses a blockchain-based platform to give patients access to their own data through some access permissions built into the blockchain. The patient may decide to grant access to their EHRs to any third party, which may reduce their paperwork, given that patients normally have to carry a bundle of papers with them when they seek out different health care providers for consultation. With blockchain technology, regardless of the time and institution, health care providers can easily gain access to all of a patient's medical data. Patients become more committed to their own health care because they are directly involved in the management of their health records through blockchain technology.

The second application that would integrate EMRs is blockchain-based privacy-preserving data sharing (BPDS), which was developed by Liu et al [21]. This uses the Ethereum blockchain platform, which reduces the risk of medical data leakage and secures data sharing in health care.

Fan et al [19] developed MedBlock, a blockchain-based information management system that is implemented in health care to enhance efficiency and secure electronic medical data sharing using blockchain. Another blockchain-based EMR is FHIRChain [24], which encapsulates the Health Level Seven Fast Healthcare Interoperability Resources (FHIR) standard for shared clinical data. Zhang et al [24] used blockchain technology via the FHIRChain-based decentralized app to share clinical data that focused on health care record management and digital health identities to verify participants for remote cancer care in a case study of collaborative decision-making [24]. Cichosz et al [25] proposed NEM multi-signature blockchain contracts to be used for the management and sharing of medical data of diabetes patients, which aimed to achieve access control and data privacy.

Li et al [20] presented a medical data preservation system based on the real-world blockchain platform Ethereum, which provides a trustworthy storage solution to ensure the primitiveness and verifiability of stored data. Mendes et al [22] presented a Smart Ambient Assisted Living environment, which uses blockchain technology to enhance data privacy and cognitive security in the health care sector. As noted above, out of the 22 included papers, 3 (14%) were high-quality papers, 9 (41%) were moderate-quality papers, and 10 (45%) were low-quality papers.

Therefore, the conclusions made are convincing based on the high or moderate quality of a high percentage of papers.

Biomedical Research and Education: Clinical Research

A total of 23% (5/22) of the selected papers in this study indicated that blockchain could be applied in biomedical research and education fields. Blockchain technology has been used extensively in biomedical research and education to preserve data privacy, integrity, sharing, record sharing, and record keeping, especially in clinical trials [23]. Nugent et al [29] presented blockchain smart contracts that prevent falsification of data and underreporting of unwanted results of clinical research, which enhances trust in the data and clinical trials.

Angeletti et al [27] proposed a proof-of-concept implementation of consent traceability in clinical trials using Ethereum to secure and ensure unfalsifiable data. Every piece of data or consent included in the blockchain system is time-stamped and publicly transparent. This is achieved through cryptographic validation. All plans, consent, protocols, and possible outcomes can be stored on blockchain even before the inception of clinical trials, which prevents any corruption and undesirable study results.

Kleinaki et al [28] presented a blockchain-based notarization service that uses smart contracts to seal biomedical database queries and the respective results, which ensures data transparency. Maslove et al [14] proposed BlockTrial, a web-based interface system that allow users to run trial-related smart contracts on the Ethereum network in clinical data management, thereby enhancing the reliability and transparency of complex data in clinical trials. The tamperproof characteristics of blockchain prevents the manipulation of data in clinical trials.

Remote Patient Monitoring

Remote patient monitoring was another blockchain use case in the health care sector. Generally, remote patient monitoring includes the gathering of biomedical data from the body and mobile devices to enable the monitoring of patient status remotely outside of traditional health care environments, such as hospitals.

Liang et al [30] presented a Hyperledger-based implementation of blockchain in mHealth that enables data collection and sharing between health care stakeholders, ensuring both data transparency and accessibility. Saravanan et al [31] proposed an end-to-end secured system, a new health care paradigm (ie, Secured Mobile-Enabled Assisting Device for Diabetics), through smart contracts to facilitate cryptographic security and formalized data access in which to monitor diabetes patients. The author stated that blockchain was engaged in a mobile-enabled assisting device that was developed to monitor diabetes patients. Ichikawa et al [16] presented a tamper-resistant mHealth system using blockchain technology where a mobile device is used to gather EMRs, which are then sent to the blockchain-based Hyperledger Fabric network to ensure secure management of the data.

Cichosz et al [25] proposed NEM multi-signature blockchain contracts for assisting diabetes patients in monitoring and transmitting their vital parameters or data by sensor device to a blockchain-based platform where the data are collected, stored, and analyzed. In emergency cases, such as abnormal blood glucose levels or missing dosages, an alert via a social network, such as Facebook or WhatsApp, will be sent to the care provider. The data can be communicated continuously by using mobile devices as a gateway with blockchain technology, which could save patients from any untoward consequences.

Drug or Pharmaceutical Supply Chains

Drug or pharmaceutical supply chains are one of the use cases of blockchain technology in the health care sector, particularly health-related supply chain management. Drug or pharmaceutical supply chains involve the introduction of new drugs into the market, ensuring the safety and validity of medical products sold to end customers [32]. Blockchain has been applied in this field to allocate a safe and secure platform and to address the most common problems faced in the pharmaceutical industry, such as delivery of substandard or counterfeit medication, which may have a negative impact on patients.

In this systematic review, only 1 paper out of 22 (5%) presented the implementation of a blockchain-based application for pharmaceutical supply chain management. Bocek et al [32] presented a real-world demonstration and evaluation of blockchain technology in the pharmaceutical supply chain, where ambient temperature sensors with blockchain technology were used to record temperatures at which drugs were stored and transported; such temperature measurements were immutably kept in a public blockchain for transparent inspection, which could also decrease the operational cost in a pharmaceutical supply chain.

Health Insurance Claims

Health insurance is necessary for everyone to get affordable medical treatment. Blockchain's characteristics, such as immutability, decentralization, transparency, and auditability of records, can benefit the process of health insurance claims in the health care sector. Nevertheless, only 1 paper from the 22 (5%) selected primary papers focused on this application. Zhou et al [33] developed a blockchain-based medical insurance storage system that is displayed using the Ethereum blockchain platform. The medical insurance data of a patient can be encrypted and immutably stored on blockchain, which enhances credibility and eliminates the involvement of third parties in the management of patients' health insurance [33].

Health Data Analytics

Only 1 paper out of 22 (5%) presented the use of blockchain technology in health data analytics. Blockchain in collaboration with other emerging technologies, such as deep- and transfer-learning techniques, was used to identify predictive analytics of health care data. Kotsiuba et al [34] stated that blockchain provides a unique opportunity to overcome the problems related to the analysis and security of medical data. Using blockchain technology, a decentralized health data ecosystem was presented that protected medical data confidentiality, produced an effective shared health infrastructure, and increased the basis of clinical data collection.

Other Functionalities

Of the 22 selected primary papers, 1 (5%) study by Talukder et al [35] included the relevant research perspective but could not be classified under any of the identified uses of blockchain. This study presented a blockchain consensus protocol that provides accurate medical decisions and reduces the disease burden by using Ethereum, based on the proof-of-disease consensus protocol. All functionalities of health data interoperability, including EMRs, patient health records, and health information exchange databases, can be achieved by this system.

Reasons for the Application of Blockchain Technology in the Current Health Care System

Overview

The main reasons (N=34) for the application or implementation of blockchain technology in the health care sector in this study's selected papers were identified and categorized into the following groups: data integrity (n=14, 41%), access control (n=11, 32%), data logging (n=6, 18%), data versioning (n=2, 6%), and nonrepudiation (n=1, 3%).

Data Integrity

Data integrity is defined as the accuracy and consistency of the data or information stored in a system, which acts as an important component of information security. Data integrity was achieved by using blockchain technology in the health care sector. Li et al [20] implemented the blockchain-based platform Ethereum to maintain the originality and variability of stored data in the system while preserving user privacy. The lifelong maintenance of data in blockchain was achieved with the proof-of-primitiveness data concept, and the system can validate the data where it is identical to the original data. The data can be restored and verified through blockchain if it has been damaged.

Kotsiuba et al [34] also presented a decentralized health data system using blockchain that secures the collection and confidentially of medical or clinical data. In the study by Zhang et al [24], data integrity was enhanced by using an FHIRChain-based decentralized app, which used blockchain technology and digital health identities in remote cancer care to validate the participants in a case study of clinical data sharing. With the application of public key cryptography, this decentralized app improves the trust of participants and enables the users to share specific and structured pieces of information, rather than an entire document. Thereby, it increases the readability of data and flexibility of sharing options.

Cichosz et al [25] implemented a blockchain-based platform to enhance the management and sharing of diabetes data in an easy and secure way, which can be achieved by a decentralization of blockchain. According to a study by Omar et al [26], the integrity, security, privacy, and accountability of data in health care are achieved through a privacy-preserving platform using blockchain technology. To ensure encryption of patient data and pseudonymity, a cryptographic function was used. The decentralization of data enabled by the peer-to-peer network in blockchain technology helps to reduce cyberattacks and preserve the health care data set.

The proof-of-concept implementation of patient-facing and researcher-facing systems using blockchain technology to enhance data integrity was demonstrated by Maslove et al [14] and Angeletti et al [27]. Maslove et al [14] demonstrated that the proof-of-concept implementation using blockchain technology in clinical trials secures original personal data, and this data would not be shared publicly before an agreement is reached. In regard to the use of blockchain technology in clinical trials, the clinical research institute can also guarantee that the data obtained are authentic and useful.

Angeletti et al [27] stated that the integrity of the data collected in clinical trials was enhanced by the application of blockchain technology, specifically blockchain-based smart contracts, which act as the foundation to promote trust throughout clinical research. The proof-of-concept implementation in clinical research enhances the interaction of researchers and patients.

Blockchain in digital health technologies has also been particularly used in mHealth, which includes remote patient monitoring to ensure the safe and precise preservation of medical information to improve data integrity. Ichikawa et al [16] concluded from their study that the usage of blockchain technology in mHealth improves data transparency and accessibility without the involvement of third parties, due to the tamperproof and decentralized characteristics of blockchain technology.

Access Control

According to Azaria et al [17], access control is defined as an individual having full authority in deciding who can access their medical data, as well as when and how much of their own medical data can be accessed using blockchain technology. Access control may lead to patients' direct involvement in controlling their own medical data usage. The distributed ledger, which is one of the characteristics of blockchain technology, ensures efficient access and retrieval of EMRs [18].

Fan et al [19] used a proof of concept with an application programming interface using blockchain technology, which allows a permission system where each patient is able to view, control, and specify who can access their records.

A study by Cunningham and Ainsworth [18] found that the EMRs that included a patient's full medical history from many different hospitals could be easily accessed by the patient using a blockchain-based information management system, which enhances the outcome of treatment by avoiding the segregation of medical data from different hospitals. An access protocol was implemented that prevented unauthorized users from obtaining any sensitive data or information.

For access control and the preservation of data, Fan et al [19] used the blockchain-based platform concept with NEM multi-signature blockchain contracts, which ensured privacy control of health data. With this concept, patients are in control of their own data and have the power to decide who can access their personal data. For instance, an older adult patient could share access of their medical data with their adult child.

According to Cichosz et al [25], patients were able to access their own medical data through smart blockchain contracts, which may lead to secure data sharing. Through BPDS, which consists of data access permission implemented by Liu et al [21], patients have full control over their medical records or data, without jeopardizing their privacy. Furthermore, a user can use patient data with permission from the patient. The owner of the data in blockchain is capable of revoking his or her access permission, in case of a violation of access rules.

According to Liu et al [21], health records that are centrally stored are more vulnerable to cyberattacks. Therefore, Nagasubramanian et al [23] presented a keyless signature infrastructure (KSI) blockchain technology for securing EHRs that ensures authentication and integrity of health records. In a KSI blockchain system, the signed data are stored and can be operated without a network connection, and no third parties are required to preserve data in this system.

Data access by health care professionals can be achieved through smart contracts of blockchain technology with cryptographic security. Using blockchain technology, Saravanan et al [31] implemented a mobile-based secure health care system that can predict a patient's diabetes status in real time. In case of emergency, the doctor can access a patient's health record and prescribe them with a suitable medication dosage using this technology system. This blockchain system is used to store data related to health care and securely connect with third parties.

Data Logging

Data logging is defined as an operation of gathering and storing information over a period of time. It allows tracking of all types of interactions, such as storage, access, or modification of data, files, or applications in a system. Data logging can be achieved by the application of blockchain technology in the health care sector.

In clinical trials, Nugent et al [29] demonstrated blockchain technology using an Ethereum smart contract to enhance the trustworthiness, reliability, and transparency of data management. The cryptographic and tamperproof characteristics of blockchain prevent all forms of manipulation and enhance the data logging of complex clinical trial data management, so more informed decisions can be made by medical professionals. An mHealth care system using blockchain technology was implemented by Liang et al [30] that ensured gathering, sharing, and collaboration of data between the health care providers and individuals in a secure way.

Bocek et al [32] stated in their study that the application of blockchain technology in pharmaceutical supply chain management ensures data logging. They demonstrated the use of an Internet of Things sensor device (modum.io AG) that uses blockchain technology to ensure the verification of compliance with quality control temperature requirements. This device was used to monitor and store the temperature of products, enhance data immutability, and facilitate public accessibility of temperature records of pharmaceutical products, especially during transportation. Data provenance was ensured using blockchain technology that can prove the origin of products in a supply chain.

Zhou et al [33] stated that blockchain technology acts as a tamperproof and decentralized technology to record data, which enhances users' trust in a health insurance system, especially with the implementation of a blockchain-based medical insurance storage system. For instance, the data about each patient's spending was stored and secured in the blockchain by the hospital, which helped the insurance company obtain information about the total amount of spending by the patient; however, third parties, including the insurance company, cannot modify or delete the data and do not have the authority to access a patient's personal medical data.

Data Versioning

Data versioning is defined as saving new copies of the data when any modification is made to the existing data. This helps to keep track of the data and ensure easy retrieval of any specific version of the respective stored data in a system. Kleinaki et al [28] implemented a blockchain-based notarization service that uses smart digital contracts to secure data in the biomedical research sector. A study by Mendes et al [22] showed that after the retrieval process, retrieved data cannot be modified, which ensures the integrity and nonrepudiation of the data. Using blockchain technology, data versioning was achieved where medical evidence of different versions of data retrieved from a biomedical database were securely stored and saved, along with content that is continually updated. In this study, this was mostly used for decision support in the health care sector.

Nonrepudiation

Nonrepudiation guarantees the validity of data in a particular health care system, which cannot not be denied by anyone and ensures the originality and integrity of data. A study by Angeletti et al [27] used blockchain technology to collect, store, and track clinical trial consent in a secure, unfalsifiable, and publicly verifiable way; this consent was originally time-stamped with the application of proof of concept, leading to the nonrepudiation of data. The authentication system ensures that the clinical trial consent is accessible and transparent for patients, while traceable for stakeholders. A single document in open format was used and accounted for the whole time-stamped consent collection process. This document cannot be corrupted and is considered a robust proof of data.

Study Limitations

One limitation of this systematic review study was that there were no published studies on the safety of blockchain technology in health care, so the safety aspect of blockchain technology cannot be reviewed. In addition, there were few papers published on the negative aspects of implementation of blockchain technology in health care. Most studies only published the positive aspects, which may have led to bias.

Future Directions

Blockchain technology is still a new technology that has not been widely implemented in the health care sector. This study can be a guide for future research, implementation, and evaluation of blockchain technology in this sector. More research should be carried out regarding the implementation of blockchain technology in real health care environments for better understanding, characterization, and evaluation.

Researchers should also focus on carrying out research on the safety of implementing blockchain technology in health care.

Conclusions

This systematic review has presented an overview of the use and characteristics of blockchain technology in the health care sector. The findings show that blockchain technology research and application in the health care sector is still in its infancy but growing rapidly. Blockchain technology has started to develop from cryptocurrencies, such as Bitcoin, into various general-purpose technologies in many industries, including health care. According to the selected papers in this study, EMRs, biomedical research and education, remote patient monitoring, drug or pharmaceutical supply chains, health insurance claims, and health data analytics are the most common uses of blockchain technology in health care. The main reasons for the application of blockchain technology are to enhance data integrity, access control, logging, data versioning, and nonrepudiation of patient health records or other health information in health care settings.

Authors' Contributions

DE, LCM, CSL, FSB, and CST contributed to the study design, data extraction, quality assessment, analysis and interpretation of data, and drafting of the manuscript. KWG and SFY contributed to the study design, quality assessment, analysis and interpretation of data, and revision of the paper. KSL, ZH, ACI, and MJL contributed to the study design, data extraction, and analysis and interpretation of data. All authors proofread and approved the submitted version of the paper. The paper contents have not been previously presented elsewhere. All authors have read and approved this manuscript for publication and report no financial disclosures.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Databases and search terms used.
[DOCX File , 14 KB - medinform_v10i1e17278_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist for the systematic review of the use of blockchain technology in the health care sector.
[DOCX File , 29 KB - medinform_v10i1e17278_app2.docx]

References

1. Norris AC. Current trends and challenges in health informatics. Health Informatics J 2016 Jul 25;8(4):205-213. [doi: 10.1177/146045820200800407]
2. Nguyen Q, Naguib R, Ghani MA, Bali R, Marshall I, Phuong N, et al. An analysis of the healthcare informatics and systems in Southeast Asia: A current perspective from seven countries. Int J Electron Healthc 2008;4(2):184. [doi: 10.1504/ijeh.2008.019792]
3. Georgiou A. Health informatics and evidence-based medicine - More than a marriage of convenience? Health Informatics J 2016 Sep 14;7(3-4):127-130. [doi: 10.1177/146045820100700303]
4. Ismail NI, Abdullah NH, Shamsuddin A. Adoption of hospital information system (HIS) in Malaysian public hospitals. Procedia Soc Behav Sci 2015 Jan;172:336-343. [doi: 10.1016/j.sbspro.2015.01.373]
5. Peterson K, Deeduvanu R, Kanjamala P, Boles K. A blockchain-based approach to health information exchange networks. HealthIT.gov. 2016. URL: https://www.healthit.gov/sites/default/files/12-55-blockchain-based-approach-final.pdf [accessed 2019-08-08]
6. Zhang P, White J, Schmidt DC, Lenz G. Applying software patterns to address interoperability challenges in blockchain-based healthcare apps. ArXiv Preprint posted online on June 5, 2017 [FREE Full text]
7. Böhme R, Christin N, Edelman B, Moore T. Bitcoin: Economics, technology, and governance. J Econ Perspect 2015 May 01;29(2):213-238. [doi: 10.1257/jep.29.2.213]
8. Dwyer GP. The economics of Bitcoin and similar private digital currencies. J Financ Stab 2015 Apr;17:81-91. [doi: 10.1016/j.jfs.2014.11.006]
9. Mettler M. Blockchain technology in healthcare: The revolution starts here. In: Proceedings of the 18th International Conference on e-Health Networking, Applications and Services. 2016 Presented at: 18th International Conference on e-Health Networking, Applications and Services; Sepember 14-17, 2016; Munich, Germany. [doi: 10.1109/healthcom.2016.7749510]
10. Yli-Huumo J, Ko D, Choi S, Park S, Smolander K. Where is current research on blockchain technology?—A systematic review. PLoS One 2016;11(10):e0163477 [FREE Full text] [doi: 10.1371/journal.pone.0163477] [Medline: 27695049]

11. Dybå T, Dingsøyr T. Empirical studies of agile software development: A systematic review. Inf Softw Technol 2008 Aug;50(9-10):833-859. [doi: 10.1016/j.infsof.2008.01.006]
12. Petersen K, Feldt R, Mujtaba S, Mattsson M. Systematic mapping studies in software engineering. In: Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering. 2008 Presented at: 12th International Conference on Evaluation and Assessment in Software Engineering; June 26-27, 2008; Bari, Italy. [doi: 10.14236/ewic/ease2008.8]
13. Hölbl M, Kompara M, Kamišalić A, Nemec Zlatolas L. A systematic review of the use of blockchain in healthcare. Symmetry 2018 Oct 10;10(10):470. [doi: 10.3390/sym10100470]
14. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: A proof-of-concept study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
15. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: Transparency and traceability of consent. F1000Res 2017;6:66 [FREE Full text] [doi: 10.12688/f1000research.10531.5] [Medline: 29167732]
16. Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
17. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: Using blockchain for medical data access and permission management. In: Proceedings of the 2nd International Conference on Open and Big Data. 2016 Presented at: 2nd International Conference on Open and Big Data; August 22-24, 2016; Vienna, Austria p. 22-24. [doi: 10.1109/obd.2016.11]
18. Cunningham J, Ainsworth J. Enabling patient control of personal electronic health records through distributed ledger technology. Stud Health Technol Inform 2017;245:45-48. [Medline: 29295049]
19. Fan K, Wang S, Ren Y, Li H, Yang Y. MedBlock: Efficient and secure medical data sharing via blockchain. J Med Syst 2018 Jun 21;42(8):136. [doi: 10.1007/s10916-018-0993-7] [Medline: 29931655]
20. Li H, Zhu L, Shen M, Gao F, Tao X, Liu S. Blockchain-based data preservation system for medical data. J Med Syst 2018 Jun 28;42(8):141. [doi: 10.1007/s10916-018-0997-3] [Medline: 29956058]
21. Liu J, Li X, Ye L, Zhang H, Du X, Guizani M. BPDS: A blockchain based privacy-preserving data sharing for electronic medical records. In: Proceedings of the 2018 IEEE Global Communications Conference. 2018 Presented at: 2018 IEEE Global Communications Conference; December 9-13, 2018; Abu Dhabi, United Arab Emirates p. 9-13. [doi: 10.1109/glocom.2018.8647713]
22. Mendes D, Rodrigues IP, Fonseca C, Lopes MJ, García-Alonso JM, Berrocal J. Anonymized distributed PHR using blockchain for openness and non-repudiation guarantee. Stud Health Technol Inform 2018;255:170-174. [Medline: 30306930]
23. Nagasubramanian G, Sakthivel RK, Patan R, Gandomi AH, Sankayya M, Balusamy B. Securing e-health records using keyless signature infrastructure blockchain technology in the cloud. Neural Comput Appl 2018 Nov 30;32(3):639-647. [doi: 10.1007/s00521-018-3915-1]
24. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: Applying blockchain to securely and scalably share clinical data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
25. Cichosz SL, Stausholm MN, Kronborg T, Vestergaard P, Hejlesen O. How to use blockchain for diabetes health care data and access management: An operational concept. J Diabetes Sci Technol 2019 Mar;13(2):248-253 [FREE Full text] [doi: 10.1177/1932296818790281] [Medline: 30047789]
26. Omar AA, Bhuiyan MZA, Basu A, Kiyomoto S, Rahman MS. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Gener Comput Syst 2019 Jun;95:511-521. [doi: 10.1016/j.future.2018.12.044]
27. Angeletti F, Chatzigiannakis I, Vitaletti A. The role of blockchain and IoT in recruiting participants for digital clinical trials. In: Proceedings of the 25th International Conference on Software, Telecommunications and Computer Networks. 2017 Presented at: 25th International Conference on Software, Telecommunications and Computer Networks; September 21-23, 2017; Split, Croatia. [doi: 10.23919/SOFTCOM.2017.8115590]
28. Kleinaki A, Mytis-Gkometh P, Drosatos G, Efraimidis PS, Kaldoudi E. A blockchain-based notarization service for biomedical knowledge retrieval. Comput Struct Biotechnol J 2018;16:288-297 [FREE Full text] [doi: 10.1016/j.csbj.2018.08.002] [Medline: 30181840]
29. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. F1000Res 2016;5:2541 [FREE Full text] [doi: 10.12688/f1000research.9756.1] [Medline: 28357041]
30. Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. In: Proceedings of the 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications. 2017 Presented at: 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications; October 8-13, 2017; Montreal, QC. [doi: 10.1109/PIMRC.2017.8292361]
31. Saravanan M, Shubha R, Mary A. SMEAD: A secured mobile enabled assisting device for diabetics monitoring. In: Proceedings of the 11th IEEE International Conference on Advanced Networks and Telecommunications Systems. 2018 Presented at: 11th IEEE International Conference on Advanced Networks and Telecommunications Systems; December 17-20, 2017; Bhubaneswar, India. [doi: 10.1109/ants.2017.8384099]
32. Bocek T, Rodrigues B, Strasser T, Stiller B. Blockchains everywhere - A use-case of blockchains in the pharma supply-chain. In: Proceedings of the IFIP/IEEE Symposium on Integrated Network and Service Management. 2017 Presented at: IFIP/IEEE

Symposium on Integrated Network and Service Management; May 8-12, 2017; Lisbon, Portugal. [doi: 10.23919/inm.2017.7987376]
33. Zhou L, Wang L, Sun Y. MIStore: A blockchain-based medical insurance storage system. J Med Syst 2018 Jul 02;42(8):149 [FREE Full text] [doi: 10.1007/s10916-018-0996-4] [Medline: 29968202]
34. Kotsiuba I, Velvkzhanin A, Yanovich Y, Bandurova I, Dyachenko Y, Zhygulin V. Decentralized e-Health architecture for boosting healthcare analytics. In: Proceedings of the 2nd World Conference on Smart Trends in Systems, Security and Sustainability. 2019 Presented at: 2nd World Conference on Smart Trends in Systems, Security and Sustainability; October 30-31, 2018; London, UK. [doi: 10.1109/worlds4.2018.8611621]
35. Talukder A, Chaitanya M, Arnold D, Sakurai K. Proof of disease: A blockchain consensus protocol for accurate medical decisions and reducing the disease burden. In: Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovations. 2018 Presented at: 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovations; October 7-11, 2018; Guangzhou, China. [doi: 10.1109/smartworld.2018.00079]

Abbreviations

BPDS: blockchain-based privacy-preserving data sharing
EHR: electronic health record
EMR: electronic medical record
FHIR: Fast Healthcare Interoperability Resources
HI: health informatics
HIS: hospital information system
IT: information technology
KSI: keyless signature infrastructure
MeSH: Medical Subject Headings
mHealth: mobile health
MIT: Massachusetts Institute of Technology
NEM: New Economy Movement
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Q1: question 1
Q2: question 2
Q3: question 3
Q4: question 4

Edited by G Eysenbach; submitted 02.12.19; peer-reviewed by F Riaz, W Zhang, T Ueno, C Esposito; comments to author 24.04.20; revised version received 12.11.20; accepted 28.09.21; published 20.01.22.

Please cite as:
Elangovan D, Long CS, Bakrin FS, Tan CS, Goh KW, Yeoh SF, Loy MJ, Hussain Z, Lee KS, Idris AC, Ming LC
The Use of Blockchain Technology in the Health Care Sector: Systematic Review
JMIR Med Inform 2022;10(1):e17278
URL: https://medinform.jmir.org/2022/1/e17278
doi:10.2196/17278
PMID:35049516

Review

Digital Health Interventions to Enhance Prevention in Primary Care: Scoping Review

Van C Willis[1*], PhD; Kelly Jean Thomas Craig[1*], PhD; Yalda Jabbarpour[2], MD; Elisabeth L Scheufele[1], MS, MD; Yull E Arriaga[1], MD; Monica Ajinkya[2], MD; Kyu B Rhee[1], MPP, MD; Andrew Bazemore[3], MPH, MD

[1]Center for Artificial Intelligence, Research, and Evaluation, IBM Watson Health, Cambridge, MA, United States

[2]Policy Studies in Family Medicine and Primary Care, The Robert Graham Center, American Academy of Family Physicians, Washington, DC, United States

[3]The American Board of Family Medicine, Lexington, KY, United States

[*]these authors contributed equally

Corresponding Author:
Kelly Jean Thomas Craig, PhD
Center for Artificial Intelligence, Research, and Evaluation
IBM Watson Health
75 Binney St
Cambridge, MA, 02142
United States
Phone: 1 9702613366
Email: kelly.jean.craig@ibm.com

Abstract

Background: Disease prevention is a central aspect of primary care practice and is comprised of primary (eg, vaccinations), secondary (eg, screenings), tertiary (eg, chronic condition monitoring), and quaternary (eg, prevention of overmedicalization) levels. Despite rapid digital transformation of primary care practices, digital health interventions (DHIs) in preventive care have yet to be systematically evaluated.

Objective: This review aimed to identify and describe the scope and use of current DHIs for preventive care in primary care settings.

Methods: A scoping review to identify literature published from 2014 to 2020 was conducted across multiple databases using keywords and Medical Subject Headings terms covering primary care professionals, prevention and care management, and digital health. A subgroup analysis identified relevant studies conducted in US primary care settings, excluding DHIs that use the electronic health record (EHR) as a retrospective data capture tool. Technology descriptions, outcomes (eg, health care performance and implementation science), and study quality as per Oxford levels of evidence were abstracted.

Results: The search yielded 5274 citations, of which 1060 full-text articles were identified. Following a subgroup analysis, 241 articles met the inclusion criteria. Studies primarily examined DHIs among health information technologies, including EHRs (166/241, 68.9%), clinical decision support (88/241, 36.5%), telehealth (88/241, 36.5%), and multiple technologies (154/241, 63.9%). DHIs were predominantly used for tertiary prevention (131/241, 54.4%). Of the core primary care functions, comprehensiveness was addressed most frequently (213/241, 88.4%). DHI users were providers (205/241, 85.1%), patients (111/241, 46.1%), or multiple types (89/241, 36.9%). Reported outcomes were primarily clinical (179/241, 70.1%), and statistically significant improvements were common (192/241, 79.7%). Results were summarized across the following 5 topics for the most novel/distinct DHIs: population-centered, patient-centered, care access expansion, panel-centered (dashboarding), and application-driven DHIs. The quality of the included studies was moderate to low.

Conclusions: Preventive DHIs in primary care settings demonstrated meaningful improvements in both clinical and nonclinical outcomes, and across user types; however, adoption and implementation in the US were limited primarily to EHR platforms, and users were mainly clinicians receiving alerts regarding care management for their patients. Evaluations of negative results, effects on health disparities, and many other gaps remain to be explored.

(JMIR Med Inform 2022;10(1):e33518) doi:10.2196/33518

KEYWORDS

digital technology; primary health care; preventive medicine; telemedicine; clinical decision support systems

Introduction

The Institute of Medicine declared primary care to be "essential health care" and the central feature of an effective health care system [1]. Primary care has the potential to enhance quality, reduce costs, and increase equity and access to care by providing first contact and easy access to comprehensive, continuous, and coordinated medical care for patients [2] and populations, as articulated in the 4Cs framework by Dr Barbara Starfield [3]. Prevention of diseases and their complications ranks among primary care's most fundamental functions; when performed effectively, primary care prevention can decrease mortality and morbidity in both chronic and acute conditions [4]. Various practitioners, including physicians, nurses, physician assistants, and pharmacists, recognize its value, but preventive services are often underutilized [5], despite guideline recommendations provided by the US Preventive Services Task Force [6].

Many studies have investigated the sources of suboptimal preventive health service delivery. Among the major barriers to preventive care implementation by clinicians is time. Studies have shown that 8.6 hours per working day are needed for a clinician to fully satisfy the US Preventive Services Taskforce preventive care recommendations for their patients [7]. A steady growth in competing demands across the management of acute, chronic, and preventive needs and an aging population with increasing comorbidities make it nearly impossible for a clinician to provide recommended preventive services without support. Innovations in care delivery, such as the patient-centered medical home [8], use of community health workers [9], and integration of primary care with public health [10], can help reduce this burden on clinicians, but with the rapid evolution of information technology, digital health interventions (DHIs) to address prevention are crucial.

DHIs are delivered via digital technologies to support a variety of health system needs and are used both formally and informally by providers, patients, and population stakeholders. Examples of these technologies include mobile wireless health devices (mobile health [mHealth]) using SMS or smartphone apps, telehealth systems for remote clinical services, wireless medical devices, software as a medical device (eg, clinical decision support), medical imaging, health information technology (HIT), and patient portals. Other digital health facets, such as advanced data analytics and artificial intelligence (AI), may be used as standalone interventions or integrated components within digital technologies. Digital health technologies may or may not be regulated by the US Food and Drug Administration (FDA) or recognized by the World Health Organization (WHO).

DHIs can support primary (eg, timely receipt of vaccinations), secondary (eg, completion of indicated screenings), tertiary (eg, routine monitoring of chronic conditions), and quaternary (eg, prevention of overmedicalization) prevention. DHIs have provided meaningful outcomes via the incorporation of care management programs, disease registries, and behavioral change interventions to improve medication adherence, promote weight loss, support smoking and substance abuse cessation, and enhance mental health [11]. Moreover, DHIs have been effectively used to address racial, ethnic, and socioeconomic health disparities [12]. In addition, the COVID-19 pandemic has accelerated the adoption of DHIs, such as telehealth services, and raised the possibility of longer-term incorporation of such technologies by a primary care community that has traditionally lagged hospital and acute care peers.

Although prior studies have examined the impact of individual DHIs on preventive service receipt, no comprehensive review of these modalities exists to date. A scoping review with a subgroup analysis was conducted to understand how DHIs are being used in US primary care settings to enhance and support the delivery of preventive care.

Methods

Study Design

A scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [13] to identify studies that examined patients/consumers, providers, and/or population stakeholders in primary care settings (eg, limited to outpatient, ambulatory care, and long-term care) that used at least one digital health technology as an intervention for prevention (primary [eg, timely receipt of vaccinations], secondary [eg, completion of indicated screenings], tertiary [eg, routine monitoring of chronic conditions], and quaternary [eg, prevention of overmedicalization]) and reported beneficial outcomes on health, health care performance, and implementation science. The protocol is available upon request.

Search Strategy

Systematic search queries of MEDLINE via PubMed, Embase, and the Cochrane Library were used to identify references published or available online between January 1, 2014, and July 19, 2020 (Multimedia Appendices 1-7). Studies were limited to primary designs or systematic reviews (with the same inclusion criteria) published in English with abstracts. The rationale for this search cutoff time frame was based upon a high threshold of eligible providers achieving meaningful use of certified electronic health record (EHR) technology, whereby 82.8% of office-based physicians had adopted any EHR [14].

Screening Process

To ensure screener alignment, dual review of 20% of randomized titles and abstracts was followed by group resolution of conflicts. All remaining titles and abstracts underwent single review, and full-text articles were examined by 2 independent reviewers for relevance against the inclusion/exclusion criteria (Multimedia Appendix 8), with third-party adjudication provided for any discrepancies in eligibility. Results were tracked in DistillerSR (Evidence Partners).

The eligible population criteria included studies that examined patients/consumers, providers (both licensed and unlicensed), and/or population stakeholders (eg, payers, employers, communities, health systems, and the government) in outpatient care, ambulatory care, and long-term settings of primary care. Interventions had to target primary, secondary, tertiary, or quaternary prevention using at least one FDA/WHO approved or nonregulated digital health technology facet (eg, telehealth, mHealth, HIT, data analytics, and AI). No comparisons were required. Outcomes of interest included health (eg, individual- or population-level outcomes), health care performance (eg, as per the Agency for Healthcare Research and Quality [AHRQ]: access, quality, utilization, and efficiency, with measures categorized as structural, process, or outcomes including clinical/physiological, surrogate/intermediate, patient-centered, or patient-reported), and DHI implementation (eg, taxonomy as per Proctor et al: acceptability, adoption, appropriateness, costs, feasibility, fidelity, penetration, and sustainability) [15]. Only English-language primary studies or systematic reviews with the same inclusion criteria published between January 2014 and July 2020 were included. For definitions and descriptions of terms, see Multimedia Appendix 8 and Multimedia Appendix 9. Notable exclusion criteria for interventions included DHIs associated with treatment or diagnosis (except for preventive screenings), medical imaging for diagnosis, and telehealth using only noncellular telephone communication. Studies conducted in critical care (eg, intensive care unit) or inpatient (eg, hospital admission) settings were excluded.

Data Extraction

After a series of data form piloting and discussions by all extractors to identify gaps in data extraction forms and ensure consistency in the application of definitions, data were abstracted into standardized forms within DistillerSR (Multimedia Appendix 10) for synthesis by a single reviewer. All fields of the data extraction forms for each article were examined for completeness by a second reviewer. Many data categorizations were not mutually exclusive, resulting in percentages totaling more than 100%.

Subgroup Analysis and Data Synthesis

Following title and abstract screening, the large scope (>1000 titles) of the remaining included studies prohibited full-text review of all preventive DHIs identified globally. To narrow the scope of the geography and interventions under review, a subgroup analysis was performed; geography limits were set to only include studies conducted in the US. Additionally, it was apparent that a large volume of records focused on data analysis methods tangential to the development of DHIs. As such, studies that only used EHRs as a retrospective data capture tool were excluded. Two examples of excluded studies are a retrospective analysis of EHRs to determine the prevalence of a preventable disease and a study on the use of diagnostic telemedicine referral to a dermatologist.

Content analysis of extracted technology descriptions was performed to identify recurrent topics and more clearly understand the types of DHIs evaluated in the included studies according to *a priori* research questions in the protocol. This analysis yielded a list of articles selected to represent innovative or unique DHIs and their implementation in the final data set. Selected technologies were then narratively synthesized into 5 topical groups (eg, population-centered, patient-centered, care access expansion, panel-centered [dashboarding], and app-driven) to provide a framework for their analysis. Selected outcome (eg, health, health care performance, and implementation science) results from these articles were then extracted by a single reviewer to provide additional context regarding the impact of these DHIs beyond the directionality of their results. Details presented from this synthesis are not exhaustive, and key use cases have been highlighted in the results.

Study Quality

Study quality was assessed using the Oxford levels of evidence [16], which allow for the categorization of evidence quality across heterogeneous study types. Examples of the study types comprising these evidence levels include (in increasing quality) expert opinion, case series, systematic reviews of case-control studies, individual cohort studies, randomized controlled trials (RCTs) with narrow confidence intervals, and systematic reviews of RCTs.

Results

Literature searches yielded 5274 unique citations, of which 1060 articles were eligible for full-text screening. A subgroup analysis was conducted to limit geography to US–only settings and exclude DHIs that evaluated EHRs as retrospective data capture tools. These applied limits resulted in 310 articles for full-text review, of which 241 articles [17-257] were included for the subgroup analysis (Figure 1). Abstractions of the included articles can be found in Multimedia Appendix 11. An overview of the study design and key findings is provided in Figure 2. The types of DHI articles covered included HIT (166/241, 68.9%), clinical decision support (88/241, 36.5%), telehealth (88/241, 36.5%), mHealth (35/241, 14.5%), patient portals (16/241, 6.6%), wireless medical devices (6/241, 2.5%), medical imaging (2/241, 0.8%), and other DHIs (31/241, 12.9%) (see Multimedia Appendix 9 for a description of each). The integration of multiple types of technologies was commonly applied to support DHIs (154/241, 63.9%) in practice. The most commonly identified combination of technology was the use of clinical decision support algorithms and mHealth to support more advanced care using HIT-related data.

Figure 1. The flow diagram illustrates the flow of information through the different phases of the scoping review, including the number of records identified, included and excluded records, and the reasons for exclusion.

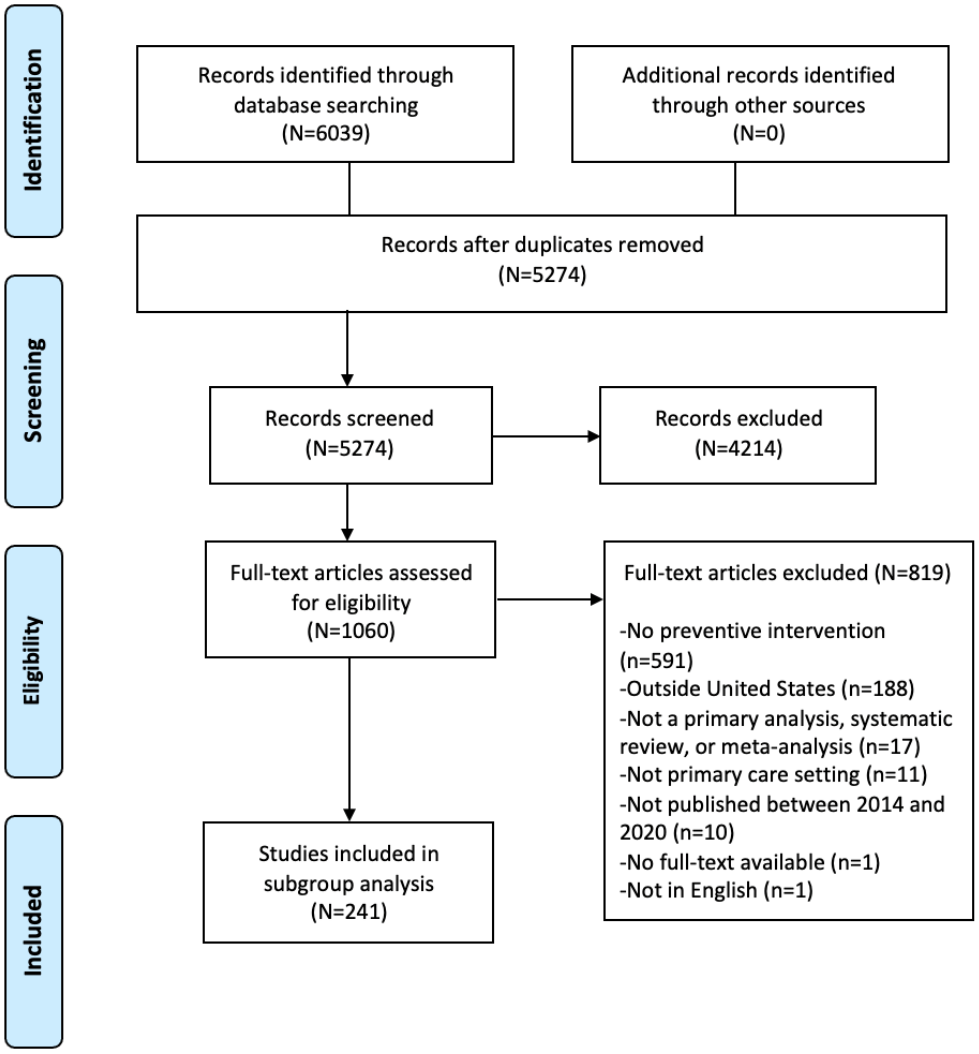
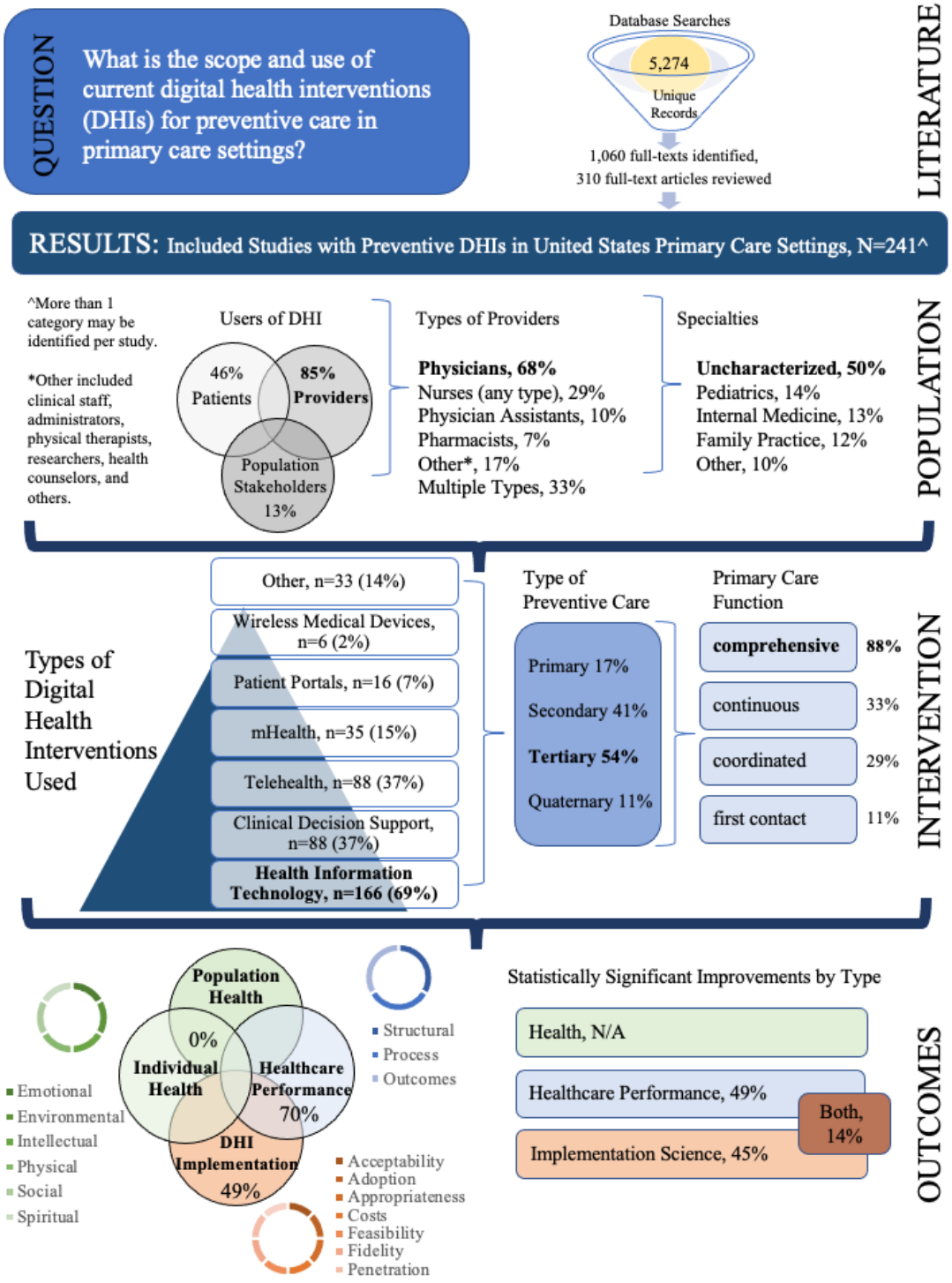


Figure 2. Summary of the study design and key findings. Scoping review study design and summarization of results across the categories of study population, intervention, and outcomes. N/A, not applicable.



The DHIs predominantly addressed tertiary prevention (131/241, 54.4%), followed by secondary (97/241, 40.3%), primary (40/241, 16.6%), and quaternary prevention (27/241, 11.2%), and a combination of prevention levels (43/241, 17.8%). The 4Cs primary care model by Dr Starfield was used as a framework to identify how DHIs supported delivery of preventive care; a large number of articles evaluated DHIs that demonstrated improvements in comprehensiveness of care (213/241, 88.4%), continuous care (76/241, 31.5%), coordinated care (69/241, 28.6%), and first contact care (26/241, 10.8%). The continuum of comprehensive care by DHIs included proactive anticipatory care, self-management support for patients, community resources for patients, longer patient visits to improve communication and clinician documentation, coding practices to improve accuracy, preventive care best practices (eg, immunizations, disease prevention and management, and reduction of overmedicalization), support for the increased scope of clinician practice, and knowledge-seeking practices.

DHI users were identified as providers (205/241, 85.1%), patients/consumers (111/241, 46.1%), others (31/241, 12.9%), or spanning multiple types (89/241, 36.9%). The types of

providers using DHIs included physicians (163/241, 67.3%), nurses of any type (71/241, 29.5%), physician assistants (24/241, 10.0%), pharmacists (16/241, 6.6%), others (42/241, 17.4%), and multiple types (79/241, 32.8%). The "others" provider type included various clinic staff, administrators, technicians, physical therapists, researchers, health counselors, etc. The DHI user physician specialty characterization was as follows: uncharacterized (121/241, 50.2%), pediatrics (34/241, 14.1%), internal medicine (32/241, 13.2%), family practice (30/241, 12.4%), and others (25/241, 10.4%). Notably, 27 (11.2%) articles involved study settings with a mix of user types among majority Latino, African American, and Asian American populations, but only 6 (2%) of them discussed health disparities as the primary focus of their DHIs.

Primary and secondary outcomes for DHIs were predominantly clinical; 169 (70.1%) articles addressed clinical (eg, health care performance) outcomes, whereas 119 (49.4%) addressed nonclinical (eg, implementation science) outcomes. No identified studies examined health domain-related (eg, outcomes related to dimensions of wellness such as environmental, emotional, intellectual, physical, social, and spiritual) outcomes. A statistically significant improvement in relevant measured outcomes was identified in 192 (79.7%) articles, with 117 (48.5%) articles reporting improved health care performance outcomes (eg, preventive care/screening rates, validated tool scores, and medication adherence), 109 (45.2%) articles reporting improved implementation science outcomes (eg, intervention acceptability, adoption, and cost), and 34 (14.1%) articles reporting improvement in both. Among articles demonstrating statistically significant improvements in outcomes, 16 (6.6%) and 15 (6.2%) showed benefits for racial/ethnic groups specifically in health care and implementation science outcomes, respectively, with 4 (1.7%) articles identifying benefits for racial/ethnic groups in both. Moreover, 39 (16.2%) articles demonstrated only nonsignificant beneficial findings, while 8 (3.3%) articles provided no beneficial findings and only 1 (0.4%) article reported harm resulting from a DHI (in this case, limited to a portion of a subpopulation, whereas other populations received benefit).

Given that DHIs are frequently implemented as a combination of technologies, a content analysis was conducted to understand how DHIs identified in the included studies are collectively and uniquely being leveraged in care settings to impact prevention. Five topics were identified following content analysis that represent the most novel or distinct DHIs from the reviewed studies as follows: population-centered, patient-centered, care access expansion, panel-centered (dashboarding), and app-driven. Selected abstractions for the articles matching these topics are presented in Tables 1-5.

Table 1. Population-centered digital interventions for primary care.

First author, year	Study design	Description of technology	Sample size	Selected outcomes
Nagykaldi, 2014 [172]	Pre-post	Linking of a regional health system, hospital organization, and preventive services reminder system via HIE[a].	346 patients (20% ethnic minorities)	12%-36% increase in preventive service documentation and delivery (*P*<.001). 9.6% increase in medication reconciliation (*P*<.001).
Nagykaldi, 2017 [171]	Pre-post	Wellness coordinator connection to HIE organizations, PCPs[b], county health departments, and hospitals for preventive care outreach for rural communities.	9138 rural patients	3%-215% increase in delivery of 10 preventive services over 12 months (*P*=.004). 80% ROI[c] for selective preventive services (range, 32%-122%). 40% ROI on wellness coordinator employment cost.
Fanizza, 2018 [81]	Open label nonrandomized	Pharmacist connection to the state HIE for comprehensive medication review after discharge and communication with prescribers.	40 patients	25.2% decrease in overall 30-day readmission rates (*P*=.03). 22.7% decrease in 30-day readmission rates for initial diagnosis (*P*=.009).
Shade, 2015 [199]	Pre-post	Clinic link to the state surveillance system providing alerts when out-of-care HIV patients present in the ED[d] or other settings.	6 sites serving underserved communities	OR[e] 2.61 (95% CI 2.11-3.21) for care retention (*P*=.001). OR 1.24 (95% CI 1.03-1.49) for being on ART[f] (*P*=.02). OR 4.16 (95% CI 2.54-6.80) for undetectable viral load (*P*<.001).

[a]HIE: health information exchange.

[b]PCP: primary care provider.

[c]ROI: return on investment.

[d]ED: emergency department.

[e]OR: odds ratio.

[f]ART: antiretroviral therapy.

Table 2. Selected patient-centered digital health interventions for primary care (direct engagement).

First author, year	Study design	Description of technology	Sample size	Selected outcomes
Grant, 2015 [97]	RCT[a]	Informatics surveillance and reminder system connected to EHR[b] lab test orders that generates mailed letters requesting patient completion of labs for hyperlipidemia, diabetes, and HTN[c] monitoring.	4038 patients	aHR[d] 1.26 (95% CI 0.99-1.62) for decreased time to LDL[e] goal. aHR 1.15 (95% CI 1.01-1.32) for earlier LDL lab assessment.
Hess, 2014 [110]	Observational cohort	PHR[f] delivering active notifications regarding gaps in preventive chronic disease monitoring until patient logs on to the PHR or closes the prevention gap.	584 patients	58% of all prevention gaps were closed over 12 months. 61% of notified patients accessed the PHR or closed the triggering care gap after the 1st message and 73% after the 2nd message.
Hojat, 2020 [112]	Controlled trial	EHR bulk-ordered HCV[g] antibody testing plus automatic PHR messages requesting patients to go to the lab.	1024 patients	14% increase in completed HCV tests (P<.001; OR[h] 1.7, 95% CI 1.2-2.1). Only 3.5% of patients responded to PHR messages, and repeat messaging had no effect on completion.
Langford, 2019 [138]	Observational cohort	SMS text message contact to help underserved patients with diabetes find their optimal basal insulin dose.	113 patients	84% of patients reached optimal insulin dose. Age, copay status, and initial fasting blood glucose were significantly associated with 100% SMS response (P≤.03).
Mehta, 2018 [163]	RCT	Patient portal message containing either opt-in or opt-out for FIT[i] colorectal cancer screening test.	127 patients	28% higher FIT completion rate for patients receiving opt-out messages.
Quanbeck 2018 [185]	Observational cohort	Patient discussion board, interactive modules for health tracking, and self-management and coping with cravings for addiction management. Clinician web portal for patient-generated data.	268 patients	44% reduction in risky drinking days (P=.04), and 34% reduction in illicit drug use days (P=.01), over 12 months. 53%-60% of patients accessed the intervention during the final week of the implementation period.
Smallwood, 2017 [211]	RCT	Patient portal decision support tool for fracture risk and prevention. Includes educational information, risk calculation, and a treatment decision values elicitation exercise.	50 patients	Improved decision quality (P<.001) and conflict (P<.001) scores after the intervention. 25.7% (P=.046) increase in treatment decisions 3 months after the intervention.
Turvey, 2016 [234]	RCT	Patient portal link to a downloadable and printable CCD[j] for sharing with non-VA[k] providers for continuity of care.	52 patients	73% increase in the proportion of patients sharing the CCD with non-VA providers with training on accessing the CCD (P<.001). No improvement in medication reconciliations, but significant reduction in duplicate laboratory tests ordered by non-VA providers (P=.02).
Woo, 2016 [241]	Pilot	Daily customized spinal cord injury/disorder disease management questions delivered to patients via a data messaging device. Provider web portal with patient responses and risk level ratings.	33 patients	Average total response rate of 56%, ranging from 10% to 93%. Nearly 20% decrease in the DUSOI[l] score over 6 months.
Yakovchenko, 2019 [245]	RCT	Customized SMS reminder messages about HCV treatment appointments, labs, adherence, and motivation.	71 patients	Lower distress about failing treatment (P=.05) and better medication adherence (P=.06). 96% of texters vs 94% of nontexters achieved SVR[m].

[a]RCT: randomized controlled trial.

[b]EHR: electronic health record.

[c]HTN: hypertension.

[d]aHR: adjusted hazard ratio.

[e]LDL: low-density lipoprotein.

[f]PHR: personalized health record.

[g]HCV: hepatitis C virus.

[h]OR: odds ratio.

[i]FIT: fecal immunochemical test.

[j]CCD: continuity of care document.

[k]VA: Veterans Affairs.

[l]DUSOI: Duke Severity of Illness Checklist.

[m]SVR: sustained virologic response.

Table 3. Selected care access expansion digital health interventions (virtual care/telehealth).

First author, year	Study design	Description of technology	Sample size	Selected outcomes
Aikens, 2015 [20]	Observational cohort	Weekly IVR[a] calls for depression self-management. Option to designate a lay support person to receive email reports summarizing reported symptoms and providing problem-tailored support guidance.	221 patients	Increases of 20% in the per-week aOR[b] for medication adherence and 16% for depression remission compared with controls.
Coker, 2019 [58]	RCT[c]	Telehealth-enhanced referral to a CMHC[d] using informational videos, SMS text messages, and telehealth screening at the primary care clinic.	342 Latino children	aOR 3.02 (95% CI 1.47-6.22) for completing CMHC visits compared with controls. Telehealth referrals took longer to complete screening but reported greater satisfaction with referral than controls.
Halterman, 2018 [103]	RCT	Videoconference telemedicine visit in a school health office for asthma baseline and medication; follow-up telemedicine assessments every 4-6 weeks.	400 urban students	0.69 (95% CI 0.15-1.22) more symptom-free days per 2 weeks (P=.01). aOR 0.52 (95% CI 0.32-0.84) for asthma-related ED[e] visit or hospitalization.
Osofsky, 2017 [178]	Pre-experimental time series	Onsite and/or telemedicine behavioral-based trauma treatment delivered in primary care clinics.	235 patients	4.5-point decrease in the PCL-C[f] score (P=.001), and 1.8-point decrease in the PHQ-15[g] score (P=.001).
Perry, 2018 [181]	RCT	Live video telemedicine asthma education at school for a child, caregiver(s), and school nurse; telemonitoring of patient-reported symptoms; PCP[h] prompts with guideline-based asthma management.	393 rural African American students	No change in symptom-free days, quality of life, or lung function. 42% increase in peak flow meter use compared with controls (P<.01) and 19% increase in medication adherence (P=.03) over 6 months.
Reeves, 2016 [186]	Pre-post	Implementation of EHRs[i] in the school system for the asthma care program; messaging connection to PCP EHR systems; school nurse asthma template for PCP messaging.	33 students	39.4% decrease in asthma inpatient admissions (P<.001) and 18.2% decrease in exacerbations (P<.05) over 12 months.
Richter, 2015 [189]	RCT	Live video telehealth for tobacco cessation delivered in primary care clinics.	566 patients	No difference in biochemically verified prevalence, prolonged abstinence, quit attempts, or number of cigarettes smoked per day compared with phone counseling.

[a]IVR: interactive voice response.

[b]aOR: adjusted odds ratio.

[c]RCT: randomized controlled trial.

[d]CMHC: community mental health clinic.

[e]ED: emergency department.

[f]PCL-C: posttraumatic stress disorder checklist-civilian version.

[g]PHQ-15: 15-item patient health questionnaire.

[h]PCP: primary care physician.

[i]EHR: electronic health record.

Table 4. Panel-centered digital health interventions for primary care (dashboarding).

First author, year	Study design	Description of technology	Sample size	Selected outcomes
Allen, 2017 [24]	RCT[a]	Culturally sensitive team model using an electronic diabetes dashboard providing alerts and reports for each patient regarding clinical and behavioral factors and social distress.	399 Latino patients	Social distress score decrease of 0.6 (controls) vs 1.6 (intervention) over 6 months (*P*=.01).
Duquaine, 2015 [75]	Observational cohort	CDS[b] for tobacco use and interventions for smoking cessation; quarterly communications with practice-specific and overall program performance.	19 clinics treating low-income and Medicaid patients	Successful implementation at all sites. Change in EHR[c] documentation of prevalence and cessation rates (NR[d]).
Fiks, 2015 [85]	Open-label nonrandomized	Quarterly feedback reports summarizing personal, practice, and network rates of missed HPV[e] vaccine opportunities.	227 PCPs[f]	5.7% (95% CI 3.8-7.7) increase in HPV vaccination compared with controls.
Kapoor, 2018 [127]	Observational cohort	Emailed report of the proportion of atrial fibrillation patients receiving anticoagulation therapy compared to peers plus EHR message 1 day before visits with anticoagulation eligible patients.	5406 patients	Providers reviewed emails (45%) and EHR messages (96%), demonstrating feasibility. No change in the percentage of patients receiving anticoagulation therapy compared with controls after 3 months.
Zimmerman, 2017 [255]; Nowalk, 2016 [174]; Zimmerman, 2017 [257]; Lin, 2016 [145]; Zimmerman, 2017 [256]	RCT and pre-post	4 Pillars Immunization Toolkit and Practice Transformation Program. Web-based dashboard providing and tracking strategies for increasing practice vaccination rates, including EHR prompts, digital outreach, and standing order programs.	25 clinics [255]; 24 clinics [174]; 11 clinics [257]; 25 clinics [145]; 22 clinics [256]	2.7% to 10.2% statistically significant increases in vaccination rates for intervention and control sites during RCT studies. −1.9% to 17.1% statistically significant increases in vaccination rates for active intervention groups during year 2 of the pre-post study.

[a]RCT: randomized controlled trial.

[b]CDS: clinical decision support.

[c]EHR: electronic health record.

[d]NR: not reported.

[e]HPV: human papillomavirus.

[f]PCP: primary care physician.

Table 5. Selected app-driven digital health interventions for primary care.

First author, year	Study design	Description of technology	Sample size	Selected outcomes
Bennett, 2018 [35]	RCT[a]	App using IVR[b] and SMS text messaging to collect patient behavior change data and weight via a smart scale, provide tailored patient feedback based on goal progression, and generate EHR[c] counseling recommendations for clinicians.	351 patients	−4.4 kg (95% CI −5.5 to −3.3) weight loss at 6 months (P<.001); −3.8 kg (95% CI −5.0 to −2.5) weight loss at 12 months (P<.001). Participants completing ≥80% of interactions lost significantly more weight than less engaged participants (P<.01).
Brayboy, 2016 [45]	Pre-post	iPhone-compatible app for providing trusted, age-appropriate, straightforward sexual health information and resources to teenage girls.	20 teenage girls	3.4%-4.2% improvement in sexual health topic knowledge. 58.8% increase in the perception that they or other teenage girls would use the app (P<.001).
Dahne, 2019 [64]	RCT	Self-help app adaptation of Brief Behavioral Apptivation, including education, identification of values, daily mood monitoring, and social support including gamification, to reinforce continued use.	52 patients	63% greater decrease on BDI-II[d] assessment after treatment compared with usual care. 70% of participants continued to use the app 1 month after enrollment, and 50% continued to use it at 2 months.
Gustafson, 2014 [99]	RCT	Smartphone app to support alcoholism recovery using alerts for trigger locations, audio-guided relaxation, PRO[e] measurement, and clinician notification, as well as a panic button for contacting support persons.	349 patients	1.37 (95% CI 0.46-2.27) fewer risky drinking days than controls over 12 months (P=.003). OR 1.65 (95% CI 1.05-2.57) for abstinence prevalence over 12 months (P=.03).
Leddy, 2019 [140]	RCT	Home smartphone urinalysis test to complete proteinuria screening for HTN[f] management. SMS text message link for downloading the app, obtaining the home testing kit, and receiving PCP[g] notification of abnormal results.	999 patients	10.9% increase in proteinuria screening completion (P<.001). 89% of home test patients preferred home testing over a visit to the physician's office.
Lv, 2017 [149]	Pre-post	Dashboard of patient's personalized action plan, treatment goals, and self-monitoring data combined with a wireless BP[h] monitor, smartphone, study app, pedometer, and web messaging system.	147 patients	55.9% increase in the proportion of patients meeting office BP goals (<140/90 mmHg) at 6 months (P<.001). 46.2% increase in the proportion of patients meeting home BP goals (<135/85 mmHg) at 6 months (P<.001).
Ofili, 2018 [176]	Pre-post	App with diabetes curriculum, goal identification and tracking, connectivity to consumer devices (eg, activity monitors), and health coach consultation.	287 patients	Improvements in SBP[i] (6 mmHg), blood glucose (15 mg/dL), and physical activity (0.56 miles/day) at 12 weeks (all P<.01), which continued through 52 weeks.
Yu, 2018 [249]	Pre-post	App delivering a guided cognitive behavioral program for generalized anxiety disorder along with in-app coach pairing and messaging.	63 patients	3.6-point mean reduction on GAD-7[j] over 2 months for patients with baseline GAD-7 ≥8 (P<.001).

[a]RCT: randomized controlled trial.

[b]IVR: interactive voice response.

[c]EHR: electronic health record.

[d]BDI-II: Beck Depression Inventory II.

[e]PRO: patient-reported outcome.

[f]HTN: hypertension.

[g]PCP: primary care provider.

[h]BP: blood pressure.

[i]SBP: systolic blood pressure.

[j]GAD-7: Generalized Anxiety Disorder-7.

Primary prevention targets focused on the use of population-centered [171,172] and panel-centered [85,145,175,255-257] DHIs to improve adolescent [256,257] or adult [145,171,172,174,255] vaccination rates for human papillomavirus [256,257], influenza [145,171,172,257], pneumococcal disease [171,172,255,257], and Tdap (tetanus, diphtheria, and pertussis) [174,257].

All the above DHIs that targeted primary prevention had statistically significant health care [145,172,255-257] or implementation [171,174] outcomes following the intervention.

Identifying return on investment (ROI) and value on investment can be large barriers for DHI implementation; however, both were satisfied when a community wellness registry was connected to EHRs via a health information exchange (HIE). This pilot study demonstrated the feasibility and cost-effectiveness of technology implementation in a community-based model with a mean ROI of 80% (range, 32% to 122%) for the improved delivery of 10 selective preventive services (mean increase 35%, range 3% to 215%; *P*=.04) in rural settings [171].

Patient-centered [112,163,211] and population-centered [171,172] DHIs supported secondary prevention by examining measures that led to early diagnosis and treatment using direct-to-patient messaging in an EHR [112,163], decision aids embedded in patient portals [211], and an intelligent HIE using clinical decision support [171,172]. These DHIs improved screening rates for cancer (eg, breast [172] and colorectal [163,171,172]), hepatitis C virus (HCV) [112], and osteoporosis [171,211]. Only 1 study in this grouping did not have significant improvements following the DHI, which may be due to the more invasive and costly colonoscopy procedure itself rather than the ineffectiveness of the EHR portal messaging intervention to improve colorectal cancer screening [163]. However, an advanced EHR that used population analytics and bulk laboratory ordering to directly engage patients for universal HCV screening nearly doubled testing (odds ratio 1.7, 95% CI 1.2-2.1) in the intervention group [112].

Tertiary prevention for chronic disease management was supported primarily by care access expansion [20,58,103,178,181,189], app-driven [35,64,140,149,176,249], and patient-centered [38,97,241,245] approaches. Overall, DHIs decreased disease severity and associated comorbidities; lowered the numbers of emergency department visits, hospitalizations, and 30-day readmissions; increased the receipt of follow-up care; improved medication adherence in the identified studies [20,35,58,64,97,103,149,176,199,245,249]; and improved the quality or effectiveness of health services by technology implementation [24,38,103,140,171,178,186,189,249]. Disease areas targeted by DHIs included diabetes [38,97,171,176], hypertension [97,140,149], asthma [103,181,186], obesity [35,171], cardiovascular disease [123], HIV [199], HCV [245], and hyperlipidemia [97]. Management of behavioral health included smoking cessation support [75,171,189], promotion of physical activity [171], substance abuse management [99,185], and sexual health education [45] using mHealth technology. Mental health [58] (eg, depression [20,64], anxiety [249], posttraumatic stress disorder [178], and social distress [24]) improved following digital interventions. Notably, telehealth and mHealth were leveraged predominantly to support mental health interventions with care access expansion [20,58,178] and app-driven [64,249] technologies to improve patient function, minimize illness impacts, and decrease associated complications. Behavioral and mental health conditions and other chronic diseases often occur concurrently [258]. Two studies [178,185] integrated behavioral or mental health DHIs for chronic condition care, but only 1 study [178] reported outcome measures for both mental and physical health, whereby both improved significantly.

Notably, none of the studies identified by the content analysis examined DHIs for quaternary prevention.

Using the Oxford levels of evidence [16], the quality of the included studies was moderate to low overall due to many studies (101/241, 41.9%) presenting level 4 evidence (eg, case series, poor quality cohort, and case-control studies) and the remainder displaying level 1b (eg, individual RCTs with a narrow CI; 46/241, 19.1%), 2b (eg, individual cohorts including low-quality RCTs; 58/241, 24.1%), 2c (eg, outcomes research and ecological studies; 17/241, 7.1%), 3b (eg, individual case-control studies; 17/241, 7.1%), or 5 (eg, expert opinions; 3/241, 1.2%) evidence.

Discussion

Principal Findings

Amidst the rapid digital transformation of the primary care delivery system in response to the COVID-19 pandemic, this is the first comprehensive summary on DHIs in use by interdisciplinary clinicians (eg, physicians, pharmacists, psychiatrists, etc) in primary care. This scoping review and its subgroup analysis summarized a growing evidence base and rendered a collection of potentially successful strategies for patients, providers, and population stakeholders to improve outcomes for health, health care performance, and implementation science through the use of DHIs. Moreover, important scientific gaps were identified in the contemporary evaluation and knowledge of DHIs leveraged in primary care, particularly the scarcity of the evaluation of DHIs in health disparities and evaluation of the negative effects of DHIs.

A few major themes emerged from our analysis of the extracted data. First, the digital health technologies identified and reviewed were highly concentrated in a narrow range of HIT, most specifically around EHRs/electronic medical records, particularly with the use of alerts to help clinicians make appropriate clinical decisions. Though understandable given their high use and decade-long attention to increasing adoption via “meaningful use” in primary care [259,260], the absence of DHI literature involving other platforms was telling. Despite unprecedented attention to telehealth implementation due to the COVID-19 pandemic response, little evidence of effective implementation of this specific DHI exists to guide primary care telehealth use for health care delivery in the US. A few studies did examine more innovative uses of technology, particularly for the delivery of mental and behavioral health (Tables 1-5). As HIT continues to rapidly evolve and health care is delivered in more innovative ways due to the COVID-19 pandemic, more research should focus on novel DHIs applied to primary care.

Second, despite prevention being 1 of 6 mechanisms underpinning primary care’s beneficial impact on population health [261] and an early target for DHIs, studies evaluating prevention were predominantly focused on secondary or tertiary preventive interventions. Most would agree that disease prevention offers the greatest yield for population health and is amenable to DHIs via mobile and online apps, clinical kiosks, and electronic patient portals [262,263]. Primary prevention

interventions, such as immunizations, rely on effective patient counseling and education, which can be difficult and time-consuming to document and capture in EHRs (the predominant type of intervention found in our review). This finding may be a reflection of physician roles in the US. Traditionally, the role of primary prevention has relied on public health professionals [264], and although primary care physicians are increasing their ability to address the needs of the community, most physicians are still focused on the needs of the individual [265]. As the intersection of public health and primary care becomes more urgent to strategically improve individual and population health, future studies should examine the role of DHI adoption and implementation in their integration.

Third, DHIs enhanced core primary care functions by contributing to the comprehensiveness of care provided. This was an unexpected finding given that DHIs are often thought of in the context of first contact through patient portals, coordination through electronic referrals and linked EHRs, and continuity through HIEs and sharing of documents. Many of the articles reviewed discussed the use of DHIs to identify patients in need of services and alert clinicians to provide them. For example, multiple studies described EHR alerts that would prompt clinicians to order viral hepatitis C testing for patients with indications for screening (Multimedia Appendix 11) [83,93,112,116,133,150,164,173,207,246]. Other studies shared examples of how patients could be trained to provide services for themselves (Table 2) or how DHIs could be used to offer additional clinical services (Table 3). Thus, it makes sense that comprehensiveness, or the provision of a robust set of services to a patient, would be improved with DHIs. In an era where comprehensiveness of care is said to be declining in primary care [266-271], DHIs may provide an innovative solution for primary care practices to increase and enhance the services they provide.

Finally, while the development and release of health apps continue to increase, few evaluations of app-driven DHIs were identified in our study (Table 5). This may be in part because many apps lacked integration with primary care or other technology systems or because of the evolving standards for evaluating these types of interventions, as evidenced by the recent establishment of the FDA's Digital Health Center of Excellence [272]. Most app-driven DHIs included in our study were patient-facing and focused on helping to better involve patients in their care. However, app-driven DHIs are also capable of providing an overwhelming amount of data to providers. Balancing data collection features from apps by adding functionalities, such as thresholds triggering clinical alerts/feedback, designing patient-counseling suggestions based on gathered data, and pairing with timely coaching/contact is important to enhance the clinical relevance and quality of these tools. As the development and clinical adoption of app-driven DHIs continue to expand, rigorous investigation of their safety, efficacy, and value in primary care is urgently needed.

Limitations

These results should be interpreted in the context of a few limitations. The findings are limited to studies conducted in US settings, which prohibits the generalization of their applicability and use at a global scale. Review of the use of DHIs in non-US primary care settings should be prioritized in future work. Further, due to the heterogeneity of identified interventions, it is not possible to provide head-to-head comparisons. The large heterogeneity of DHIs is an additional reason why our synthesis focused on the novel and distinct DHIs that are collectively used in primary care practice rather than presenting evidence collated by distinct DHI technologies. Other limitations include single screening of titles and abstracts, English language restriction, and lack of gray literature evaluation. Data extraction for each article was not confirmed by a secondary reviewer, leaving room for bias in the interpretation of the articles. For example, it was left up to each reviewer to determine the type of prevention the DHI was addressing, or which primary care function (eg, the 4Cs by Dr Starfield: first contact, comprehensiveness, continuity, and coordination) the DHI enhanced. However, careful and collaborative definition of our processes and outcomes prior to extraction (ie, types of prevention or primary care functions) should minimize this bias. Lastly, we intentionally selected a quality assessment tool rather than a risk of bias tool, as we only planned to measure the extent that methodological safeguards (ie, internal validity) against bias were implemented. A risk of bias assessment would have offered a bias judgement (ie, estimation of intervention effects) on such a quality assessment, and judgement of the evidence may have shifted with this approach. It is important to consider that even when a study implements all possible safeguards in a tool, it may not be unbiased, and conversely, a study applying no safeguards is not necessarily biased [273].

Conclusions

Gayle Stephens noted in 1965 that "One of the paradoxes of our time is that the healing relationship seems most in jeopardy at a time when we need it most," commenting on the range of "forces which threaten to depersonalize the meeting of a doctor and patient" [274]. That paradox remains in an age where technology is often seen as distracting rather than enhancing care. Through further adoption of DHIs with evidence of effectiveness, providers and patients/consumers can enhance primary care by improving the delivery of preventive services and promoting more comprehensive care. Yet, relying solely on EHR alerts may not lead to substantial improvements in health care in the US. Moreover, rigorous and prospective evaluations of the potential negative effects of these DHIs, particularly for clinical end users of these technologies, will be needed to ensure holistic improvement of health care. Innovative DHIs should undergo evaluation in well-designed studies to generate evidence and establish best practices that can be replicated and scaled in diverse primary care settings. Given the ability of technology to amplify existing health disparities and biases, the development of DHIs that can help overcome health disparities and the evaluation of the benefits and harms of current DHIs on health disparities are imperative. In addition, DHIs that allow integration of public health with primary care will be essential for rapid and effective responses to health and health care challenges, such as the COVID-19 pandemic, in an increasingly technology-driven health care environment.

Acknowledgments

We would like to thank Dr Brian Antono for his screening support. Funding was provided by IBM Watson Health.

Authors' Contributions

Conceptualization: KJTC, YJ, AB, and KBR; formal analysis: KJTC, VCW, YEA, ELS, YJ, and MA; methodology: KJTC; project administration: KJTC; supervision: KJTC; validation: KJTC and VCW; writing–original draft: KJTC, YJ, VCW, and AB; writing–review and editing: all authors.

Conflicts of Interest

VCW was employed by IBM Corporation. KJTC is employed by IBM Corporation. YJ has no conflicts. ELS was employed by IBM Corporation. YEA was employed by IBM Corporation. MA has no conflicts. KBR was employed by IBM Corporation. AB has no conflicts.

Multimedia Appendix 1
Search outline using the PCC (participants, concept, and context) framework.
[DOCX File , 26 KB - medinform_v10i1e33518_app1.docx]

Multimedia Appendix 2
MEDLINE search via PubMed.
[DOCX File , 28 KB - medinform_v10i1e33518_app2.docx]

Multimedia Appendix 3
Cochrane Library search.
[DOCX File , 28 KB - medinform_v10i1e33518_app3.docx]

Multimedia Appendix 4
Embase search.
[DOCX File , 28 KB - medinform_v10i1e33518_app4.docx]

Multimedia Appendix 5
Updated MEDLINE search via PubMed.
[DOCX File , 29 KB - medinform_v10i1e33518_app5.docx]

Multimedia Appendix 6
Updated Cochrane Library search.
[DOCX File , 28 KB - medinform_v10i1e33518_app6.docx]

Multimedia Appendix 7
Updated Embase search.
[DOCX File , 29 KB - medinform_v10i1e33518_app7.docx]

Multimedia Appendix 8
Inclusion/exclusion criteria.
[DOCX File , 27 KB - medinform_v10i1e33518_app8.docx]

Multimedia Appendix 9
Description of digital health intervention categories.
[DOCX File , 29 KB - medinform_v10i1e33518_app9.docx]

Multimedia Appendix 10
Data extraction forms.
[DOCX File , 806 KB - medinform_v10i1e33518_app10.docx]

Multimedia Appendix 11
Abstracted results from included articles.
[DOCX File , 59 KB - medinform_v10i1e33518_app11.docx]

References

1. Institute of Medicine (US) Committee on the Future of Primary Care. In: Donaldson MS, Yordy KD, Lohr KN, Vanselow NA, editors. Primary Care: America's Health in a New Era. Washington, DC: National Academies Press; 1996.
2. Starfield B. Primary Care: Balancing Health Needs, Services, and Technology. Oxford, United Kingdom: Oxford University Press, USA; 1998.
3. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q 2005;83(3):457-502 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00409.x] [Medline: 16202000]
4. Hensrud DD. Clinical preventive medicine in primary care: background and practice: 1. Rationale and current preventive practices. Mayo Clin Proc 2000 Feb;75(2):165-172. [doi: 10.4065/75.2.165] [Medline: 10683656]
5. Rubin HR. Overcoming barriers to preventive care. J Gen Intern Med 2000 Jun;15(6):434-436 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.115600.x] [Medline: 10886481]
6. Petitti DB, Teutsch SM, Barton MB, Sawaya GF, Ockene JK, DeWitt T, U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: insufficient evidence. Ann Intern Med 2009 Feb 03;150(3):199-205. [doi: 10.7326/0003-4819-150-3-200902030-00010] [Medline: 19189910]
7. Privett N, Guerrier S. Estimation of the Time Needed to Deliver the 2020 USPSTF Preventive Care Recommendations in Primary Care. Am J Public Health 2021 Jan;111(1):145-149. [doi: 10.2105/AJPH.2020.305967] [Medline: 33211585]
8. Ferrante JM, Balasubramanian BA, Hudson SV, Crabtree BF. Principles of the patient-centered medical home and preventive services delivery. Ann Fam Med 2010 Mar 08;8(2):108-116 [FREE Full text] [doi: 10.1370/afm.1080] [Medline: 20212297]
9. Roland KB, Milliken EL, Rohan EA, DeGroff A, White S, Melillo S, et al. Use of Community Health Workers and Patient Navigators to Improve Cancer Outcomes Among Patients Served by Federally Qualified Health Centers: A Systematic Literature Review. Health Equity 2017;1(1):61-76 [FREE Full text] [doi: 10.1089/heq.2017.0001] [Medline: 28905047]
10. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. The Lancet 2014 Jul 5;384(9937):45-52. [doi: 10.1016/S0140-6736(14)60648-6] [Medline: 24996589]
11. Thomas Craig KJ, Morgan LC, Chen C, Michie S, Fusco N, Snowdon JL, et al. Systematic review of context-aware digital behavior change interventions to improve health. Transl Behav Med 2021 May 25;11(5):1037-1048 [FREE Full text] [doi: 10.1093/tbm/ibaa099] [Medline: 33085767]
12. Thomas Craig KJ, Fusco N, Lindsley K, Snowdon J, Willis V, Arriaga Y, et al. Rapid review: Identification of digital health interventions in atherosclerotic-related cardiovascular disease populations to address racial, ethnic, and socioeconomic health disparities. Cardiovascular Digital Health Journal 2020 Nov;1(3):139-148 [FREE Full text] [doi: 10.1016/j.cvdhj.2020.11.001]
13. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
14. Office-based Physician Electronic Health Record Adoption. HealthIT. URL: https://www.healthit.gov/data/quickstats/office-based-physician-electronic-health-record-adoption [accessed 2022-01-03]
15. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
16. Oxford Centre for Evidence-Based Medicine: Levels of Evidence (March 2009). CEBM. URL: https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ [accessed 2020-03-27]
17. Adams WG, Phillips BD, Bacic JD, Walsh KE, Shanahan CW, Paasche-Orlow MK. Automated conversation system before pediatric primary care visits: a randomized trial. Pediatrics 2014 Sep 04;134(3):e691-e699. [doi: 10.1542/peds.2013-3759] [Medline: 25092938]
18. Afshar AR, Oldenburg CE, Stewart JM. A Novel Hybrid Fixed and Mobile Ultra-Widefield Imaging Program for Diabetic Teleretinopathy Screening. Ophthalmol Retina 2019 Jul;3(7):576-579. [doi: 10.1016/j.oret.2019.03.007] [Medline: 31076334]
19. Ahern DK, Parker D, Eaton C, Rafferty C, Wroblewski J, Goldman R. Patient-facing Technology for Identification of COPD in Primary Care. J Innov Health Inform 2016 Jul 15;23(2):824 [FREE Full text] [doi: 10.14236/jhi.v23i2.824] [Medline: 27869583]
20. Aikens JE, Trivedi R, Heapy A, Pfeiffer PN, Piette JD. Potential Impact of Incorporating a Patient-Selected Support Person into mHealth for Depression. J Gen Intern Med 2015 Jun 10;30(6):797-803 [FREE Full text] [doi: 10.1007/s11606-015-3208-7] [Medline: 25666218]
21. Akenroye AT, Kumthekar AA, Alevizos MK, Mowrey WB, Broder A. Implementing an Electronic Medical Record-Based Reminder for Cardiovascular Risk Screening in Rheumatoid Arthritis. Arthritis Care Res (Hoboken) 2017 May 10;69(5):625-632. [doi: 10.1002/acr.22966] [Medline: 27390217]

22. Albu JB, Sohler N, Li R, Li X, Young E, Gregg EW, et al. An Interrupted Time Series Analysis to Determine the Effect of an Electronic Health Record-Based Intervention on Appropriate Screening for Type 2 Diabetes in Urban Primary Care Clinics in New York City. Diabetes Care 2017 Aug 15;40(8):1058-1064 [FREE Full text] [doi: 10.2337/dc16-2133] [Medline: 28620094]
23. Aleem S, Torrey W, Duncan M, Hort S, Mecchella J. Depression screening optimization in an academic rural setting. International Journal of Health Care Quality Assurance 2015;28(7):709-725. [doi: 10.1108/ijhcqa-01-2015-0012]
24. Allen NA, Zagarins S, Welch G. Reducing Social Distress for Chronic Disease Patients in Primary Care: An Intervention for Latino Type 2 Diabetes Patients Seen at Community Health Centers. Diabetes Spectr 2017 Nov 27;30(4):288-296 [FREE Full text] [doi: 10.2337/ds16-0031] [Medline: 29151719]
25. Anderson D, Villagra VG, Coman E, Ahmed T, Porto A, Jepeal N, et al. Reduced Cost Of Specialty Care Using Electronic Consultations For Medicaid Patients. Health Aff (Millwood) 2018 Dec;37(12):2031-2036. [doi: 10.1377/hlthaff.2018.05124] [Medline: 30633678]
26. Apter AJ, Localio AR, Morales KH, Han X, Perez L, Mullen AN, et al. Home visits for uncontrolled asthma among low-income adults with patient portal access. J Allergy Clin Immunol 2019 Sep;144(3):846-853.e11 [FREE Full text] [doi: 10.1016/j.jaci.2019.05.030] [Medline: 31181221]
27. Bachhuber MA, O'Grady MA, Chung H, Neighbors CJ, DeLuca J, D'Aloia EM, et al. Delivery of screening and brief intervention for unhealthy alcohol use in an urban academic Federally Qualified Health Center. Addict Sci Clin Pract 2017 Dec 07;12(1):33 [FREE Full text] [doi: 10.1186/s13722-017-0100-2] [Medline: 29212532]
28. Bae J, Ford E, Kharrazi H, Huerta T. Electronic medical record reminders and smoking cessation activities in primary care. Addict Behav 2018 Feb;77:203-209 [FREE Full text] [doi: 10.1016/j.addbeh.2017.10.009] [Medline: 29065376]
29. Bae J, Ford EW, Wu S, Huerta T. Electronic reminder's role in promoting human papillomavirus vaccine use. Am J Manag Care 2017 Nov 01;23(11):e353-e359 [FREE Full text] [Medline: 29182355]
30. Baker JW, Forkum W, McNeal J. Utilizing clinical video telehealth to improve access and optimize pharmacists' role in diabetes management. J Am Pharm Assoc (2003) 2019 Apr;59(2S):S63-S66. [doi: 10.1016/j.japh.2019.01.004] [Medline: 30738723]
31. Bar-Shain DS, Stager MM, Runkle AP, Leon JB, Kaelber DC. Direct messaging to parents/guardians to improve adolescent immunizations. J Adolesc Health 2015 May;56(5 Suppl):S21-S26 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.11.023] [Medline: 25863550]
32. Barclay C, Viswanathan M, Ratner S, Tompkins J, Jonas DE. Implementing Evidence-Based Screening and Counseling for Unhealthy Alcohol Use with Epic-Based Electronic Health Record Tools. Jt Comm J Qual Patient Saf 2019 Aug;45(8):566-574 [FREE Full text] [doi: 10.1016/j.jcjq.2019.05.009] [Medline: 31378277]
33. Barton AB, Okorodudu DE, Bosworth HB, Crowley MJ. Clinical Inertia in a Randomized Trial of Telemedicine-Based Chronic Disease Management: Lessons Learned. Telemed J E Health 2018 Oct;24(10):742-748. [doi: 10.1089/tmj.2017.0184] [Medline: 29341850]
34. Beasley JM, Kirshner L, Wylie-Rosett J, Sevick MA, DeLuca L, Chodosh J. BRInging the Diabetes prevention program to GEriatric populations (BRIDGE): a feasibility study. Pilot Feasibility Stud 2019 Nov 11;5(1):129 [FREE Full text] [doi: 10.1186/s40814-019-0513-7] [Medline: 31741744]
35. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an App and Provider Counseling for Obesity Treatment in Primary Care. American Journal of Preventive Medicine 2018 Dec;55(6):777-786. [doi: 10.1016/J.AMEPRE.2018.07.005]
36. Benson GA, Sidebottom A, Sillah A, Boucher JL, Miedema MD, Knickelbine T, et al. Reach and effectiveness of the HeartBeat Connections telemedicine pilot program. J Telemed Telecare 2017 Feb 13;24(3):216-223. [doi: 10.1177/1357633x17692723]
37. Berkley H, Barnes M, Carnahan D, Hayhurst J, Bockhorst A, Neville J. Public Health Surveillance via Template Management in Electronic Health Records: Tri-Service Workflow's Rapid Response to an Infectious Disease Crisis. Military Medicine 2017 Mar;182(S1):340-345. [doi: 10.7205/milmed-d-16-00128]
38. Bhat S, Kroehl ME, Trinkley KE, Chow Z, Heath LJ, Billups SJ, et al. Evaluation of a Clinical Pharmacist-Led Multidisciplinary Antidepressant Telemonitoring Service in the Primary Care Setting. Popul Health Manag 2018 Oct;21(5):366-372 [FREE Full text] [doi: 10.1089/pop.2017.0144] [Medline: 29211661]
39. Bondurant KL, Wheeler JG, Bursac Z, Holmes T, Tilford JM. Comparison of Office-Based Versus Outsourced Immunization Recall Services. Clin Pediatr (Phila) 2017 Jun;56(6):555-563. [doi: 10.1177/0009922816673307] [Medline: 27798388]
40. Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S. Validation of a Novel Electronic Health Record Patient Portal Advance Care Planning Delivery System. J Med Internet Res 2018 Jun 26;20(6):e208 [FREE Full text] [doi: 10.2196/jmir.9203] [Medline: 29945860]
41. Bose-Brill S, Kretovics M, Ballenger T, Modan G, Lai A, Belanger L, et al. Testing of a tethered personal health record framework for early end-of-life discussions. Am J Manag Care 2016 Jul 01;22(7):e258-e263 [FREE Full text] [Medline: 27442309]

42. Bouskill K, Smith-Morris C, Bresnick G, Cuadros J, Pedersen ER. Blind spots in telemedicine: a qualitative study of staff workarounds to resolve gaps in diabetes management. BMC Health Serv Res 2018 Aug 07;18(1):617 [FREE Full text] [doi: 10.1186/s12913-018-3427-9] [Medline: 30086743]
43. Brantley AD, Page KM, Zack B, Friedrich KR, Wendell D, Robinson WT, et al. Making the Connection: Using Videoconferencing to Increase Linkage to Care for Incarcerated Persons Living with HIV Post-release. AIDS Behav 2019 Jan;23(Suppl 1):32-40. [doi: 10.1007/s10461-018-2115-4] [Medline: 29680934]
44. Bratic JS, Cunningham RM, Belleza-Bascon B, Watson SK, Guffey D, Boom JA. Longitudinal Evaluation of Clinical Decision Support to Improve Influenza Vaccine Uptake in an Integrated Pediatric Health Care Delivery System, Houston, Texas. Appl Clin Inform 2019 Oct;10(5):944-951 [FREE Full text] [doi: 10.1055/s-0039-3400748] [Medline: 31860114]
45. Brayboy LM, Sepolen A, Mezoian T, Schultz L, Landgren-Mills BS, Spencer N, et al. Girl Talk: A Smartphone Application to Teach Sexual Health Education to Adolescent Girls. J Pediatr Adolesc Gynecol 2017 Feb;30(1):23-28 [FREE Full text] [doi: 10.1016/j.jpag.2016.06.011] [Medline: 27393638]
46. Breitenstein SM, Laurent S, Pabalan L, Risser HJ, Roper P, Saba MT, et al. Implementation findings from an effectiveness-implementation trial of tablet-based parent training in pediatric primary care. Fam Syst Health 2019 Dec;37(4):282-290 [FREE Full text] [doi: 10.1037/fsh0000447] [Medline: 31621349]
47. Brown CM, Perkins J, Blust A, Kahn RS. A neighborhood-based approach to population health in the pediatric medical home. J Community Health 2015 Feb 13;40(1):1-11. [doi: 10.1007/s10900-014-9885-z] [Medline: 24923727]
48. Brunette MF, Gunn W, Alvarez H, Finn PC, Geiger P, Ferron JC, et al. A pre-post pilot study of a brief, web-based intervention to engage disadvantaged smokers into cessation treatment. Addict Sci Clin Pract 2015 Feb 01;10:3 [FREE Full text] [doi: 10.1186/s13722-015-0026-5] [Medline: 25638283]
49. Buist DSM, Knight Ross N, Reid RJ, Grossman DC. Electronic health risk assessment adoption in an integrated healthcare system. Am J Manag Care 2014;20(1):62-69 [FREE Full text] [Medline: 24512165]
50. Bulbin D, Denio AE, Berger A, Brown J, Maynard C, Sharma T, et al. Improved Gout Outcomes in Primary Care Using a Novel Disease Management Program: A Pilot Study. Arthritis Care Res (Hoboken) 2018 Nov;70(11):1679-1685. [doi: 10.1002/acr.23544] [Medline: 29438607]
51. Burdick T, Kessler R. Development and use of a clinical decision support tool for behavioral health screening in primary care clinics. Appl Clin Inform 2017 Dec 21;08(02):412-429. [doi: 10.4338/aci-2016-04-ra-0068]
52. Carswell SB, Mitchell SG, Gryczynski J, Lertch E. Computerizing NIAAA's Best Practices for Youth Screening and Brief Intervention: A Proof-of-Concept Pilot Study of an Automated Alcohol Screening and Intervention Resource Tool. J Drug Educ 2020 Dec 16;49(1-2):3-14. [doi: 10.1177/0047237919894960] [Medline: 31841034]
53. Cawthon C, Mion LC, Willens DE, Roumie CL, Kripalani S. Implementing Routine Health Literacy Assessment in Hospital and Primary Care Patients. The Joint Commission Journal on Quality and Patient Safety 2014 Feb;40(2):68-76. [doi: 10.1016/s1553-7250(14)40008-4]
54. Chambers EC, Wylie-Rosett J, Blank AE, Ouziel J, Hollingsworth N, Riley RW, et al. Increasing Referrals to a YMCA-Based Diabetes Prevention Program: Effects of Electronic Referral System Modification and Provider Education in Federally Qualified Health Centers. Prev Chronic Dis 2015 Nov 05;12:E189 [FREE Full text] [doi: 10.5888/pcd12.150294] [Medline: 26542141]
55. Chouteau WA, Allen SR. Implementation of a Portable Medical Summary for Adolescents and Young Adults With Medical Complexity in Transition to Adult Health Care. J Pediatr Nurs 2019;48:35-41. [doi: 10.1016/j.pedn.2019.05.012] [Medline: 31229878]
56. Chung E, Gubernick R, LaNoue M, Abatemarco D. Child Abuse and Neglect Risk Assessment: Quality Improvement in a Primary Care Setting. Acad Pediatr 2019 Mar;19(2):227-235 [FREE Full text] [doi: 10.1016/j.acap.2018.09.011] [Medline: 30273689]
57. Church E, Banks R, Wilson B, Arfons L, Perez F, Jump R. Improving pneumococcal vaccine uptake in veterans with chronic lymphocytic leukemia through a virtual clinic. Curr Oncol 2018 Feb 01;25(1):e95-e98 [FREE Full text] [doi: 10.3747/co.25.3756] [Medline: 29507501]
58. Coker TR, Porras-Javier L, Zhang L, Soares N, Park C, Patel A, et al. A Telehealth-Enhanced Referral Process in Pediatric Primary Care: A Cluster Randomized Trial. Pediatrics 2019 Mar 15;143(3):e20182738. [doi: 10.1542/peds.2018-2738] [Medline: 30770523]
59. No authors listed. Collins BN, Lepore SJ, Winickoff JP, et al. An Office-Initiated Multilevel Intervention for Tobacco Smoke Exposure: A Randomized Trial. 2018;141(Suppl 1):S75-S86. Pediatrics 2018 Jun;141(6):e20180894 [FREE Full text] [doi: 10.1542/peds.2018-0894] [Medline: 29853625]
60. Condren M, Honey BL, Carter SM, Ngo N, Landsaw J, Bryant C, et al. Influence of a systems-based approach to prescribing errors in a pediatric resident clinic. Acad Pediatr 2014 Sep;14(5):485-490. [doi: 10.1016/j.acap.2014.03.018] [Medline: 25169160]
61. Conklin TM. Addressing alcohol misuse in primary care. J Am Assoc Nurse Pract 2020 Aug;32(8):583-588. [doi: 10.1097/jxx.0000000000000259]

62. Cordova D, Alers-Rojas F, Lua FM, Bauermeister J, Nurenberg R, Ovadje L, et al. The Usability and Acceptability of an Adolescent mHealth HIV/STI and Drug Abuse Preventive Intervention in Primary Care. Behav Med 2018 May 25;44(1):36-47 [FREE Full text] [doi: 10.1080/08964289.2016.1189396] [Medline: 27223646]
63. Curtis L, Mullen R, Russell A, Fata A, Bailey S, Makoul G, et al. An efficacy trial of an electronic health record-based strategy to inform patients on safe medication use: The role of written and spoken communication. Patient Educ Couns 2016 Sep;99(9):1489-1495 [FREE Full text] [doi: 10.1016/j.pec.2016.07.004] [Medline: 27444235]
64. Dahne J, Lejuez CW, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot Randomized Trial of a Self-Help Behavioral Activation Mobile App for Utilization in Primary Care. Behav Ther 2019 Jul;50(4):817-827 [FREE Full text] [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
65. Dean KM, Hatfield LA, Jena AB, Cristman D, Flair M, Kator K, et al. Preliminary Data on a Care Coordination Program for Home Care Recipients. J Am Geriatr Soc 2016 Sep 10;64(9):1900-1903. [doi: 10.1111/jgs.14351] [Medline: 27506164]
66. DeCamp LR, Godage SK, Valenzuela Araujo D, Dominguez Cortez J, Wu L, Psoter KJ, et al. A Texting Intervention in Latino Families to Reduce ED Use: A Randomized Trial. Pediatrics 2020 Jan;145(1) [FREE Full text] [doi: 10.1542/peds.2019-1405] [Medline: 31879276]
67. DeSantes K, Dodge A, Eickhoff J, Peterson AL. Improving Universal Pediatric Lipid Screening. J Pediatr 2017 Sep;188(5):87-90. [doi: 10.1016/j.jpeds.2017.05.030] [Medline: 28595766]
68. Devries J, Rafie S, Polston G. Implementing an overdose education and naloxone distribution program in a health system. J Am Pharm Assoc (2003) 2017 Apr;57(2S):S154-S160. [doi: 10.1016/j.japh.2017.01.002] [Medline: 28233681]
69. Dexter AS, Pope JF, Erickson D, Fontenot C, Ollendike E, Walker E. Cooking Education Improves Cooking Confidence and Dietary Habits in Veterans. Diabetes Educ 2019 Aug;45(4):442-449. [doi: 10.1177/0145721719848429] [Medline: 31072223]
70. Diaz MCG, Wysocki T, Crutchfield JH, Franciosi JP, Werk LN. Provider-Focused Intervention to Promote Comprehensive Screening for Adolescent Idiopathic Scoliosis by Primary Care Pediatricians. Am J Med Qual 2019;34(2):182-188. [doi: 10.1177/1062860618792667] [Medline: 30095983]
71. Dischinger HR, Cheng E, Mann AD, Grueber TM, Hawk S, Davis LA, et al. Decisional support to prevent adverse drug reactions of long latency: pilot randomized controlled intervention for glucocorticoid-induced diabetes. J Eval Clin Pract 2015 Aug 07;21(4):614-619. [doi: 10.1111/jep.12351] [Medline: 25851076]
72. Dixon BE, Simonaitis L, Perkins SM, Wright A, Middleton B. Measuring agreement between decision support reminders: the cloud vs. the local expert. BMC Med Inform Decis Mak 2014 Apr 10;14:31 [FREE Full text] [doi: 10.1186/1472-6947-14-31] [Medline: 24720863]
73. Dombkowski KJ, Cowan AE, Reeves SL, Foley MR, Dempsey AF. The impacts of email reminder/recall on adolescent influenza vaccination. Vaccine 2017 May 25;35(23):3089-3095. [doi: 10.1016/j.vaccine.2017.04.033] [Medline: 28455173]
74. Donovan JL, Kanaan AO, Gurwitz JH, Tjia J, Cutrona SL, Garber L, et al. A Pilot Health Information Technology-Based Effort to Increase the Quality of Transitions From Skilled Nursing Facility to Home: Compelling Evidence of High Rate of Adverse Outcomes. J Am Med Dir Assoc 2016 Apr 01;17(4):312-317. [doi: 10.1016/j.jamda.2015.11.008] [Medline: 26723801]
75. Duquaine D, Farley SM, Sacks R, Mandel-Ricci J, Silfen SL, Shih SC. Designing a quality improvement program with electronic health records: New York City's Health eQuits. Am J Med Qual 2015 Jan 29;30(2):141-148. [doi: 10.1177/1062860613520406] [Medline: 24477313]
76. Dwinnells R, Misik L. An Integrative Behavioral Health Care Model Using Automated SBIRT and Care Coordination in Community Health Care. J Prim Care Community Health 2017 Oct 16;8(4):300-304 [FREE Full text] [doi: 10.1177/2150131917740245] [Medline: 29144194]
77. Eckman MH, Lip GY, Wise RE, Speer B, Sullivan M, Walker N, et al. Using an Atrial Fibrillation Decision Support Tool for Thromboprophylaxis in Atrial Fibrillation: Effect of Sex and Age. J Am Geriatr Soc 2016 May 25;64(5):1054-1060 [FREE Full text] [doi: 10.1111/jgs.14099] [Medline: 27225358]
78. Eckman MH, Lip GY, Wise RE, Speer B, Sullivan M, Walker N, et al. Impact of an Atrial Fibrillation Decision Support Tool on thromboprophylaxis for atrial fibrillation. Am Heart J 2016 Jun;176:17-27. [doi: 10.1016/j.ahj.2016.02.009] [Medline: 27264216]
79. Emerson JF, Welch M, Rossman WE, Carek S, Ludden T, Templin M, et al. A Multidisciplinary Intervention Utilizing Virtual Communication Tools to Reduce Health Disparities: A Pilot Randomized Controlled Trial. Int J Environ Res Public Health 2015 Dec 22;13(1):ijerph13010031 [FREE Full text] [doi: 10.3390/ijerph13010031] [Medline: 26703661]
80. Engel CC, Litz B, Magruder KM, Harper E, Gore K, Stein N, et al. Delivery of self training and education for stressful situations (DESTRESS-PC): a randomized trial of nurse assisted online self-management for PTSD in primary care. Gen Hosp Psychiatry 2015 Jul;37(4):323-328 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.04.007] [Medline: 25929985]
81. Fanizza FA, Ruisinger JF, Prohaska ES, Melton BL. Integrating a health information exchange into a community pharmacy transitions of care service. J Am Pharm Assoc (2003) 2018;58(4):442-449. [doi: 10.1016/j.japh.2018.02.012] [Medline: 29625912]
82. Federman A, Sarzynski E, Brach C, Francaviglia P, Jacques J, Jandorf L, et al. Challenges optimizing the after visit summary. Int J Med Inform 2018 Dec;120:14-19 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.09.009] [Medline: 30409339]

83. Federman A, Kil N, Kannry J, Andreopolous E, Toribio W, Lyons J. An Electronic Health Record-based Intervention to Promote Hepatitis C Virus Testing Among Adults Born Between 1945 and 1965: a Cluster-randomized Trial. Medical Care 2017;55(6):590-597. [doi: 10.1097/mlr.0000000000000715]
84. Fiks AG, DuRivage N, Mayne SL, Finch S, Ross ME, Giacomini K, et al. Adoption of a Portal for the Primary Care Management of Pediatric Asthma: A Mixed-Methods Implementation Study. J Med Internet Res 2016 Jun 29;18(6):e172 [FREE Full text] [doi: 10.2196/jmir.5610] [Medline: 27357835]
85. Fiks AG, Luan X, Mayne SL. Improving HPV Vaccination Rates Using Maintenance-of-Certification Requirements. Pediatrics 2016 Mar 03;137(3):e20150675. [doi: 10.1542/peds.2015-0675] [Medline: 26908681]
86. Fiore M, Adsit R, Zehner M, McCarthy D, Lundsten S, Hartlaub P, et al. An electronic health record-based interoperable eReferral system to enhance smoking Quitline treatment in primary care. J Am Med Inform Assoc 2019 Aug 01;26(8-9):778-786 [FREE Full text] [doi: 10.1093/jamia/ocz044] [Medline: 31089727]
87. Fitzpatrick SL, Dickins K, Avery E, Ventrelle J, Shultz A, Kishen E, et al. Effect of an obesity best practice alert on physician documentation and referral practices. Transl Behav Med 2017 Dec 26;7(4):881-890 [FREE Full text] [doi: 10.1007/s13142-017-0514-0] [Medline: 28653221]
88. Flocke S, Seeholzer E, Lewis S, Gill I, Ordillas E, Rose J, et al. Designing for Sustainability: An Approach to Integrating Staff Role Changes and Electronic Health Record Functionality Within Safety-Net Clinics to Address Provision of Tobacco Cessation Care. Jt Comm J Qual Patient Saf 2019 Dec;45(12):798-807 [FREE Full text] [doi: 10.1016/j.jcjq.2019.09.003] [Medline: 31648946]
89. Flynn DM, Eaton LH, McQuinn H, Alden A, Meins AR, Rue T, et al. TelePain: Primary Care Chronic Pain Management through Weekly Didactic and Case-based Telementoring. Contemp Clin Trials Commun 2017 Dec;8:162-166 [FREE Full text] [doi: 10.1016/j.conctc.2017.10.004] [Medline: 29497708]
90. Frank JW, Carey EP, Fagan KM, Aron DC, Todd-Stenberg J, Moore BA, et al. Evaluation of a telementoring intervention for pain management in the Veterans Health Administration. Pain Med 2015 Jun;16(6):1090-1100. [doi: 10.1111/pme.12715] [Medline: 25716075]
91. Fried TR, Niehoff KM, Street RL, Charpentier PA, Rajeevan N, Miller PL, et al. Effect of the Tool to Reduce Inappropriate Medications on Medication Communication and Deprescribing. J Am Geriatr Soc 2017 Oct;65(10):2265-2271 [FREE Full text] [doi: 10.1111/jgs.15042] [Medline: 28804870]
92. Ganz DA, Kim S, Zingmond DS, Ramirez KD, Roth CP, Jennings LA, et al. Effect of a falls quality improvement program on serious fall-related injuries. J Am Geriatr Soc 2015 Jan 17;63(1):63-70 [FREE Full text] [doi: 10.1111/jgs.13154] [Medline: 25597558]
93. Geboy AG, Nichols WL, Fernandez SJ, Desale S, Basch P, Fishbein DA. Leveraging the electronic health record to eliminate hepatitis C: Screening in a large integrated healthcare system. PLoS One 2019 May 23;14(5):e0216459 [FREE Full text] [doi: 10.1371/journal.pone.0216459] [Medline: 31120906]
94. Gill JM, Chen YX, Grimes A, Klinkman MS. Using electronic health record-based tools to screen for bipolar disorder in primary care patients with depression. J Am Board Fam Med 2012 May 08;25(3):283-290 [FREE Full text] [doi: 10.3122/jabfm.2012.03.110217] [Medline: 22570391]
95. Gold J, Reyes-Gastelum D, Turner J, Pediatric Asthma Quality Team, Davies HD. A quality improvement study using fishbone analysis and an electronic medical records intervention to improve care for children with asthma. Am J Med Qual 2014;29(1):70-77. [doi: 10.1177/1062860613484171] [Medline: 23574643]
96. Goleman MJ, Dolce M, Morack J. Quality Improvement Initiative to Improve Human Papillomavirus Vaccine Initiation at 9 Years of Age. Acad Pediatr 2018 Sep;18(7):769-775. [doi: 10.1016/j.acap.2018.05.005] [Medline: 29842924]
97. Grant RW, Ashburner JM, Jernigan MC, Chang J, Borowsky LH, Chang Y, et al. Randomized trial of a health IT tool to support between-visit-based laboratory monitoring for chronic disease medication prescriptions. J Gen Intern Med 2015 May 6;30(5):619-625 [FREE Full text] [doi: 10.1007/s11606-014-3152-y] [Medline: 25560319]
98. Greene BR, McManus K, Redmond SJ, Caulfield B, Quinn CC. Digital assessment of falls risk, frailty, and mobility impairment using wearable sensors. npj Digit. Med 2019 Dec 11;2(1):125 [FREE Full text] [doi: 10.1038/s41746-019-0204-z] [Medline: 31840096]
99. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A Smartphone Application to Support Recovery From Alcoholism. JAMA Psychiatry 2014 May 01;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
100. Ha B, O'Sullivan DL, Diamond CA, Plumb AJ, Sleeth JS, Greer FR, et al. Improving Rates of Screening for Anemia in Infancy. Clin Pediatr (Phila) 2017 Nov 28;57(9):1064-1068. [doi: 10.1177/00099228177446608]
101. Haas JS, Baer HJ, Eibensteiner K, Klinger EV, St. Hubert S, Getty G, et al. A Cluster Randomized Trial of a Personalized Multi-Condition Risk Assessment in Primary Care. American Journal of Preventive Medicine 2017 Jan;52(1):100-105. [doi: 10.1016/j.amepre.2016.07.013]
102. Haas JS, Linder JA, Park ER, Gonzalez I, Rigotti NA, Klinger EV, et al. Proactive Tobacco Cessation Outreach to Smokers of Low Socioeconomic Status. JAMA Intern Med 2015 Feb 01;175(2):218-226 [FREE Full text] [doi: 10.1001/jamainternmed.2014.6674] [Medline: 25506771]

103. Halterman JS, Fagnano M, Tajon RS, Tremblay P, Wang H, Butz A, et al. Effect of the School-Based Telemedicine Enhanced Asthma Management (SB-TEAM) Program on Asthma Morbidity. JAMA Pediatr 2018 Mar 01;172(3):e174938 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.4938] [Medline: 29309483]
104. Hark L, Acito M, Adeghate J, Henderer J, Okudolo J, Malik K, et al. Philadelphia Telemedicine Glaucoma Detection and Follow-up Study: Ocular Findings at Two Health Centers. Journal of Health Care for the Poor and Underserved 2018;29(4):1400-1415. [doi: 10.1353/hpu.2018.0103]
105. Hark LA, Katz LJ, Myers JS, Waisbourd M, Johnson D, Pizzi LT, et al. Philadelphia Telemedicine Glaucoma Detection and Follow-up Study: Methods and Screening Results. American Journal of Ophthalmology 2017 Sep;181:114-124. [doi: 10.1016/j.ajo.2017.06.024]
106. Hark LA, Myers JS, Pasquale LR, Razeghinejad MR, Maity A, Zhan T, et al. Philadelphia Telemedicine Glaucoma Detection and Follow-up Study. Journal of Glaucoma 2019;28(4):294-301. [doi: 10.1097/ijg.0000000000001207]
107. Hatef E, Alexander M, Vanderver B, Fagan P, Albert M. Assessment of annual diabetic eye examination using telemedicine technology among underserved patients in primary care setting. Middle East Afr J Ophthalmol 2017;24(4):207. [doi: 10.4103/meajo.meajo_19_16]
108. Hay JW, Lee P, Jin H, Guterman JJ, Gross-Schulman S, Ell K, et al. Cost-Effectiveness of a Technology-Facilitated Depression Care Management Adoption Model in Safety-Net Primary Care Patients with Type 2 Diabetes. Value in Health 2018 May;21(5):561-568. [doi: 10.1016/j.jval.2017.11.005] [Medline: 29753353]
109. Hayek S, Nieva R, Corrigan F, Zhou A, Mudaliar U, Mays D, et al. End-of-Life Care Planning: Improving Documentation of Advance Directives in the Outpatient Clinic Using Electronic Medical Records. Journal of Palliative Medicine 2014 Dec;17(12):1348-1352. [doi: 10.1089/jpm.2013.0684] [Medline: 24988497]
110. Hess R, Fischer GS, Sullivan SM, Dong X, Weimer M, Zeith C, et al. Patterns of Response to Patient-Centered Decision Support Through a Personal Health Record. Telemedicine and e-Health 2014 Nov;20(11):984-989. [doi: 10.1089/tmj.2013.0332]
111. Heyworth L, Kleinman K, Oddleifson S, Bernstein L, Frampton J, Lehrer M, et al. Comparison of interactive voice response, patient mailing, and mailed registry to encourage screening for osteoporosis: a randomized controlled trial. Osteoporos Int 2014 Feb 25;25(5):1519-1526. [doi: 10.1007/s00198-014-2629-1]
112. Hojat L, Avery A, Greco PJ, Kaelber DC. Doubling Hepatitis C Virus Screening in Primary Care Using Advanced Electronic Health Record Tools—A Non-Randomized Controlled Trial. J GEN INTERN MED 2019 Dec 02;35(2):498-504. [doi: 10.1007/s11606-019-05536-z]
113. Honaker SM, Dugan T, Daftary A, Davis S, Saha C, Baye F, et al. Unexplained Practice Variation in Primary Care Providers' Concern for Pediatric Obstructive Sleep Apnea. Academic Pediatrics 2018 May;18(4):418-424. [doi: 10.1016/j.acap.2018.01.011]
114. Honaker SM, Street A, Daftary AS, Downs SM. The Use of Computer Decision Support for Pediatric Obstructive Sleep Apnea Detection in Primary Care. Journal of Clinical Sleep Medicine 2019 Mar 15;15(03):453-462. [doi: 10.5664/jcsm.7674]
115. Hoopes M, Angier H, Raynor L, Suchocki A, Muench J, Marino M, et al. Development of an algorithm to link electronic health record prescriptions with pharmacy dispense claims. Journal of the American Medical Informatics Association 2018;25(10):1322-1330. [doi: 10.1093/jamia/ocy095]
116. Jain M, Rich N, Ahn C, Turner B, Sanders J, Adamson B, et al. Evaluation of a Multifaceted Intervention to Reduce Health Disparities in Hepatitis C Screening: A Pre‐Post Analysis. Hepatology 2019 May 31:A [FREE Full text] [doi: 10.1002/hep.30638]
117. Jani PD, Forbes L, Choudhury A, Preisser JS, Viera AJ, Garg S. Evaluation of Diabetic Retinal Screening and Factors for Ophthalmology Referral in a Telemedicine Network. JAMA Ophthalmol 2017 Jul 01;135(7):706. [doi: 10.1001/jamaophthalmol.2017.1150]
118. Jenssen BP, Bryant-Stephens T, Leone FT, Grundmeier RW, Fiks AG. Clinical Decision Support Tool for Parental Tobacco Treatment in Primary Care. Pediatrics 2016 Apr 14;137(5):e20154185. [doi: 10.1542/peds.2015-4185]
119. Jenssen BP, Muthu N, Kelly MK, Baca H, Shults J, Grundmeier RW, et al. Parent eReferral to Tobacco Quitline: A Pragmatic Randomized Trial in Pediatric Primary Care. American Journal of Preventive Medicine 2019 Jul;57(1):32-40. [doi: 10.1016/j.amepre.2019.03.005]
120. Jetelina KK, Woodson TT, Gunn R, Muller B, Clark KD, DeVoe JE, et al. Evaluation of an Electronic Health Record (EHR) Tool for Integrated Behavioral Health in Primary Care. J Am Board Fam Med 2018 Sep 10;31(5):712-723. [doi: 10.3122/jabfm.2018.05.180041]
121. Jones Q, Johnston B, Biola H, Gomez S, Crowder C. Implementing standardized substance use disorder screening in primary care. Journal of the American Academy of Physician Assistants 2018;31(10):42-45. [doi: 10.1097/01.jaa.0000795888.70937.34]
122. Kamo N, Williams B, Smith D, Blackmore C. Improving the Electronic Capture of Advance Care Directives in a Healthcare Delivery System. J Am Geriatr Soc 2017 Feb 09;65(5):973-979 [FREE Full text] [doi: 10.1111/jgs.14695]
123. Kapoor A, Amroze A, Golden J, Crawford S, O'Day K, Elhag R, et al. SUPPORT‐AF: Piloting a Multi‐Faceted, Electronic Medical Record‐Based Intervention to Improve Prescription of Anticoagulation. JAHA 2018 Sep 04;7(17):e009946. [doi: 10.1161/jaha.118.009946]

124. Karas D, Sondike S, Fitzgibbon J, Redding M, Brown M. Using a Clinical Decision Support Tool to Increase Chlamydia Screening Across a Large Primary Care Pediatric Network. Clin Pediatr (Phila) 2018 Sep 28;57(14):1638-1641. [doi: 10.1177/0009922818803397]
125. Kavanagh P, Sobota A, McClure E, Sprinz P, Adams W. Using an Electronic Health Record-Based Registry to Improve Pediatric Sickle Cell Care. Journal of Clinical Outcomes Management 2014;21(4):159-168.
126. Keck JW, Roper KL, Hieronymus LB, Thomas AR, Huang Z, Westgate PM, et al. Primary Care Cluster RCT to Increase Diabetes Prevention Program Referrals. American Journal of Preventive Medicine 2020 Jul;59(1):79-87. [doi: 10.1016/j.amepre.2020.02.008]
127. Kershaw C, Taylor JL, Horowitz G, Brockmeyer D, Libman H, Kriegel G, et al. Use of an electronic medical record reminder improves HIV screening. BMC Health Serv Res 2018 Jan 10;18(1):14. [doi: 10.1186/s12913-017-2824-9]
128. Kim RH, Day SC, Small DS, Snider CK, Rareshide CAL, Patel MS. Variations in Influenza Vaccination by Clinic Appointment Time and an Active Choice Intervention in the Electronic Health Record to Increase Influenza Vaccination. JAMA Netw Open 2018 Sep 14;1(5):e181770. [doi: 10.1001/jamanetworkopen.2018.1770]
129. Kinahan KE, Kircher S, Altman J, Rademaker A, Salsman JM, Didwania A, et al. Promoting the Shared-Care Model for Adolescent and Young Adults With Cancer: Optimizing Referrals and Care Coordination With Primary Care Providers. J Natl Compr Canc Netw 2016 Dec 31;15(1):38-44. [doi: 10.6004/jnccn.2017.0005]
130. Knierim KE, Hall TL, Dickinson LM, Nease DE, de la Cerda DR, Fernald D, et al. Primary Care Practices' Ability to Report Electronic Clinical Quality Measures in the EvidenceNOW Southwest Initiative to Improve Heart Health. JAMA Netw Open 2019 Aug 07;2(8):e198569. [doi: 10.1001/jamanetworkopen.2019.8569]
131. Krishnan A, Finkelstein EA, Levine E, Foley P, Askew S, Steinberg D, et al. A Digital Behavioral Weight Gain Prevention Intervention in Primary Care Practice: Cost and Cost-Effectiveness Analysis. J Med Internet Res 2019 May 17;21(5):e12201. [doi: 10.2196/12201]
132. Kroenke K, Baye F, Lourens SG, Evans E, Weitlauf S, McCalley S, et al. Automated Self-management (ASM) vs. ASM-Enhanced Collaborative Care for Chronic Pain and Mood Symptoms: the CAMMPS Randomized Clinical Trial. J GEN INTERN MED 2019 Jun 21;34(9):1806-1814. [doi: 10.1007/s11606-019-05121-4]
133. Kruger DL, Rein DB, Kil N, Jordan C, Brown KA, Yartel A, et al. Implementation of Birth-Cohort Testing for Hepatitis C Virus. Health Promotion Practice 2016 Aug 19;18(2):283-289. [doi: 10.1177/1524839916661495]
134. Kukafka R, Fang J, Vanegas A, Silverman T, Crew KD. Pilot study of decision support tools on breast cancer chemoprevention for high-risk women and healthcare providers in the primary care setting. BMC Med Inform Decis Mak 2018 Dec 17;18(1):134. [doi: 10.1186/s12911-018-0716-5]
135. Kukafka R, Jeong IC, Finkelstein J. Optimizing Decision Support for Tailored Health Behavior Change Applications. Stud Health Technol Inform 2015;216:108-112. [Medline: 26262020]
136. Ladapo JA, Lyons H, Yau M, Rich P, Newton D, Bruce-Mensah K, et al. Enhanced Assessment of Chest Pain and Related Symptoms in the Primary Care Setting Through the Use of a Novel Personalized Medicine Genomic Test. Am J Med Qual 2014 May 05;30(4):345-352. [doi: 10.1177/1062860614532517]
137. Landis SE, Galvin SL. Implementation and Assessment of a Fall Screening Program in Primary Care Practices. J Am Geriatr Soc 2014 Dec 17;62(12):2408-2414. [doi: 10.1111/jgs.13137]
138. Langford AT, Wang B, Orzeck-Byrnes NA, Aidasani SR, Hu L, Applegate M, et al. Sociodemographic and clinical correlates of key outcomes from a Mobile Insulin Titration Intervention (MITI) for medically underserved patients. Patient Education and Counseling 2019 Mar;102(3):520-527. [doi: 10.1016/j.pec.2018.09.016]
139. Lanpher MG, Askew S, Bennett GG. Health Literacy and Weight Change in a Digital Health Intervention for Women: A Randomized Controlled Trial in Primary Care Practice. Journal of Health Communication 2016 Apr 04;21(sup1):34-42. [doi: 10.1080/10810730.2015.1131773]
140. Leddy J, Green JA, Yule C, Molecavage J, Coresh J, Chang AR. Improving proteinuria screening with mailed smartphone urinalysis testing in previously unscreened patients with hypertension: a randomized controlled trial. BMC Nephrol 2019 Apr 18;20(1). [doi: 10.1186/s12882-019-1324-z]
141. Lepore S, Collins B, Coffman D, Winickoff J, Nair U, Moughan B, et al. Kids Safe and Smokefree (KiSS) Multilevel Intervention to Reduce Child Tobacco Smoke Exposure: Long-Term Results of a Randomized Controlled Trial. IJERPH 2018 Jun 12;15(6):1239. [doi: 10.3390/ijerph15061239]
142. Levy S, Henderson L, McAlpine C. Growing up with confidence: using telehealth to support continence self-care deficits amongst young people with complex needs. jhi 2014 May 27;21(3):113-117. [doi: 10.14236/jhi.v21i3.58]
143. Lewis KH, Skelton JA, Hsu F, Ezouah P, Taveras EM, Block JP. Implementing a novel electronic health record approach to track child sugar-sweetened beverage consumption. Preventive Medicine Reports 2018 Sep;11:169-175. [doi: 10.1016/j.pmedr.2018.06.007]
144. Liebschutz JM, Xuan Z, Shanahan CW, LaRochelle M, Keosaian J, Beers D, et al. Improving Adherence to Long-term Opioid Therapy Guidelines to Reduce Opioid Misuse in Primary Care. JAMA Intern Med 2017 Sep 01;177(9):1265. [doi: 10.1001/jamainternmed.2017.2468]

145. Lin CJ, Nowalk MP, Pavlik VN, Brown AE, Zhang S, Raviotta JM, et al. Using the 4 pillars™ practice transformation program to increase adult influenza vaccination and reduce missed opportunities in a randomized cluster trial. BMC Infect Dis 2016 Nov 3;16(1):623. [doi: 10.1186/s12879-016-1940-1]
146. Litke J, Spoutz L, Ahlstrom D, Perdew C, Llamas W, Erickson K. Impact of the clinical pharmacy specialist in telehealth primary care. American Journal of Health-System Pharmacy 2018;75(13):982-986. [doi: 10.2146/ajhp170633]
147. Liu Y, Zupan NJ, Swearingen R, Jacobson N, Carlson JN, Mahoney JE, et al. Identification of barriers, facilitators and system-based implementation strategies to increase teleophthalmology use for diabetic eye screening in a rural US primary care clinic: a qualitative study. BMJ Open 2019 Feb 18;9(2):e022594. [doi: 10.1136/bmjopen-2018-022594]
148. Looman W, Antolick M, Cady R, Lunos S, Garwick A, Finkelstein S. Effects of a Telehealth Care Coordination Intervention on Perceptions of Health Care by Caregivers of Children With Medical Complexity: A Randomized Controlled Trial. Journal of Pediatric Health Care 2015 Jul;29(4):352-363 [FREE Full text] [doi: 10.1016/j.pedhc.2015.01.007]
149. Lv N, Xiao L, Simmons ML, Rosas LG, Chan A, Entwistle M. Personalized Hypertension Management Using Patient-Generated Health Data Integrated With Electronic Health Records (EMPOWER-H): Six-Month Pre-Post Study. J Med Internet Res 2017 Sep 19;19(9):e311 [FREE Full text] [doi: 10.2196/jmir.7831] [Medline: 28928111]
150. MacLean CD, Berger C, Cangiano ML, Ziegelman D, Lidofsky SD. Impact of electronic reminder systems on hepatitis C screening in primary care. J Viral Hepat 2018 Mar 30;25(8):939-944. [doi: 10.1111/jvh.12885]
151. Maddali MV, Mehtani NJ, Converse C, Kapoor S, Pham P, Li JZ, et al. Development and Validation of HIV-ASSIST, an Online, Educational, Clinical Decision Support Tool to Guide Patient-Centered ARV Regimen Selection. J Acquir Immune Defic Syndr 2019 Oct 01;82(2):188-194. [doi: 10.1097/QAI.0000000000002118] [Medline: 31513553]
152. Mahabee-Gittens EM, Dexheimer JW, Tabangin M, Khoury JC, Merianos AL, Stone L, et al. An Electronic Health Record−Based Strategy to Address Child Tobacco Smoke Exposure. American Journal of Preventive Medicine 2018 Jan;54(1):64-71. [doi: 10.1016/j.amepre.2017.08.011]
153. Mahoney MC, Erwin DO, Twarozek AM, Saad-Harfouche FG, Rodriguez EM, Sun X, et al. Leveraging technology to promote smoking cessation in urban and rural primary care medical offices. Preventive Medicine 2018 Sep;114:102-106 [FREE Full text] [doi: 10.1016/j.ypmed.2018.06.016] [Medline: 29953897]
154. Mainous AG, Carek PJ, Lynch K, Tanner RJ, Hulihan MM, Baskin J, et al. Effectiveness of Clinical Decision Support Based Intervention in the Improvement of Care for Adult Sickle Cell Disease Patients in Primary Care. J Am Board Fam Med 2018 Sep 10;31(5):812-816. [doi: 10.3122/jabfm.2018.05.180106]
155. Mann DM, Palmisano J, Lin JJ. A pilot randomized trial of technology-assisted goal setting to improve physical activity among primary care patients with prediabetes. Preventive Medicine Reports 2016 Dec;4:107-112. [doi: 10.1016/j.pmedr.2016.05.012]
156. Mansberger SL, Sheppler C, Barker G, Gardiner SK, Demirel S, Wooten K, et al. Long-term Comparative Effectiveness of Telemedicine in Providing Diabetic Retinopathy Screening Examinations. JAMA Ophthalmol 2015 May 01;133(5):518. [doi: 10.1001/jamaophthalmol.2015.1]
157. Marcelin JR, Tan EM, Marcelin A, Scheitel M, Ramu P, Hankey R, et al. Assessment and improvement of HIV screening rates in a Midwest primary care practice using an electronic clinical decision support system: a quality improvement study. BMC Med Inform Decis Mak 2016 Jul 4;16(1):76. [doi: 10.1186/s12911-016-0320-5]
158. Maxwell LG, McFarland MS, Baker JW, Cassidy RF. Evaluation of the Impact of a Pharmacist-Led Telehealth Clinic on Diabetes-Related Goals of Therapy in a Veteran Population. Pharmacotherapy 2016 Mar 11;36(3):348-356. [doi: 10.1002/phar.1719]
159. McAdam-Marx C, Tak C, Petigara T, Jones NW, Yoo M, Briley MS, et al. Impact of a guideline-based best practice alert on pneumococcal vaccination rates in adults in a primary care setting. BMC Health Serv Res 2019 Jul 10;19(1):474. [doi: 10.1186/s12913-019-4263-2]
160. McGrath K, Foster K, Doggett P, Altshuler M, Osborne-Wu J, Castellan C, et al. EMR-Based Intervention Improves Lead Screening at an Urban Family Medicine Practice. Fam Med 2016 Nov;48(10):801-804 [FREE Full text] [Medline: 27875603]
161. McLendon SF, Wood FG, Stanley N. Enhancing diabetes care through care coordination, telemedicine, and education: Evaluation of a rural pilot program. Public Health Nurs 2019 Mar 13;36(3):310-320. [doi: 10.1111/phn.12601]
162. Meeker D, Linder JA, Fox CR, Friedberg MW, Persell SD, Goldstein NJ, et al. Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices. JAMA 2016 Feb 09;315(6):562. [doi: 10.1001/jama.2016.0275]
163. Mehta S, Khan T, Guerra C, Reitz C, McAuliffe T, Volpp K. A Randomized Controlled Trial of Opt-in Versus Opt-Out Colorectal Cancer Screening Outreach. American Journal of Gastroenterology 2018;113(12):1848-1854. [doi: 10.1038/s41395-018-0151-3]
164. Mera J, Vellozzi C, Hariri S, Carabin H, Drevets DA, Miller A, et al. Identification and Clinical Management of Persons with Chronic Hepatitis C Virus Infection — Cherokee Nation, 2012–2015. MMWR Morb. Mortal. Wkly. Rep 2016 May 13;65(18):461-466. [doi: 10.15585/mmwr.mm6518a2]
165. Militello L, Diiulio J, Borders M, Sushereba C, Saleem J, Haverkamp D, et al. Evaluating a Modular Decision Support Application For Colorectal Cancer Screening. Appl Clin Inform 2017 Dec 20;26(01):162-179. [doi: 10.4338/aci-2016-09-ra-0152]

166. Mitchell SL, Shaffer ML, Cohen S, Hanson LC, Habtemariam D, Volandes AE. An Advance Care Planning Video Decision Support Tool for Nursing Home Residents With Advanced Dementia. JAMA Intern Med 2018 Jul 01;178(7):961. [doi: 10.1001/jamainternmed.2018.1506]
167. Molleda L, Bahamon M, St. George SM, Perrino T, Estrada Y, Correa Herrera D, et al. Clinic Personnel, Facilitator, and Parent Perspectives of eHealth Familias Unidas in Primary Care. Journal of Pediatric Health Care 2017 May;31(3):350-361 [FREE Full text] [doi: 10.1016/j.pedhc.2016.11.001] [Medline: 28012799]
168. Mundt MP, Zakletskaia LI. Putting the pieces together: EHR communication and diabetes patient outcomes. Am J Manag Care 2018 Oct;24(10):462-468 [FREE Full text] [Medline: 30325187]
169. Mussman GM, Vossmeyer MT, Brady PW, Warrick DM, Simmons JM, White CM. Improving the reliability of verbal communication between primary care physicians and pediatric hospitalists at hospital discharge. J. Hosp. Med 2015 May 29;10(9):574-580. [doi: 10.1002/jhm.2392]
170. Mussulman L, Ellerbeck EF, Cupertino AP, Preacher KJ, Spaulding R, Catley D, et al. Design and participant characteristics of a randomized-controlled trial of telemedicine for smoking cessation among rural smokers. Contemporary Clinical Trials 2014 Jul;38(2):173-181. [doi: 10.1016/j.cct.2014.04.008]
171. Nagykaldi ZJ, Scheid D, Zhao D, Mishra B, Greever-Rice T. An Innovative Community-based Model for Improving Preventive Care in Rural Counties. J Am Board Fam Med 2017 Sep 18;30(5):583-591. [doi: 10.3122/jabfm.2017.05.170035]
172. Nagykaldi Z, Yeaman B, Jones M, Mold J, Scheid D. HIE-i-health information exchange with intelligence. Journal of Ambulatory Care Management 2014;37(1):20-31. [doi: 10.1097/jac.0000000000000002]
173. Nitsche B, Miller SC, Giorgio M, Berry CA, Muir A. Improving Hepatitis C Identification: Technology Alone Is Not the Answer. Health Promotion Practice 2017 Sep 11;19(4):506-512. [doi: 10.1177/1524839917725501]
174. Nowalk MP, Lin CJ, Pavlik VN, Brown AE, Zhang S, Moehling KK, et al. Using the 4 Pillars™ Practice Transformation Program to increase adult Tdap immunization in a randomized controlled cluster trial. Vaccine 2016 Sep;34(41):5026-5033. [doi: 10.1016/j.vaccine.2016.07.053]
175. O'Connor PJ, Magid DJ, Sperl-Hillen JM, Price DW, Asche SE, Rush WA, et al. Personalised physician learning intervention to improve hypertension and lipid control: randomised trial comparing two methods of physician profiling. BMJ Qual Saf 2014 Sep 16;23(12):1014-1022. [doi: 10.1136/bmjqs-2014-002807]
176. Ofili EO, Pemu PE, Quarshie A, Mensah EA, Rollins L, Ojutalayo F, et al. Democratizing Discovery Health with N=Me. Trans Am Clin Climatol Assoc 2018;129:215-234 [FREE Full text] [Medline: 30166716]
177. Okah A, Williams K, Talib N, Mann K. Promoting Oral Health in Childhood: A Quality Improvement Project. Pediatrics 2018 Jun 25;141(6):e20172396. [doi: 10.1542/peds.2017-2396] [Medline: 29802117]
178. Osofsky HJ, Osofsky JD, Hansel TC, Flynn T. The Louisiana Mental and Behavioral Health Capacity Project Trauma-Informed Integrated Care Model and Improved Posttraumatic Stress Outcomes. J Public Health Manag Pract 2017 Dec;23 Suppl 6 Suppl, Gulf Region Health Outreach Program:S25-S31. [doi: 10.1097/PHH.0000000000000652] [Medline: 28961649]
179. Ozanne EM, Howe R, Omer Z, Esserman LJ. Development of a personalized decision aid for breast cancer risk reduction and management. BMC Med Inform Decis Mak 2014 Jan 14;14(1):4 [FREE Full text] [doi: 10.1186/1472-6947-14-4] [Medline: 24422989]
180. Peralta CA, Frigaard M, Rolon L, Seal K, Tuot D, Senyak J, et al. Screening for CKD To Improve Processes of Care among Nondiabetic Veterans with Hypertension. CJASN 2020 Feb 07;15(2):174-181. [doi: 10.2215/cjn.05050419]
181. Perry TT, Halterman JS, Brown RH, Luo C, Randle SM, Hunter CR, et al. Results of an asthma education program delivered via telemedicine in rural schools. Ann Allergy Asthma Immunol 2018 Apr;120(4):401-408 [FREE Full text] [doi: 10.1016/j.anai.2018.02.013] [Medline: 29471032]
182. Peters KM, Sadler G, Miller E, Radovic A. An Electronic Referral and Social Work Protocol to Improve Access to Mental Health Services. Pediatrics 2018 Nov;142(5). [doi: 10.1542/peds.2017-2417] [Medline: 30282783]
183. Petersen A, Mermelstein R, Berg KM, Baker TB, Smith SS, Jorenby D, et al. Offering smoking treatment to primary care patients in two Wisconsin healthcare systems: Who chooses smoking reduction versus cessation? Prev Med 2017 Dec;105:332-336 [FREE Full text] [doi: 10.1016/j.ypmed.2017.10.015] [Medline: 28988997]
184. Phillips LS, Barb D, Yong C, Tomolo AM, Jackson SL, Olson DE, et al. Translating What Works: A New Approach to Improve Diabetes Management. J Diabetes Sci Technol 2015 Jul 09;9(4):857-864 [FREE Full text] [doi: 10.1177/1932296815576000] [Medline: 25759182]
185. Quanbeck A, Gustafson DH, Marsch LA, Chih M, Kornfield R, McTavish F, et al. Implementing a Mobile Health System to Integrate the Treatment of Addiction Into Primary Care: A Hybrid Implementation-Effectiveness Study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
186. Reeves K, Taylor Y, Tapp H, Ludden T, Shade L, Burton B, et al. Evaluation of a Pilot Asthma Care Program for Electronic Communication between School Health and a Healthcare System's Electronic Medical Record. Appl Clin Inform 2017 Dec 18;07(04):969-982 [FREE Full text] [doi: 10.4338/aci-2016-02-ra-0022]
187. Regan ME. Implementing an evidence-based clinical decision support tool to improve the detection, evaluation, and referral patterns of adult chronic kidney disease patients in primary care. J Am Assoc Nurse Pract 2017 Dec;29(12):741-753. [doi: 10.1002/2327-6924.12505] [Medline: 28875593]

188. Richardson LP, Zhou C, Gersh E, Spielvogle H, Taylor JA, McCarty CA. Effect of Electronic Screening With Personalized Feedback on Adolescent Health Risk Behaviors in a Primary Care Setting: A Randomized Clinical Trial. JAMA Netw Open 2019 May 03;2(5):e193581 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.3581] [Medline: 31074815]
189. Richter KP, Shireman TI, Ellerbeck EF, Cupertino AP, Catley D, Cox LS, et al. Comparative and cost effectiveness of telemedicine versus telephone counseling for smoking cessation. J Med Internet Res 2015 May 08;17(5):e113 [FREE Full text] [doi: 10.2196/jmir.3975] [Medline: 25956257]
190. Ripley-Moffitt C, Neutze D, Gwynne M, Goldstein AO. Patient care outcomes of a tobacco use registry in an academic family practice. J Am Board Fam Med 2015 Mar 06;28(2):205-213 [FREE Full text] [doi: 10.3122/jabfm.2015.02.140121] [Medline: 25748761]
191. Rollman BL, Belnap BH, Mazumdar S, Abebe KZ, Karp JF, Lenze EJ, et al. Telephone-Delivered Stepped Collaborative Care for Treating Anxiety in Primary Care: A Randomized Controlled Trial. J Gen Intern Med 2017 Mar;32(3):245-255 [FREE Full text] [doi: 10.1007/s11606-016-3873-1] [Medline: 27714649]
192. Rose BL, Leung S, Gustin J, Childers J. Initiating Advance Care Planning in Primary Care: A Model for Success. J Palliat Med 2019 Apr;22(4):427-431. [doi: 10.1089/jpm.2018.0380] [Medline: 30481086]
193. Satre DD, Leibowitz AS, Leyden W, Catz SL, Hare CB, Jang H, et al. Interventions to Reduce Unhealthy Alcohol Use among Primary Care Patients with HIV: the Health and Motivation Randomized Clinical Trial. J Gen Intern Med 2019 Oct 11;34(10):2054-2061 [FREE Full text] [doi: 10.1007/s11606-019-05065-9] [Medline: 31187344]
194. Scheuner MT, Hamilton AB, Peredo J, Sale TJ, Austin C, Gilman SC, et al. A cancer genetics toolkit improves access to genetic services through documentation and use of the family history by primary-care clinicians. Genet Med 2014 Jan 13;16(1):60-69. [doi: 10.1038/gim.2013.75] [Medline: 23765051]
195. Schiff GD, Klinger E, Salazar A, Medoff J, Amato MG, John Orav E, et al. Screening for Adverse Drug Events: a Randomized Trial of Automated Calls Coupled with Phone-Based Pharmacist Counseling. J Gen Intern Med 2019 Feb;34(2):285-292 [FREE Full text] [doi: 10.1007/s11606-018-4672-7] [Medline: 30291602]
196. Schlittenhardt M, Smith SC, Ward-Smith P. Tele-Continence Care: A Novel Approach for Providers. Urol Nurs 2016;36(5):217-223. [Medline: 29240333]
197. Senft N, Everson J. eHealth Engagement as a Response to Negative Healthcare Experiences: Cross-Sectional Survey Analysis. J Med Internet Res 2018 Dec 05;20(12):e11034 [FREE Full text] [doi: 10.2196/11034] [Medline: 30518513]
198. Sequist TD, Holliday AM, Orav EJ, Bates DW, Denker BM. Physician and patient tools to improve chronic kidney disease care. Am J Manag Care 2018 Apr 01;24(4):e107-e114 [FREE Full text] [Medline: 29668213]
199. Shade SB, Steward WT, Koester KA, Chakravarty D, Myers JJ. Health information technology interventions enhance care completion, engagement in HIV care and treatment, and viral suppression among HIV-infected patients in publicly funded settings. J Am Med Inform Assoc 2015 Apr;22(e1):e104-e111 [FREE Full text] [doi: 10.1136/amiajnl-2013-002623] [Medline: 25030033]
200. Shah SD, Prine L, Waltermaurer E, Rubin SE. Feasibility study of family planning services screening as clinical decision support at an urban Federally Qualified Health Center network. Contraception 2019 Jan;99(1):27-31. [doi: 10.1016/j.contraception.2018.10.004] [Medline: 30336133]
201. Shaibi GQ, Kullo IJ, Singh DP, Sharp RR, De Filippis E, Cuellar I, et al. Developing a Process for Returning Medically Actionable Genomic Variants to Latino Patients in a Federally Qualified Health Center. Public Health Genomics 2018;21(1-2):77-84 [FREE Full text] [doi: 10.1159/000494488] [Medline: 30522109]
202. Shaikh U, Nettiksimmons J, Joseph JG, Tancredi D, Romano PS. Collaborative practice improvement for childhood obesity in rural clinics: the Healthy Eating Active Living Telehealth Community of Practice (HEALTH COP). Am J Med Qual 2014 Oct 29;29(6):467-475. [doi: 10.1177/1062860613506252] [Medline: 24170936]
203. Shane-McWhorter L, Lenert L, Petersen M, Woolsey S, McAdam-Marx C, Coursey JM, et al. The Utah Remote Monitoring Project: improving health care one patient at a time. Diabetes Technol Ther 2014 Oct;16(10):653-660 [FREE Full text] [doi: 10.1089/dia.2014.0045] [Medline: 24991923]
204. Sharifi M, Adams WG, Winickoff JP, Guo J, Reid M, Boynton-Jarrett R. Enhancing the electronic health record to increase counseling and quit-line referral for parents who smoke. Acad Pediatr 2014 Sep;14(5):478-484. [doi: 10.1016/j.acap.2014.03.017] [Medline: 25169159]
205. Sharifi M, Franz C, Horan CM, Giles CM, Long MW, Ward ZJ, et al. Cost-Effectiveness of a Clinical Childhood Obesity Intervention. Pediatrics 2017 Nov 31;140(5):e20162998 [FREE Full text] [doi: 10.1542/peds.2016-2998] [Medline: 29089403]
206. Shuen JA, Wilson MP, Kreshak A, Mullinax S, Brennan J, Castillo EM, et al. Telephoned, Texted, or Typed Out: A Randomized Trial of Physician-Patient Communication After Emergency Department Discharge. J Emerg Med 2018 Oct;55(4):573-581 [FREE Full text] [doi: 10.1016/j.jemermed.2018.07.023] [Medline: 30181075]
207. Sidlow R, Msaouel P. Improving Hepatitis C Virus Screening Rates in Primary Care: A Targeted Intervention Using the Electronic Health Record. Journal for Healthcare Quality 2015;37(5):319-323. [doi: 10.1097/jhq.0000000000000010]
208. Sigel E, Harpin SB, Tung G. Increasing documentation and referral for youth at risk for violence through the primary health care setting. Clin Pediatr (Phila) 2015 May 09;54(5):451-457. [doi: 10.1177/0009922814553431] [Medline: 25305258]

209. Siminerio L, Ruppert K, Huber K, Toledo FGS. Telemedicine for Reach, Education, Access, and Treatment (TREAT): linking telemedicine with diabetes self-management education to improve care in rural communities. Diabetes Educ 2014 Sep;40(6):797-805. [doi: 10.1177/0145721714551993] [Medline: 25253624]
210. Skolarus LE, Piette JD, Pfeiffer PN, Williams LS, Mackey J, Hughes R, et al. Interactive Voice Response-An Innovative Approach to Post-Stroke Depression Self-Management Support. Transl Stroke Res 2017 Feb 9;8(1):77-82 [FREE Full text] [doi: 10.1007/s12975-016-0481-7] [Medline: 27394917]
211. Smallwood AJ, Schapira MM, Fedders M, Neuner JM. A pilot randomized controlled trial of a decision aid with tailored fracture risk tool delivered via a patient portal. Osteoporos Int 2017 Feb;28(2):567-576. [doi: 10.1007/s00198-016-3767-4] [Medline: 27647529]
212. Smania M. Use of a Point-of-Care Tool to Improve Nurse Practitioner BRCA Knowledge. CJON 2016 Jun 1;20(3):327-331. [doi: 10.1188/16.cjon.327-331]
213. Smith KJ, Kuo S, Zgibor JC, McTigue KM, Hess R, Bhargava T, et al. Cost effectiveness of an internet-delivered lifestyle intervention in primary care patients with high cardiovascular risk. Prev Med 2016 Jun;87:103-109. [doi: 10.1016/j.ypmed.2016.02.036] [Medline: 26921656]
214. Smith NM, Satyshur RD. Pediatric Diabetes Telemedicine Program Improves Access to Care for Rural Families: Role of APRNs. Pediatr Nurs 2016;42(6):294-299. [Medline: 29406653]
215. Sobota AE, Kavanagh PL, Adams WG, McClure E, Farrell D, Sprinz PG. Improvement in influenza vaccination rates in a pediatric sickle cell disease clinic. Pediatr Blood Cancer 2015 Apr 24;62(4):654-657 [FREE Full text] [doi: 10.1002/pbc.25390] [Medline: 25545967]
216. Sood A, Watts SA, Johnson JK, Hirth S, Aron DC. Telemedicine consultation for patients with diabetes mellitus: a cluster randomised controlled trial. J Telemed Telecare 2018 Jul;24(6):385-391. [doi: 10.1177/1357633X17704346] [Medline: 28406066]
217. Sperl-Hillen J, Crain A, Margolis K, Ekstrom H, Appana D, Amundson G, et al. Clinical decision support directed to primary care patients and providers reduces cardiovascular risk: a randomized trial. J Am Med Inform Assoc 2018 Sep 01;25(9):1137-1146 [FREE Full text] [doi: 10.1093/jamia/ocy085] [Medline: 29982627]
218. Stading JA, Phan L, Walter A, Bilslend L, White R, Qi Y. Initial Experience of Clinical Pharmacy Services Delivered by Computer Communication via Cisco Jabber Video in a US Veterans Administration Medical Center. J Pharm Technol 2014 Jun 01;30(3):76-80 [FREE Full text] [doi: 10.1177/8755122513505744] [Medline: 34860870]
219. Steinberg DM, Levine EL, Lane I, Askew S, Foley PB, Puleo E, et al. Adherence to Self-Monitoring via Interactive Voice Response Technology in an eHealth Intervention Targeting Weight Gain Prevention Among Black Women: Randomized Controlled Trial. J Med Internet Res 2014 Apr 29;16(4):e114 [FREE Full text] [doi: 10.2196/jmir.2996] [Medline: 24780934]
220. Stevens MB, Hastings SN, Powers J, Vandenberg AE, Echt KV, Bryan WE, et al. Enhancing the Quality of Prescribing Practices for Older Veterans Discharged from the Emergency Department (EQUiPPED): Preliminary Results from Enhancing Quality of Prescribing Practices for Older Veterans Discharged from the Emergency Department, a Novel Multicomponent Interdisciplinary Quality Improvement Initiative. J Am Geriatr Soc 2015 May;63(5):1025-1029. [doi: 10.1111/jgs.13404] [Medline: 25945692]
221. Stockwell MS, Marchant CD, Wodi AP, Barnett ED, Broder KR, Jakob K, et al. A multi-site feasibility study to assess fever and wheezing in children after influenza vaccines using text messaging. Vaccine 2017 Dec 15;35(50):6941-6948 [FREE Full text] [doi: 10.1016/j.vaccine.2017.10.073] [Medline: 29089191]
222. Stulberg DB, Dahlquist IH, Disterhoft J, Bello JK, Hunter MS. Increase in Contraceptive Counseling by Primary Care Clinicians After Implementation of One Key Question® at an Urban Community Health Center. Matern Child Health J 2019 Aug;23(8):996-1002. [doi: 10.1007/s10995-019-02754-z] [Medline: 31203521]
223. Sudhanthar S, Lapinski J, Turner J, Gold J, Sigal Y, Thakur K, et al. Improving oral health through dental fluoride varnish application in a primary care paediatric practice. BMJ Open Qual 2019 May 31;8(2):e000589 [FREE Full text] [doi: 10.1136/bmjoq-2018-000589] [Medline: 31259286]
224. Szilagyi PG, Serwint JR, Humiston SG, Rand CM, Schaffer S, Vincelli P, et al. Effect of provider prompts on adolescent immunization rates: a randomized trial. Acad Pediatr 2015 Mar;15(2):149-157 [FREE Full text] [doi: 10.1016/j.acap.2014.10.006] [Medline: 25748976]
225. Talmi A, Bunik M, Asherin R, Rannie M, Watlington T, Beaty B, et al. Improving developmental screening documentation and referral completion. Pediatrics 2014 Oct 01;134(4):e1181-e1188. [doi: 10.1542/peds.2012-1151] [Medline: 25180272]
226. Taveras EM, Marshall R, Sharifi M, Avalon E, Fiechtner L, Horan C, et al. Comparative Effectiveness of Clinical-Community Childhood Obesity Interventions: A Randomized Clinical Trial. JAMA Pediatr 2017 Aug 07;171(8):e171325 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.1325] [Medline: 28586856]
227. Taylor AM, Bingham J, Schussel K, Axon DR, Dickman DJ, Boesen K, et al. Integrating Innovative Telehealth Solutions into an Interprofessional Team-Delivered Chronic Care Management Pilot Program. J Manag Care Spec Pharm 2018 Aug;24(8):813-818. [doi: 10.18553/jmcp.2018.24.8.813] [Medline: 30058982]
228. Tetrault J, Holt S, Cavallo D, O'Connor P, Gordon M, Corvino J, et al. Computerized Cognitive Behavioral Therapy for Substance Use Disorders in a Specialized Primary Care Practice: a Randomized Feasibility Trial to Address the RT Component of SBIRT. Journal of Addiction Medicine 2020;14(6):e303-e309. [doi: 10.1097/adm.0000000000000663]

229. Tetuan C, Axon DR, Bingham J, Boesen K, Lipsy R, Scovis N, et al. Assessing the Effect of a Telepharmacist's Recommendations During an Integrated, Interprofessional Telehealth Appointment and Their Alignment with Quality Measures. J Manag Care Spec Pharm 2019 Dec;25(12):1334-1339. [doi: 10.18553/jmcp.2019.25.12.1334] [Medline: 31778622]
230. Thomas KE, Kisely S, Urrego F. Electronic Heath Record Prompts May Increase Screening for Secondhand Smoke Exposure. Clin Pediatr (Phila) 2018 Jan 30;57(1):27-30. [doi: 10.1177/0009922816688261] [Medline: 28135880]
231. Tieu C, Chaudhry R, Schroeder DR, Bock FA, Hanson GJ, Tung EE. Utilization of Patient Electronic Messaging to Promote Advance Care Planning in the Primary Care Setting. Am J Hosp Palliat Care 2017 Aug 17;34(7):665-670. [doi: 10.1177/1049909116650237] [Medline: 27188759]
232. Tsoh JY, Quach T, Duong TB, Sa Nan Park E, Wong C, Huang SM, et al. Interactive Mobile Doctor (iMD) to Promote Patient-Provider Discussion on Tobacco Use among Asian American Patients in Primary Care: A Pilot Study. J Am Board Fam Med 2018 Nov 09;31(6):869-880 [FREE Full text] [doi: 10.3122/jabfm.2018.06.180018] [Medline: 30413543]
233. Turner JC, Keller A, Wu H, Zimmerman M, Zhang J, Barnes LE. Utilization of primary care among college students with mental health disorders. Health Psychol 2018 Apr;37(4):385-393. [doi: 10.1037/hea0000580] [Medline: 29376665]
234. Turvey CL, Klein DM, Witry M, Klutts JS, Hill EL, Alexander B, et al. Patient Education for Consumer-Mediated HIE. A Pilot Randomized Controlled Trial of the Department of Veterans Affairs Blue Button. Appl Clin Inform 2016 Aug 03;7(3):765-776 [FREE Full text] [doi: 10.4338/ACI-2016-01-RA-0014] [Medline: 27484821]
235. Wagholikar KB, Hankey RA, Decker LK, Cha SS, Greenes RA, Liu H, et al. Evaluation of the effect of decision support on the efficiency of primary care providers in the outpatient practice. J Prim Care Community Health 2015 Jan;6(1):54-60 [FREE Full text] [doi: 10.1177/2150131914546325] [Medline: 25155103]
236. Walsh KE, Bacic J, Phillips BD, Adams WG. Misuse of Pediatric Medications and Parent-Physician Communication: An Interactive Voice Response Intervention. J Patient Saf 2017 Feb 22;17(3):e207-e213. [doi: 10.1097/pts.0000000000000375]
237. Werk LN, Diaz MC, Cadilla A, Franciosi JP, Hossain MJ. Promoting Adherence to Influenza Vaccination Recommendations in Pediatric Practice. J Prim Care Community Health 2019 Jun 11;10:2150132719853061 [FREE Full text] [doi: 10.1177/2150132719853061] [Medline: 31184255]
238. Wise K, Sebastian S, Haas-Gehres A, Moore-Clingenpeel M, Lamberjack K. Pharmacist impact on pediatric vaccination errors and missed opportunities in the setting of clinical decision support. J Am Pharm Assoc (2003) 2017;57(3):356-361 [FREE Full text] [doi: 10.1016/j.japh.2017.02.025] [Medline: 28427898]
239. Wolin KY, Steinberg DM, Lane IB, Askew S, Greaney ML, Colditz GA, et al. Engagement with eHealth Self-Monitoring in a Primary Care-Based Weight Management Intervention. PLoS One 2015 Oct 15;10(10):e0140455 [FREE Full text] [doi: 10.1371/journal.pone.0140455] [Medline: 26469065]
240. Wolver SE, Stultz JS, Aggarwal A, Thacker L, Banas C. Provider and Patient Perceptions of an External Medication History Function. J Patient Saf 2018 Dec;14(4):234-240. [doi: 10.1097/PTS.0000000000000197] [Medline: 26101998]
241. Woo C, Seton JM, Washington M, Tomlinson SC, Phrasavath D, Farrell KR, et al. Increasing specialty care access through use of an innovative home telehealth-based spinal cord injury disease management protocol (SCI DMP). The Journal of Spinal Cord Medicine 2014 Mar 12;39(1):3-12. [doi: 10.1179/2045772314y.0000000202]
242. Wood SM, McGeary A, Wilson M, Taylor A, Aumaier B, Petsis D, et al. Effectiveness of a Quality Improvement Intervention to Improve Rates of Routine Chlamydia Trachomatis Screening in Female Adolescents Seeking Primary Preventive Care. J Pediatr Adolesc Gynecol 2019 Feb;32(1):32-38 [FREE Full text] [doi: 10.1016/j.jpag.2018.10.004] [Medline: 30394335]
243. Wu RR, Myers RA, Sperber N, Voils CI, Neuner J, McCarty CA, et al. Implementation, adoption, and utility of family health history risk assessment in diverse care settings: evaluating implementation processes and impact with an implementation framework. Genet Med 2019 Feb 6;21(2):331-338 [FREE Full text] [doi: 10.1038/s41436-018-0049-x] [Medline: 29875427]
244. Yadav S, Hartkop S, Cardenas PY, Ladkany R, Halalau A, Shoichet S, et al. Utilization of a breast cancer risk assessment tool by internal medicine residents in a primary care clinic: impact of an educational program. BMC Cancer 2019 Mar 14;19(1):228 [FREE Full text] [doi: 10.1186/s12885-019-5418-6] [Medline: 30871497]
245. Yakovchenko V, Hogan TP, Houston TK, Richardson L, Lipschitz J, Petrakis BA, et al. Automated Text Messaging With Patients in Department of Veterans Affairs Specialty Clinics: Cluster Randomized Trial. J Med Internet Res 2019 Aug 04;21(8):e14750 [FREE Full text] [doi: 10.2196/14750] [Medline: 31444872]
246. Yeboah-Korang A, Beig M, Khan M, Goldstein J, Macapinlac D, Maurer D, et al. Hepatitis C Screening in Commercially Insured U.S. Birth-cohort Patients: Factors Associated with Testing and Effect of an EMR-based Screening Alert. J Transl Int Med 2018 Jun;6(2):82-89 [FREE Full text] [doi: 10.2478/jtim-2018-0012] [Medline: 29984203]
247. Yeung A, Martinson MA, Baer L, Chen J, Clain A, Williams A, et al. The Effectiveness of Telepsychiatry-Based Culturally Sensitive Collaborative Treatment for Depressed Chinese American Immigrants. J. Clin. Psychiatry 2016 Aug 24;77(08):e996-e1002. [doi: 10.4088/jcp.15m09952]
248. Young HM, Miyamoto S, Dharmar M, Tang-Feldman Y. Nurse Coaching and Mobile Health Compared With Usual Care to Improve Diabetes Self-Efficacy for Persons With Type 2 Diabetes: Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e16665 [FREE Full text] [doi: 10.2196/16665] [Medline: 32130184]

249. Yu JS, Szigethy E, Wallace M, Solano F, Oser M. Implementation of a Guided, Digital Cognitive Behavioral Program for Anxiety in Primary Care: Preliminary Findings of Engagement and Effectiveness. Telemed J E Health 2018 Nov 26;24(11):870-878. [doi: 10.1089/tmj.2017.0280] [Medline: 29480752]
250. Zera CA, Bates DW, Stuebe AM, Ecker JL, Seely EW. Diabetes Screening Reminder for Women With Prior Gestational Diabetes. Obstetrics & Gynecology 2015;126(1):109-114. [doi: 10.1097/aog.0000000000000883]
251. Zhao L, Stinnett SS, Prakalapakorn SG. Visual Acuity Assessment and Vision Screening Using a Novel Smartphone Application. J Pediatr 2019 Oct;213:203-210.e1. [doi: 10.1016/j.jpeds.2019.06.021] [Medline: 31326117]
252. Zhong X, Liang M, Sanchez R, Yu M, Budd PR, Sprague JL, et al. On the effect of electronic patient portal on primary care utilization and appointment adherence. BMC Med Inform Decis Mak 2018 Oct 16;18(1):84 [FREE Full text] [doi: 10.1186/s12911-018-0669-8] [Medline: 30326876]
253. Zieve GG, Richardson LP, Katzman K, Spielvogle H, Whitehouse S, McCarty CA. Adolescents' Perspectives on Personalized E-Feedback in the Context of Health Risk Behavior Screening for Primary Care: Qualitative Study. J Med Internet Res 2017 Jul 20;19(7):e261 [FREE Full text] [doi: 10.2196/jmir.7474] [Medline: 28729236]
254. Zimet G, Dixon BE, Xiao S, Tu W, Kulkarni A, Dugan T, et al. Simple and Elaborated Clinician Reminder Prompts for Human Papillomavirus Vaccination: A Randomized Clinical Trial. Acad Pediatr 2018 Mar;18(2S):S66-S71. [doi: 10.1016/j.acap.2017.11.002] [Medline: 29502640]
255. Zimmerman RK, Brown AE, Pavlik VN, Moehling KK, Raviotta JM, Lin CJ, et al. Using the 4 Pillars Practice Transformation Program to Increase Pneumococcal Immunizations for Older Adults: A Cluster-Randomized Trial. J Am Geriatr Soc 2017 Jan;65(1):114-122 [FREE Full text] [doi: 10.1111/jgs.14451] [Medline: 27755655]
256. Zimmerman RK, Moehling KK, Lin CJ, Zhang S, Raviotta JM, Reis EC, et al. Improving adolescent HPV vaccination in a randomized controlled cluster trial using the 4 Pillars™ practice Transformation Program. Vaccine 2017 Jan 03;35(1):109-117 [FREE Full text] [doi: 10.1016/j.vaccine.2016.11.018] [Medline: 27876200]
257. Zimmerman RK, Raviotta JM, Nowalk MP, Moehling KK, Reis EC, Humiston SG, et al. Using the 4 Pillars™ Practice Transformation Program to increase adolescent human papillomavirus, meningococcal, tetanus-diphtheria-pertussis and influenza vaccination. Vaccine 2017 Oct 27;35(45):6180-6186 [FREE Full text] [doi: 10.1016/j.vaccine.2017.09.039] [Medline: 28947346]
258. Daré LO, Bruand P, Gérard D, Marin B, Lameyre V, Boumédiène F, et al. Co-morbidities of mental disorders and chronic physical diseases in developing and emerging countries: a meta-analysis. BMC Public Health 2019 Mar 13;19(1):304 [FREE Full text] [doi: 10.1186/s12889-019-6623-6] [Medline: 30866883]
259. Casey MM, Moscovice I, McCullough J. Rural primary care practices and meaningful use of electronic health records: the role of Regional Extension Centers. J Rural Health 2014;30(3):244-251. [doi: 10.1111/jrh.12050] [Medline: 24118180]
260. Cross DA, Cohen GR, Nong P, Day A, Vibbert D, Naraharisetti R, et al. Improving EHR Capabilities to Facilitate Stage 3 Meaningful Use Care Coordination Criteria. AMIA Annu Symp Proc 2015;2015:448-455 [FREE Full text] [Medline: 26958177]
261. Shi L, Starfield B. Primary Care, Income Inequality, and Self-Rated Health in the United States: A Mixed-Level Analysis. Int J Health Serv 2016 Jun 22;30(3):541-555. [doi: 10.2190/n4m8-303m-72ua-p1k1]
262. Krist AH, Aycock RA, Etz RS, Devoe JE, Sabo RT, Williams R, et al. MyPreventiveCare: implementation and dissemination of an interactive preventive health record in three practice-based research networks serving disadvantaged patients--a randomized cluster trial. Implement Sci 2014 Dec 11;9:181 [FREE Full text] [doi: 10.1186/s13012-014-0181-1] [Medline: 25500097]
263. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, et al. Mobile App-Based Health Promotion Programs: A Systematic Review of the Literature. Int J Environ Res Public Health 2018 Dec 13;15(12) [FREE Full text] [doi: 10.3390/ijerph15122838] [Medline: 30551555]
264. Prunuske J, Chang L, Mishori R, Dobbie A, Morley CP. The extent and methods of public health instruction in family medicine clerkships. Fam Med 2014;46(7):544-548 [FREE Full text] [Medline: 25058549]
265. Institute of Medicine, Board on Population Health and Public Health Practice, Committee on Integrating Primary Care and Public Health. Primary Care and Public Health: Exploring Integration to Improve Population Health. Washington, DC: The National Academies Press; 2012.
266. Bazemore AW, Makaroff LA, Puffer JC, Parhat P, Phillips RL, Xierali IM, et al. Declining numbers of family physicians are caring for children. J Am Board Fam Med 2012;25(2):139-140 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110203] [Medline: 22403191]
267. Carek PJ. Potentially Alarming Trends in the Scope of Practice for Family Physicians. J Am Board Fam Med 2018;31(2):178-180 [FREE Full text] [doi: 10.3122/jabfm.2018.02.180022] [Medline: 29535232]
268. Coutinho AJ, Cochrane A, Stelter K, Phillips RL, Peterson LE. Comparison of Intended Scope of Practice for Family Medicine Residents With Reported Scope of Practice Among Practicing Family Physicians. JAMA 2015 Dec 08;314(22):2364-2372. [doi: 10.1001/jama.2015.13734] [Medline: 26647258]
269. Jetty A, Jabbarpour Y, Petterson S, Eden A, Bazemore A. The Declining Presence of Family Physicians in Hospital-Based Care. J Am Board Fam Med 2019 Nov 08;32(6):771-772 [FREE Full text] [doi: 10.3122/jabfm.2019.06.190152] [Medline: 31704744]

270. Tong STC, Makaroff LA, Xierali IM, Parhat P, Puffer JC, Newton WP, et al. Proportion of family physicians providing maternity care continues to decline. J Am Board Fam Med 2012 May 08;25(3):270-271 [FREE Full text] [doi: 10.3122/jabfm.2012.03.110256] [Medline: 22570387]
271. Xierali IM, Puffer JC, Tong STC, Bazemore AW, Green LA. The percentage of family physicians attending to women's gender-specific health needs is declining. J Am Board Fam Med 2012 Jul 05;25(4):406-407 [FREE Full text] [doi: 10.3122/jabfm.2012.04.110290] [Medline: 22773707]
272. Digital Health Center of Excellence. FDA. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence [accessed 2022-01-03]
273. Furuya-Kanamori L, Xu C, Hasan SS, Doi SA. Quality versus Risk-of-Bias assessment in clinical research. J Clin Epidemiol 2021 Jan 13;129(2):172-175 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.09.044] [Medline: 33422267]
274. Stephens GG. The Personal Touch: The Physician As Healer In Modern Society. J Kans Med Soc 1965 May;66:237-239. [Medline: 14288399]

Abbreviations

AI: artificial intelligence
DHI: digital health intervention
EHR: electronic health record
FDA: Food and Drug Administration
HCV: hepatitis C virus
HIE: health information exchange
HIT: health information technology
mHealth: mobile health
RCT: randomized controlled trial
ROI: return on investment
WHO: World Health Organization

Edited by C Lovis; submitted 10.09.21; peer-reviewed by K Lin, L Maaß; comments to author 24.09.21; revised version received 19.11.21; accepted 04.12.21; published 21.01.22.

Please cite as:
Willis VC, Thomas Craig KJ, Jabbarpour Y, Scheufele EL, Arriaga YE, Ajinkya M, Rhee KB, Bazemore A
Digital Health Interventions to Enhance Prevention in Primary Care: Scoping Review
JMIR Med Inform 2022;10(1):e33518
URL: https://medinform.jmir.org/2022/1/e33518
doi:10.2196/33518
PMID:35060909

Viewpoint

Impact of Electronic Health Record Interoperability on Telehealth Service Outcomes

Xinyue Zhang[1], MPA; Richard Saltman[1], PhD

Department of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Xinyue Zhang, MPA
Department of Health Policy and Management
Rollins School of Public Health
Emory University
1518 Clifton Rd, 6th Floor
Atlanta, GA, 30322
United States
Phone: 1 2068494755
Email: xinyue.zhang2@emory.edu

Abstract

This paper aims to develop a telehealth success model and discusses three critical components: (1) health information quality, (2) electronic health record system quality, and (3) telehealth service quality to ensure effective telehealth service delivery, reduce professional burnout, and enhance access to care. The paper applied a policy analysis method and discussed telehealth applications in rural health, mental health, and veterans health services. The results pointed out the fact that, although telehealth paired with semantic/organizational interoperability facilitates value-based and team-based care, challenges remain to enhance user (both patients and clinicians) experience and satisfaction. The conclusion indicates that approaches at systemic and physician levels are needed to reduce disparities in health technology adoption and improve access to telehealth care.

(JMIR Med Inform 2022;10(1):e31837) doi:10.2196/31837

KEYWORDS

Electronic Health Records; Telehealth; Telemental health; Pandemic; Health outcomes; Health Policy

Executive Summary

A telehealth platform integrated with an interoperable electronic health record (EHR) system can contribute directly toward achieving the often-discussed "quadruple aim" [1]—better health outcomes, improved patient experience, lower costs, and improved clinician experience. This paper develops a telehealth success model and discusses three critical components: (1) health information quality, (2) EHR system quality, and (3) telehealth service quality to ensure effective health care service delivery, reduce professional burnout, and enhance access to care.

Despite the benefits of telehealth in rural health, mental health, and Veterans Administration health services, disparities continue to exist in access to care. Patients without internet service, appropriate devices, or digital literacy skills experience greater challenges in accessing care via telehealth. The COVID-19 pandemic has also caused substantial financial strain on hospitals, and it is uncertain if the current reimbursement and payment model for telehealth service/devices and regulation flexibility for virtual consulting will continue after the pandemic is over.

To help integrate telehealth into clinical practice and improve patient care, health policy at the systemic level should accelerate the uptake of telehealth. On the industry level, hospitals should identify adoption strategies for different types of telehealth services and evaluate telehealth products for health care delivery. On the physician level, health providers should offer the same level of care and follow the same treatment guidelines for telehealth services as with in-person visits and ensure that their practices are compliant with applicable regulations.

Development of Telehealth

Telehealth has become a rapidly growing sector of health care delivery systems. Previous studies show evidence that telehealth tools and services increase the overall effectiveness of physicians in (1) counseling patients with chronic conditions, (2) psychotherapy support for behavioral interventions, and (3) remote monitoring of patients [2].

The shortage of health providers and increasing consumer demand (from the aging population and people diagnosed with chronic diseases) were key factors in expanding the scope and scale of telehealth services [3]. The 2019 annual report of the Association of American Medical Colleges projected a shortfall of 40,000 to 122,000 physicians in the United States over the next decade, with a shortage of 29,000 to 42,900 doctors in 2020 [4]. To use telehealth as a new strategy to stretch the physician supply, the Interstate Medical Licensure Compact standardized licensing requirements that allow physicians to practice in multiple states and provide remote digitalized services [5]. To further remove regulatory and reimbursement barriers to telehealth services, telehealth parity laws require commercial health insurers to provide equal coverage for telehealth and in-person services in 38 states and the District of Columbia [6]. By January 2017, all state Medicaid programs reimburse teleradiology, 49 cover tele–mental health services, and 36 states cover various remote telehealth services [2].

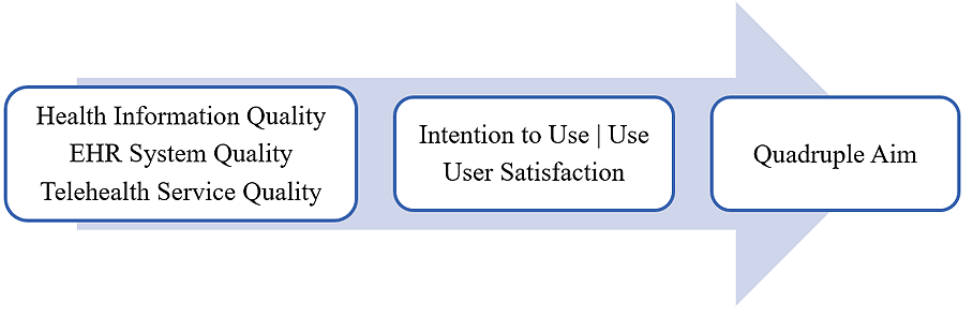
The COVID-19 pandemic accelerated the adoption of telehealth tools and services. Amid the pandemic, some hospitals are seeing 500 to 600 patients per day via video or telephone visits [7]. To enable providers to use telehealth services, Medicare implemented temporary payment flexibility to allow more beneficiaries to benefit from virtual care services and more providers to be eligible to bill for telehealth services at the same payment rate as they would receive for in-person services [8]. The Centers for Medicare and Medicaid Services (CMS) added 135 allowable services, including emergency department visits; initial nursing facility and discharge visits; home visits; and physical, occupational, and speech therapy services [9]. With these initiatives, US provider systems are rapidly deploying digitalized services for two main goals:

1. Forward triage to screen patients with COVID-19 symptoms before arrival to a health care facility so as to reduce exposure to the virus [10]
2. Continue patient care and provide virtual consultation to nonvirus patients, especially those with chronic diseases

Telehealth Success Model

This paper applies a conceptual model (Figure 1) based on Delone and McLean's [11] model of information systems success to assess the impact of hospital medical record interoperability on telehealth service outcomes. As framed by Delone and McLean [11], a sustainable information system depends on positive results from the quality of information, service, and systems, as well as interrelated measures of user satisfaction, use, and net benefits.

Figure 1. Telehealth success model. EHR: electronic health record.



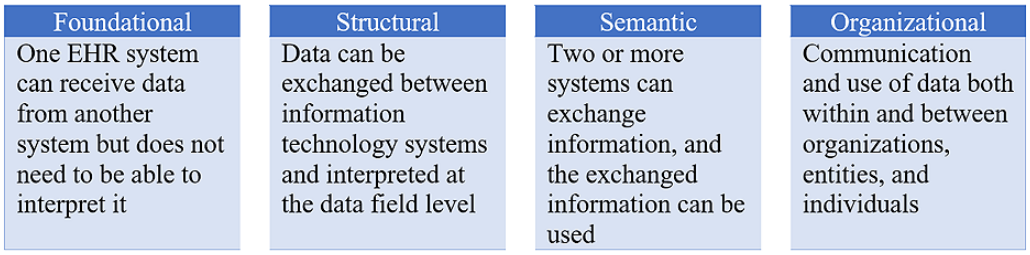
Health Information Quality

Device interoperability and data integration are key aspects of telehealth delivery. Hospital interoperability (Figure 2) covers three types of information exchange: (1) sending, receiving, and incorporating health records that support electronic referral loops; (2) electronic access for both physicians and patients to their health information; and (3) public health surveillance that collects and integrates health-related data to assist planning, implementing, and evaluating public health practice [12].

For value-based care to achieve and protect patient safety, the EHR system needs to deliver accurate and clinically appropriate data across care settings to both physicians and patients. Studies suggest that hospital sharing of diagnostic data with providers within their system is associated with lower patient mortality, and the hospital interoperability level is associated with improved process quality related to conditions of acute myocardial infarction, heart failure, and pneumonia at acute care hospitals [13-15].

However, establishing a telehealth platform in a short period of time amid pandemic conditions could put health information—patient names, address, dates, diagnoses, and more—at higher risks to safety and security [16]. The introduction of increasingly complicated technology into already complex work environments may trigger various unintended interactions that undermine or outweigh the potential benefits of the new technology [17]. Moreover, with an exponential growth in clinical data, it becomes critical to code symptoms (eg, allergies) and medications correctly to ensure patient safety and care quality [18].

Figure 2. Hospital interoperability levels. EHR: electronic health record.



EHR System Quality: Maximizing the Benefits of Telehealth

Integrating telehealth programs into a hospital's existing EHR system infrastructure helps maximize the benefits of telemedicine, as providers and staff already have experience working with the baseline system [19]. Over 95% of US hospitals reported using a certified EHR platform [12]. However, many hospitals run separate systems for doctors, labs, radiologists, and remote monitoring devices; the technical and data incompatibilities between different vendors make data sharing more vulnerable to cybersecurity threats [20]. Suboptimally integrated systems also added clerical burdens on physicians that can help lead to professional burnout. For every hour of clinical work, physicians spent 2 hours on EHR-related tasks, threatening the capacity and performance of the health system [21].

In 2020, the US Office of the National Coordinator for Health Information Technology (ONC) established requirements for a secure standards-based application programming interface (API) to support each individual patient's access and control of their electronic health information [22,23]. The increasing data volumes, new data types, and various data sources collected from telehealth services can make it difficult and labor-intensive to match or identify the correct patient between systems [24]. Thus, a wide-scale adoption of common standards would drive data sharing and make integration more consistent and efficient, thereby providing clinically useful information and mitigating physician burnout [25].

Telehealth Service Quality

Advanced interoperability—especially at semantic and organizational levels—can enable telehealth to expand access, exchange information, and provide user-centered services to both physicians and patients [26]. Current telehealth services use devices such as wearable monitors, smartphones, mobile apps, video, email, and web portals to deliver three types of care: (1) remote monitoring of patients and collecting vital signs and health data for care plan management, (2) counseling and interacting with patients at home, and (3) triaging patients to screen them to reduce exposures to viruses and thereby free up hospital resources during emergencies [27].

Telehealth Use in Clinical Practice

Rural Health Services

Approximately 80% of rural areas in the United States are classified as medically underserved and in health professional shortage areas [28,29]. These regions are lacking the physicians, registered nurses, and behavioral health providers (including psychiatrists, psychologists, and therapists) [30]. The patient-to-primary care physician ratio in rural areas is only 39.8 physicians per 100,000 people, compared to 53.3 in urban areas [31]. The shortage disproportionately impacts rural residents who tend to be older, have lower socioeconomic status, are more reliant on public insurance, and have worse health outcomes [32,33].

For rural residents, telehealth care increased access to experienced providers and high-quality care [34,35], improved continuity of care and health outcomes [36], and reduced health disparities [37]. Studies have shown that greater adoption of telehealth was associated with facilities in rural locations [38]. Between 2010 and 2017, telehealth visits have increased among rural Medicare beneficiaries, with a 425% increase for tele–mental health services [36].

Tele–Mental Health Services

There are two primary uses for tele–mental health: provider consultations with mental health specialists in primary care and emergency department settings, and the direct provision of mental health services including home-based services [37,38].

For people experiencing serious mental illness, telehealth has the potential to improve quality of life and general mental health, reduce depressive symptoms, build more confidence in managing depression, and increase satisfaction with mental health and coping skills compared to treatment offered in-person only [36,39]. For people experiencing substance use disorders (SUDs), treatments delivered through telehealth have resulted in reductions in alcohol consumption, increased tobacco cessation, and increased engagement and retention in opioid use disorder treatment [36]. Between 2016 and 2019, SUD treatment offered through telehealth increased from 13.5% to 17.4% [40].

Veterans' Health Services

As one of the early adopters of telehealth, the Veterans Health Administration (VHA) is currently the largest telehealth provider in the United States [41]. In 2018, VHA conducted over a million telehealth visits [42]. A total 10% of the visits used VA

Video Connect (VVC), a secure video teleconferencing platform that allows providers to treat veterans on their mobile devices or personal computers at a location of the veteran's choice [43,44]. The 2018 "Maintaining Internal Systems and Strengthening Integrated Outside Networks" ("MISSION") Act included mandates for VHA to establish an "Anywhere to Anywhere" telehealth network, where VHA providers in outpatient mental health and primary care service lines nationwide will be both capable and experienced with providing telehealth (VVC) to non-VA locations [43]. The demands for VA telehealth services also increased during the pandemic [45]. Tele–mental health sessions via VVC increased 42% at one VA medical center in South Carolina, from 1429 appointments in January 2020 to 2034 in March 2020 [43,45].

Patient-Centered Care

A telehealth program paired with the right EHR system can serve as a care collaboration platform and help optimize team-based care delivery [13,46]. It connects off-site specialists in the fields of cardiology, psychiatry and behavioral health, oncology, and infectious disease with patients at home or intensive care units (ICUs) [2]. Physicians can easily access and send health records from one interface to another (mobile, computer, or tablet) remotely using a protected account to diagnose and assess symptoms as an in-person consultation [47].

Telehealth interventions, particularly remote monitoring and SMS text messaging, were associated improvements in obstetric outcomes, perinatal smoking cessation, breastfeeding, and schedule optimization for high-risk obstetrics [48]. For at-risk patients with chronic disease, remote monitoring devices continuously capture physiological data such as heart rate, blood glucose, oxygen saturation level, body temperature, blood pressure, and weight over time [2,49]. Evidence also suggests telehealth services allowed physicians to better communicate with patients on treatment plans that are appropriate for their culture, race, gender, sexual orientation, and lived experience [35,48,50].

Population Health Management

An EHR system with organizational interoperability allows institutions to aggregate community-level data from disparate sources to track influenza/disease trends for population health [51,52]. For example, CMS requires hospitals to electronically report public health data such as syndromic surveillance data, electronic case reporting, reportable laboratory results, and more. Sharing critical data among health care systems, especially during a pandemic, assists public health authorities to predict clusters of outbreaks and make timely and efficient guidance for quarantine and better containment [53,54].

Postpandemic Health Care Needs and Challenges

Regulation Uncertainties

The COVID-19 pandemic in the United States is affecting different areas at different times and levels: cases spike in some states while others face the threats of both COVID-19 resurgence and seasonal flu [55]. Many hospitals are providing a combination of traditional in-person visits and telehealth services that allow remote virtual consultations to patients [34].

Early in March 2020, CMS modified policies to lift originating site restrictions and expand the type of visits allowed virtually [56]. However, after the pandemic, hospitals may shift telehealth services from urgent care and COVID-19 screenings and treatment to regular care visits [57]; it is unclear how current more flexible regulation and payment arrangements will change [56,58]. Will there be permanent polices to reimburse virtual care and remote monitoring devices? Will telehealth reimbursement rates be set at the same level with in-person visits? Will there be financial incentives to provide reliable broadband access to rural or small hospitals [59]?

With hospitals and the health care system faced with uncertainties about the duration of this pandemic and the structure of future telehealth benefits, the development of clear regulatory requirements and timetables could help reduce administrative and technological constraints associated with virtual health care delivery and encourage further investment in health care information technology (IT) infrastructure [60,61].

Hospital Financing Challenges

The COVID-19 pandemic has created substantial financial difficulties for both hospitals and the health system [62,63]. As a result of cancelled elective surgeries and nonessential medical procedures, which often generate more revenues than ICU and emergency care, US hospitals continue to experience substantial losses in revenue [64]. Expenses also have increased sharply from purchasing needed for personal protective equipment, COVID-19–associated hospitalizations, and providing additional support to frontline health workers [62]. The American Hospital Association estimates a total 4-month financial impact of US $202.6 billion in losses for US hospitals and health systems, or an average of US $50.7 billion per month [62]. These financial loss and additional system maintenance/implementation costs for telehealth and EHR systems will require decision makers to establish more effective strategies to use hospital resources and workforce [64,65].

Disparities in Telehealth Access

Although most hospitals in the United States have adopted interoperable EHR systems, there is little evidence about whether small, rural, and safety net hospitals are keeping up [66]. Compared to more technologically advanced hospitals, smaller and rural hospitals have limited broadband access [59], less interoperability and health care IT management experience [67], and staff with less technological familiarity [68,69]. Because of the uneven adoption of telemedicine services, some small clinics and postacute care facilities are unable to receive or share patient data [70,71].

Substantial disparities in access to telehealth services also remain [72]. Evidence suggests geographical disparities, profit-based discrimination, technology deployment cost, and socioeconomic factors played key roles in the telehealth use gap [59,72,73]. Moreover, people 65 years and older, with disabilities, experiencing poverty, and who are non-White are

less likely to use telehealth services because of lower smartphone or computer ownership, limited (home) broadband internet access, and low digital literacy [36]. A recent study of 134,225 completed primary care visits also reported that early adopters of online scheduling were more likely to be young, White, and commercially insured [74].

Regulatory Process for EHR Market

To ensure that EHRs cooperate effectively in an interoperable structure, substantial governmental regulation has been put into place. In the United States, CMS and the ONC regulate EHR privacy, security, and standards for hospitals or health providers and health IT developers [75,76]. In July 2021, the Interoperability and Patient Access final rule began to require CMS-regulated payers to remove industry siloes and support Patient Access API, Provider Directory API, and Payer-to-Payer Data Exchange to achieve greater semantic interoperability within the health care system while complying with existing Health Insurance Portability and Accountability Act (HIPAA) requirements [75,77]. On the technical level, CMS adopted Health Level 7 Fast Healthcare Interoperability Resources Release 4.0.1 to standardize implementing privacy and security features for provider organizations [78].

In the European Union, to facilitate cross-border EHR interoperability, the General Data Protection Regulation established explicit rules to process and protect patient health data [79]. On the technical level, the eHealth Digital Service Infrastructure has enabled provider organizations to exchange patient summaries and e-prescriptions [78]. In the Asia-Pacific region, Singapore, Japan, and Australia have instituted regulations on software qualification, software as a medical device, and presubmission consultation by regulatory authorities to facilitate EHR interoperability [80-82]. In Singapore, for example, the National Electronic Health Record system sets technical standards (including architecture, security, and operations) for the digital health market and monitors user functionality and risk to protect data security [83].

Discussion

Studies suggest that telehealth programs paired with the right EHR system enhance care access, increase patient satisfaction, and reduce medical spending [84,85], and by improving clinician experience, the integrated system can contribute to achieving the quadruple aim.

To help integrate telehealth into clinical practice and improve patient care, on a systemic level, health policies should accelerate the uptake of telehealth, including tele–mental health, to improve care quality, cost-effectiveness, and value of care [86]. Federal and state governments can use disruptive reimbursement and funding strategies on training primary and mental health care providers, workforce, licensure, and cultural sensitivity for long-term telehealth practice [34,87]. Regulators also need to assess and set standards for malpractice liability and protect patient safety and confidentiality that may result from telehealth deployment [88].

On an industry/organization level, hospitals need to identify strategies to adopt and integrate different types of telehealth services, and evaluate telehealth products for health care delivery [66,70]. Future studies are needed to provide evidence on telehealth practice guidelines and service models.

On the physician level, clinicians who provide telehealth should offer the same level of care and follow the same treatment guidelines they would follow for in-person visits [89]. Moreover, physicians should closely follow HIPAA rules, state laws, and medical board definitions to ensure their practices are compliant with applicable regulations while implementing telehealth [57].

Conflicts of Interest

None declared.

References

1. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
2. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]
3. Dall TM, Gallo PD, Chakrabarti R, West T, Semilla AP, Storm MV. An aging population and growing disease burden will require a large and specialized health care workforce by 2025. Health Aff (Millwood) 2013 Nov;32(11):2013-2020. [doi: 10.1377/hlthaff.2013.0714] [Medline: 24191094]
4. The complexities of physician supply and demand: projections from 2018 to 2033. Association of American Medical Colleges. URL: https://www.aamc.org/system/files/2020-06/stratcomm-aamc-physician-workforce-projections-june-2020.pdf [accessed 2021-02-17]
5. Frieden J. COVID-19 is making the physician shortage worse, groups say. Medpage Today. URL: https://www.medpagetoday.com/infectiousdisease/covid19/85661 [accessed 2021-01-15]
6. Abuzeineh M, Muzaale AD, Crews DC, Avery RK, Brotman DJ, Brennan DC, et al. Telemedicine in the care of kidney transplant recipients with coronavirus disease 2019: case reports. Transplant Proc 2020 Nov;52(9):2620-2625 [FREE Full text] [doi: 10.1016/j.transproceed.2020.07.009] [Medline: 32798002]

7. Redford G. Delivering more care remotely will be critical as COVID-19 races through communities. Association of American Medical Colleges. 2020 Mar 23. URL: https://www.aamc.org/news-insights/delivering-more-care-remotely-will-be-critical-covid-19-races-through-communities [accessed 2020-03-28]
8. Physicians and other clinicians: CMS flexibilities to fight COVID-19. Centers for Medicare and Medicaid Services. 2020 Mar 30. URL: https://www.endocrine.org/-/media/endocrine/files/membership/cms--covid-phys-fact-sheet.pdf [accessed 2020-05-20]
9. Verma S. Early impact of CMS expansion of Medicare telehealth during COVID-19. Health Affairs. 2020 Jul 15. URL: https://www.healthaffairs.org/do/10.1377/forefront.20200715.454789/full/ [accessed 2020-08-12]
10. COVID-19: a physician guide to keeping your practice open. American Medical Association. 2020 Dec. URL: https://www.ama-assn.org/delivering-care/public-health/covid-19-physician-guide-keeping-your-practice-open [accessed 2021-02-16]
11. DeLone H, McLean E. The DeLone and McLean model of information systems success: a ten-year update. J Manage Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
12. What is HIE? HealthIT.gov. 2020 Jul 24. URL: https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/what-hie [accessed 2020-08-19]
13. Deyo D, Khaliq A, Mitchell D, Hughes DR. Electronic sharing of diagnostic information and patient outcomes. Am J Manag Care 2018 Jan;24(1):32-37 [FREE Full text] [Medline: 29350510]
14. Appari A, Eric Johnson M, Anthony DL. Meaningful use of electronic health record systems and process quality of care: evidence from a panel data analysis of U.S. acute-care hospitals. Health Serv Res 2013 Apr;48(2 Pt 1):354-375 [FREE Full text] [doi: 10.1111/j.1475-6773.2012.01448.x] [Medline: 22816527]
15. Lammers EJ, McLaughlin CG. Meaningful use of electronic health records and Medicare expenditures: evidence from a panel data analysis of U.S. health care markets, 2010-2013. Health Serv Res 2017 Aug;52(4):1364-1386 [FREE Full text] [doi: 10.1111/1475-6773.12550] [Medline: 27546309]
16. Ozair FF, Jamshed N, Sharma A, Aggarwal P. Ethical issues in electronic health records: a general overview. Perspect Clin Res 2015;6(2):73-76 [FREE Full text] [doi: 10.4103/2229-3485.153997] [Medline: 25878950]
17. Greenhalgh T, Potts HWW, Wong G, Bark P, Swinglehurst D. Tensions and paradoxes in electronic patient record research: a systematic literature review using the meta-narrative method. Milbank Q 2009 Dec;87(4):729-788 [FREE Full text] [doi: 10.1111/j.1468-0009.2009.00578.x] [Medline: 20021585]
18. Computerized Prescriber Order Entry Medication Safety (CPOEMS): uncovering and learning from issues and errors. US Food and Drug Administration. 2015 Dec 15. URL: https://www.fda.gov/media/95234/download [accessed 2020-09-05]
19. Smith WR, Atala AJ, Terlecki RP, Kelly EE, Matthews CA. Implementation guide for rapid integration of an outpatient telemedicine program during the COVID-19 pandemic. J Am Coll Surg 2020 Aug;231(2):216-222.e2 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2020.04.030] [Medline: 32360960]
20. Williams P, Woodward A. Cybersecurity vulnerabilities in medical devices: a complex environment and multifaceted problem. Med Devices (Auckl) 2015;8:305-316. [doi: 10.2147/MDER.S50048] [Medline: 26229513]
21. Patel R, Bachu R, Adikey A, Malik M, Shah M. Factors related to physician burnout and its consequences: a review. Behav Sci (Basel) 2018 Oct 25;8(11):98 [FREE Full text] [doi: 10.3390/bs8110098] [Medline: 30366419]
22. HHS finalizes historic rules to provide patients more control of their health data. Centers for Medicare and Medicaid Services. 2020 Mar 09. URL: https://www.cms.gov/newsroom/press-releases/hhs-finalizes-historic-rules-provide-patients-more-control-their-health-data [accessed 2020-07-12]
23. Lipinski M, Knee M. 21st Century Cures Act: interoperability, information blocking, and the ONC Health IT Certification Program proposed rule. HealthIT.gov. 2019 Mar 22. URL: https://www.healthit.gov/sites/default/files/facas/2019-03-22_ONC_Cures_Act_NPRM_IB_%20HITAC_IB_Task_Force_508.pdf [accessed 2020-06-15]
24. Luxton DD, Kayl RA, Mishkind MC. mHealth data security: the need for HIPAA-compliant standardization. Telemed J E Health 2012 May;18(4):284-288. [doi: 10.1089/tmj.2011.0180] [Medline: 22400974]
25. Chervitz S, Deutsch EW, Field D, Parkinson H, Quackenbush J, Rocca-Serra P, et al. Data standards for Omics data: the basis of data sharing and reuse. Methods Mol Biol 2011;719:31-69 [FREE Full text] [doi: 10.1007/978-1-61779-027-0_2] [Medline: 21370078]
26. Interoperability in Healthcare. Healthcare Information and Management Systems Society. URL: https://www.himss.org/resources/interoperability-healthcare [accessed 2020-05-17]
27. Coronavirus (COVID-19) update: FDA allows expanded use of devices to monitor patients' vital signs remotely. US Food and Drug Administration. 2020 Mar 20. URL: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-allows-expanded-use-devices-monitor-patients-vital-signs-remotely [accessed 2020-08-12]
28. What is shortage designation? Health Resources and Services Administration: Bureau of Health Workforce. 2021 Feb. URL: https://bhw.hrsa.gov/workforce-shortage-areas/shortage-designation [accessed 2021-04-12]
29. Shortage areas. data.HRSA.gov. URL: https://data.hrsa.gov/topics/health-workforce/shortage-areas [accessed 2021-02-18]
30. The psychiatric shortage: causes and solutions. The National Council for Mental Wellbeing. 2018 Mar 01. URL: https://www.thenationalcouncil.org/wp-content/uploads/2017/03/Psychiatric-Shortage_National-Council-.pdf?daf=375ateTbd56 [accessed 2020-05-17]

31. About rural health care. National Rural Health Association. URL: https://www.ruralhealth.us/about-nrha/about-rural-health-care [accessed 2021-06-01]
32. Medicaid and rural health. Medicaid and CHIP Payment and Access Commission. 2021 Apr. URL: https://www.macpac.gov/wp-content/uploads/2021/04/Medicaid-and-Rural-Health.pdf [accessed 2021-05-01]
33. Mueller JT, McConnell K, Burow PB, Pofahl K, Merdjanoff AA, Farrell J. Impacts of the COVID-19 pandemic on rural America. Proc Natl Acad Sci U S A 2021 Jan 05;118(1):2019378118 [FREE Full text] [doi: 10.1073/pnas.2019378118] [Medline: 33328335]
34. CDC. Using telehealth to expand access to essential health services during the COVID-19 pandemic. Centers for Disease Control and Prevention. 2020 Jun 10. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/telehealth.html [accessed 2020-09-16]
35. Totten AM, Womack DM, Eden KB, McDonagh MS, Griffin JC, Grusing S, et al. Telehealth: Mapping the Evidence for Patient Outcomes From Systematic Reviews. Rockville, MD: Agency for Healthcare Research and Quality; 2016 Jun. URL: https://www.ncbi.nlm.nih.gov/books/NBK379320/
36. Telehealth for the treatment of serious mental illness and substance use disorders. Substance Abuse and Mental Health Services Administration. URL: https://store.samhsa.gov/sites/default/files/SAMHSA_Digital_Download/PEP21-06-02-001.pdf [accessed 2021-06-15]
37. Patel SY, Huskamp HA, Busch AB, Mehrotra A. Telemental health and US rural-urban differences in specialty mental health use, 2010-2017. Am J Public Health 2020 Sep;110(9):1308-1314. [doi: 10.2105/AJPH.2020.305657] [Medline: 32673109]
38. Myers CR. Using telehealth to remediate rural mental health and healthcare disparities. Issues Ment Health Nurs 2019 Mar;40(3):233-239. [doi: 10.1080/01612840.2018.1499157] [Medline: 30508400]
39. Dixon LB, Holoshitz Y, Nossel I. Treatment engagement of individuals experiencing mental illness: review and update. World Psychiatry 2016 Mar;15(1):13-20. [doi: 10.1002/wps.20306] [Medline: 26833597]
40. Uscher-Pines L, Cantor J, Huskamp HA, Mehrotra A, Busch A, Barnett M. Adoption of telemedicine services by substance abuse treatment facilities in the U.S. J Subst Abuse Treat 2020 Oct;117:108060 [FREE Full text] [doi: 10.1016/j.jsat.2020.108060] [Medline: 32811631]
41. Elliott VL. Telehealth and telemedicine: description and issues. United States Senate. 2016 Mar 29. URL: https://www.senate.gov/CRSpubs/757e3b90-ff10-497c-8e8c-ac1bdbdb3aaf.pdf [accessed 2020-09-21]
42. Spotlight on telehealth. US Department of Veterans Affairs: Health Services Research and Development. 2020 Jul. URL: https://www.hsrd.research.va.gov/news/feature/telehealth-0720.cfm [accessed 2020-08-12]
43. Myers US, Birks A, Grubaugh AL, Axon RN. Flattening the curve by getting ahead of it: how the VA healthcare system is leveraging telehealth to provide continued access to care for rural veterans. J Rural Health 2021 Jan;37(1):194-196 [FREE Full text] [doi: 10.1111/jrh.12449] [Medline: 32282955]
44. Department of Veterans Affairs (VA): a primer on telehealth. Federation of American Scientists: Project on Government Secrecy. 2019 Jul 26. URL: https://sgp.fas.org/crs/misc/R45834.pdf [accessed 2020-08-15]
45. VA Video Connect visits increase 1000% during COVID-19 pandemic. US Department of Veterans Affairs. URL: https://www.va.gov/opa/pressrel/pressrelease.cfm?id=5467 [accessed 2021-04-18]
46. Adler-Milstein J, Everson J, Lee SD. Sequencing of EHR adoption among US hospitals and the impact of meaningful use. J Am Med Inform Assoc 2014;21(6):984-991 [FREE Full text] [doi: 10.1136/amiajnl-2014-002708] [Medline: 24853066]
47. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
48. Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K, et al. Personalized telehealth in the future: a global research agenda. J Med Internet Res 2016 Mar 01;18(3):e53 [FREE Full text] [doi: 10.2196/jmir.5257] [Medline: 26932229]
49. Tomasic I, Tomasic N, Trobec R, Krpan M, Kelava T. Continuous remote monitoring of COPD patients-justification and explanation of the requirements and a survey of the available technologies. Med Biol Eng Comput 2018 Apr;56(4):547-569 [FREE Full text] [doi: 10.1007/s11517-018-1798-z] [Medline: 29504070]
50. Sharing data, saving lives. American Hospital Association. 2019 Jan. URL: https://www.aha.org/system/files/2019-01/Infographic01_18__19_Sharing-Data-Saving-Lives_infographic_FINAL.pdf [accessed 2020-08-16]
51. Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. Health Aff (Millwood) 2009;28(2):454-466. [doi: 10.1377/hlthaff.28.2.454] [Medline: 19276005]
52. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
53. McClellan M, Gottlieb S, Mostashari F, Rivers C, Silvis L. A national COVID-19 surveillance system: achieving containment. Margolis Center for Health Policy. 2020 Apr 07. URL: https://healthpolicy.duke.edu/sites/default/files/2020-06/a_national_covid_surveillance_system.pdf [accessed 2020-07-21]
54. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]

55. Maragakis L. Coronavirus second wave, third wave and beyond: what causes a COVID surge. Johns Hopkins Medicine. URL: https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus [accessed 2021-04-12]
56. Medicare coverage and payment of virtual services. Centers for Medicare and Medicaid Services. 2020 May 08. URL: https://www.cms.gov/outreach-and-educationmedicare-learning-network-mlnmlnproductsmln-multimedia/medicare-coverage-and-payment-virtual-services [accessed 2020-07-12]
57. Shachar C, Engel J, Elwyn G. Implications for telehealth in a postpandemic future: regulatory and privacy issues. JAMA 2020 Jun 16;323(23):2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]
58. COVID-19 frequently asked questions (FAQs) on Medicare Fee-for-Service (FFS) Billing. Centers for Medicare and Medicaid Services. URL: https://www.cms.gov/files/document/03092020-covid-19-faqs-508.pdf [accessed 2021-11-18]
59. Perzynski A, Roach MJ, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. J Am Med Inform Assoc 2017 Sep 01;24(5):927-932 [FREE Full text] [doi: 10.1093/jamia/ocx020] [Medline: 28371853]
60. Strategy on reducing burden relating to the use of health IT and EHRs. HealthIT.gov. 2020 Mar 11. URL: https://www.healthit.gov/topic/usability-and-provider-burden/strategy-reducing-burden-relating-use-health-it-and-ehrs [accessed 2020-08-15]
61. COVID-19 emergency declaration blanket waivers for health care providers. Centers for Medicare and Medicaid Services. 2021 May 24. URL: https://www.cms.gov/files/document/summary-covid-19-emergency-declaration-waivers.pdf [accessed 2021-06-12]
62. Hospitals and health systems face unprecedented financial pressures due to COVID-19. American Hospital Association. 2020 May. URL: https://www.aha.org/system/files/media/file/2020/05/aha-covid19-financial-impact-0520-FINAL.pdf [accessed 2020-06-01]
63. Bojdani E, Rajagopalan A, Chen A, Gearin P, Olcott W, Shankar V, et al. COVID-19 pandemic: impact on psychiatric care in the United States. Psychiatry Res 2020 May 06;289:113069. [doi: 10.1016/j.psychres.2020.113069] [Medline: 33242812]
64. Khullar D, Bond AM, Schpero WL. COVID-19 and the financial health of US hospitals. JAMA 2020 Jun 02;323(21):2127-2128. [doi: 10.1001/jama.2020.6269] [Medline: 32364565]
65. Druss BG, Cohen AN, Brister T, Cotes RO, Hendry P, Rolin D, et al. Supporting the mental health workforce during and after COVID-19. Psychiatr Serv 2021 Oct 01;72(10):1222-1224. [doi: 10.1176/appi.ps.202000509] [Medline: 33882690]
66. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148 [FREE Full text] [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
67. Cohen MF. Impact of the HITECH financial incentives on EHR adoption in small, physician-owned practices. Int J Med Inform 2016 Oct;94:143-154. [doi: 10.1016/j.ijmedinf.2016.06.017] [Medline: 27573322]
68. Garrett P, Brown CA, Hart-Hester S, Hamadain E, Dixon C, Pierce W, et al. Identifying barriers to the adoption of new technology in rural hospitals: a case report. Perspect Health Inf Manag 2006 Oct 25;3:9 [FREE Full text] [Medline: 18066367]
69. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. P T 2017 Sep;42(9):572-575 [FREE Full text] [Medline: 28890644]
70. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic health record adoption in US hospitals: progress continues, but challenges persist. Health Aff (Millwood) 2015 Dec;34(12):2174-2180. [doi: 10.1377/hlthaff.2015.0992] [Medline: 26561387]
71. Pylypchuk Y, Alvarado CS, Patel V, Searcy T. Uncovering differences in interoperability across hospital size. Healthc (Amst) 2019 Dec;7(4):S2213-0764(18)30185-4. [doi: 10.1016/j.hjdsi.2019.04.001] [Medline: 31003837]
72. Mack D, Zhang S, Douglas M, Sow C, Strothers H, Rust G. Disparities in primary care EHR adoption rates. J Health Care Poor Underserved 2016 Mar;27(1):327-338 [FREE Full text] [doi: 10.1353/hpu.2016.0016] [Medline: 27587942]
73. Heisey-Grove DM, Carretta HJ. Disparities in secure messaging uptake between patients and physicians: longitudinal analysis of two national cross-sectional surveys. J Med Internet Res 2020 May 01;22(5):e12611 [FREE Full text] [doi: 10.2196/12611] [Medline: 32356775]
74. Ganguli I, Orav EJ, Lupo C, Metlay JP, Sequist TD. Patient and visit characteristics associated with use of direct scheduling in primary care practices. JAMA Netw Open 2020 Aug 03;3(8):e209637 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.9637] [Medline: 32852551]
75. Policies and technology for interoperability and burden reduction. Centers for Medicare and Medicaid Services. URL: https://www.cms.gov/Regulations-and-Guidance/Guidance/Interoperability/index [accessed 2021-11-15]
76. McCue M. Financial performance of health plans in Medicaid managed care. Medicare Medicaid Res Rev 2012;2(2):mmrr.002.02.a07 [FREE Full text] [doi: 10.5600/mmrr.002.02.a07] [Medline: 24800142]
77. CMS. Interoperability and patient access fact sheet. Centers for Medicare and Medicaid Services. 2020 Mar 09. URL: https://www.cms.gov/newsroom/fact-sheets/interoperability-and-patient-access-fact-sheet [accessed 2020-06-12]
78. Exchange of electronic health records across the EU. European Commission: Shaping Europe's Digital Future. URL: https://digital-strategy.ec.europa.eu/en/policies/electronic-health-records [accessed 2021-11-12]
79. What is GDPR, the EU's new data protection law? General Data Protection Regulation. URL: https://gdpr.eu/what-is-gdpr/ [accessed 2021-11-12]

80. Teo P, Yeoh BS, Ong SN. SARS in Singapore: surveillance strategies in a globalising city. Health Policy 2005 Jun;72(3):279-291 [FREE Full text] [doi: 10.1016/j.healthpol.2004.11.004] [Medline: 15862636]
81. Consultation on Regulatory Guidelines for Classification of Standalone Medical Mobile Applications (SaMD) and Qualification of Clinical Decision Support Software. Singapore Health Sciences Authority. URL: https://tinyurl.com/4rztun6j [accessed 2021-11-12]
82. Digital health regulation in Asia-Pacific: overview and best practices. APACMed. 2021 Jan. URL: https://apacmed.org/content/uploads/2021/01/APACMed-Digital-Health-Regulation-in-APAC.pdf [accessed 2021-02-15]
83. International health care system profiles. The Commonwealth Fund. 2020. URL: https://www.commonwealthfund.org/international-health-policy-center/system-profiles [accessed 2021-01-09]
84. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does not decrease spending. Health Aff (Millwood) 2017 Mar 01;36(3):485-491. [doi: 10.1377/hlthaff.2016.1130] [Medline: 28264950]
85. McLendon S. Interactive video telehealth models to improve access to diabetes specialty care and education in the rural setting: a systematic review. Diabetes Spectr 2017 May;30(2):124-136 [FREE Full text] [doi: 10.2337/ds16-0004] [Medline: 28588379]
86. Telepsychiatry and COVID-19: update on telehealth restrictions in response to COVID-19. American Psychiatric Association. 2020 May 01. URL: https://www.psychiatry.org/psychiatrists/practice/telepsychiatry/blog/apa-resources-on-telepsychiatry-and-covid-19 [accessed 2020-06-12]
87. List of telehealth services. Centers for Medicare and Medicaid Services. URL: https://www.cms.gov/Medicare/Medicare-General-Information/Telehealth/Telehealth-Codes [accessed 2021-05-16]
88. Fahey RA, Hino A. COVID-19, digital privacy, and the social limits on data-focused public health responses. Int J Inf Manage 2020 Dec;55:102181 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102181] [Medline: 32836638]
89. Lammers EJ, McLaughlin CG, Barna M. Physician EHR adoption and potentially preventable hospital admissions among Medicare beneficiaries: panel data evidence, 2010-2013. Health Serv Res 2016 Dec;51(6):2056-2075 [FREE Full text] [doi: 10.1111/1475-6773.12586] [Medline: 27766628]

Abbreviations

API: application programming interface
CMS: Centers for Medicare and Medicaid Services
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
ICU: intensive care unit
IT: information technology
MISSION: Maintaining Internal Systems and Strengthening Integrated Outside Networks
ONC: Office of the National Coordinator for Health Information Technology
SUD: substance use disorder
VHA: Veterans Health Administration
VVC: VA Video Connect

Edited by C Lovis; submitted 06.07.21; peer-reviewed by A Hidki; comments to author 03.10.21; revised version received 20.10.21; accepted 14.11.21; published 11.01.22.

Please cite as:
Zhang X, Saltman R
Impact of Electronic Health Record Interoperability on Telehealth Service Outcomes
JMIR Med Inform 2022;10(1):e31837
URL: https://medinform.jmir.org/2022/1/e31837
doi:10.2196/31837
PMID:34890347

Viewpoint

Technology-Enabled, Evidence-Driven, and Patient-Centered: The Way Forward for Regulating Software as a Medical Device

Jane Elizabeth Carolan[1,2,3], BMedSc, MPH; John McGonigle[3], PhD; Andrea Dennis[3], PhD; Paula Lorgelly[2], PhD; Amitava Banerjee[1,4,5], MA, MPH, DPhil

[1]Institute of Health Informatics, University College London, London, United Kingdom
[2]Institute of Epidemiology and Health Care, University College London, London, United Kingdom
[3]Perspectum Ltd, Oxford, United Kingdom
[4]University College London Hospitals NHS Foundation Trust, London, United Kingdom
[5]Barts Health NHS Trust, London, United Kingdom

Corresponding Author:
Jane Elizabeth Carolan, BMedSc, MPH
Institute of Health Informatics
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 07464345635
Email: j.carolan@ucl.ac.uk

Abstract

Artificial intelligence (AI) is a broad discipline that aims to understand and design systems that display properties of intelligence. Machine learning (ML) is a subset of AI that describes how algorithms and models can assist computer systems in progressively improving their performance. In health care, an increasingly common application of AI/ML is software as a medical device (SaMD), which has the intention to diagnose, treat, cure, mitigate, or prevent disease. AI/ML includes either "locked" or "continuous learning" algorithms. Locked algorithms consistently provide the same output for a particular input. Conversely, continuous learning algorithms, in their infancy in terms of SaMD, modify in real-time based on incoming real-world data, without controlled software version releases. This continuous learning has the potential to better handle local population characteristics, but with the risk of reinforcing existing structural biases. Continuous learning algorithms pose the greatest regulatory complexity, requiring seemingly continuous oversight in the form of special controls to ensure ongoing safety and effectiveness. We describe the challenges of continuous learning algorithms, then highlight the new evidence standards and frameworks under development, and discuss the need for stakeholder engagement. The paper concludes with 2 key steps that regulators need to address in order to optimize and realize the benefits of SaMD: first, international standards and guiding principles addressing the uniqueness of SaMD with a continuous learning algorithm are required and second, throughout the product life cycle and appropriate to the SaMD risk classification, there needs to be continuous communication between regulators, developers, and SaMD end users to ensure vigilance and an accurate understanding of the technology.

(JMIR Med Inform 2022;10(1):e34038) doi:10.2196/34038

KEYWORDS

Artificial intelligence; machine learning; algorithm; software; risk assessment; informatics

Introduction

Artificial intelligence (AI) is a broad discipline that aims to understand and design systems that display properties of intelligence [1]. Machine learning (ML) is a subset of AI that describes how algorithms and models can assist computer systems in progressively improving their performance [2]. Based on publicly available information, in late September 2021, the US Food and Drug Administration (FDA) listed (noting "initial list" only) 343 AI/ML-enabled medical devices marketed in the United States. In health care, an increasingly common application of AI and ML is software as a medical device (SaMD), which has the intention to diagnose, treat, cure, mitigate, or prevent disease [3]. Regulatory frameworks for SaMD need to be adaptive while prioritizing patient safety and effectiveness [4-6]. Regulatory challenges of SaMD include

processing submitted evidence to verify clinical effectiveness, generalizability, interoperability, data integrity, and data security. Constructing a fit-for-purpose regulatory framework for SaMD with a continuous learning algorithm is an added complexity. As regulatory agencies aim to advance health care delivery through SaMD adoption, with efforts to avoid unintended consequences, this commentary summarizes the current regulatory frameworks for SaMD. First, we describe the challenges of continuous learning algorithms, then highlight the new evidence standards and frameworks under development, and discuss the need for stakeholder engagement, concluding with 2 key steps that regulators need to address in order to optimize and realize the many benefits of SaMD.

Technology-Enabled Algorithms

ML techniques incorporate training, validation, and test data sets at different stages of model development. Algorithms are executed in a training data set and results compared with a target value. Parameters of the model are adjusted accordingly as part of this process. Identifying potential data biases (including age, ethnicity, vendor, disease prevalence) is critical, but not limited to this point. At the validation stage, the fitted model is used to predict responses for observations in the validation data set, a process of fine-tuning the model. In the test stage, the ML model is exposed to a test data set, independent of training or validation data sets, providing unbiased evaluation of the final model. AI/ML includes either "locked" or "continuous learning" algorithms. Locked algorithms consistently provide the same output for a particular input. Such algorithms may be modified to optimize performance, requiring "episodic" regulatory review if the algorithm requires additional inputs or changes in intended use or performance. Continuous learning algorithms, in their infancy in terms of SaMD, modify in real-time based on incoming real-world data, without controlled software version releases. Continuous learning algorithms pose the greatest regulatory complexity, requiring seemingly continuous oversight in the form of special controls to ensure ongoing safety and effectiveness.

Although systems with continuous learning may appear conceptually similar to systems that self-calibrate to the local environment (eg, adapting to temperature), continuous learning algorithms using modern ML techniques are qualitatively different in that portions of their algorithms, in the form of their trained networks, are being modified autonomously. This continuous learning has the potential to better handle local population characteristics, but with the risk of reinforcing existing structural biases, potentially without adequate oversight. Thus, special regulations are needed to classify these risks and accordingly, ensure appropriate human oversight.

Frameworks and Standards for the Future

Medical device regulatory agencies such as the US FDA, EU Notified Bodies, and the UK Medicines and Healthcare products Regulatory Agency (MHRA) have responsibility for protecting public health by only enabling market access for safe and effective products. Further down the line, importantly, health care budget holders then need to assess cost-effectiveness and budget impact, a potential rate-limiting step for successful market access. Lessons on successful AI/ML adoption in other industries are limited in their value given the unique health risks and benefits that health care regulators must assess. To verify claims of safety and effectiveness in the form of submitted evidence, regulators must keep pace with the complexity of algorithm models, including validation and testing stages, selected use of software of unknown pedigree, and real-world performance [7].

The FDA has outlined its proposed framework for SaMD in a total product life cycle approach [4] and released an AI/ML-based SaMD action plan [8] in response to stakeholder feedback. At the premarket submission stage, a predetermined change control plan would play a role in obtaining reasonable assurance of safety and effectiveness: developers would stipulate what anticipated algorithm modifications would occur, and how the algorithm would learn and change without compromising safety or performance. Postmarket access, periodic updates to the FDA on changes to the algorithm to enable ongoing oversight of real-world performance would be provided. Early next year, draft guidance on detailed requirements is anticipated; currently, it is not evident how much oversight should be performed by the end user(s) and manufacturer, nor how much robust data are needed to substantiate safety and effectiveness claims.

To promote rigor and transparency in design and reporting of AI-based interventions (underpinning regulatory submission evidence claims), reporting guidelines and checklists include Consolidated Standards of Reporting Trials–Artificial Intelligence (CONSORT-AI), Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence (SPIRIT-AI), The Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis-Artificial Intelligence (TRIPOD-AI), and Minimum Information About Clinical Artificial Intelligence Modeling (MI-CLAIM) [9,10]. In the UK, the National Institute for Health and Care Excellence (NICE) has also released revised evidence standards for digital health technologies [11]. Currently, there is an absence of tailored frameworks for AI/ML-based SaMD with a continuous learning algorithm; guidelines including MI-CLAIM and NICE's evidence standards framework, while valuable for locked algorithms, note that continuous learning algorithms are beyond their scope.

Globally, the International Medical Device Regulators Federation (IMDRF) aims to accelerate medical device international regulatory harmonization and has drafted key SaMD policies to complement existing international standards, particularly in terms of risk classification, converging terminology, a risk-based framework, and quality management systems. The Institute of Electrical and Electronic Engineers (IEEE) has Artificial Intelligence Medical Device Working Groups on terminology and recommended practice for the quality management of data sets. United Nations agency collaboration between the World Health Organization and the International Telecommunication Union: Focus Group on Artificial Intelligence for Health (FG-AI4H) was established to use AI to advance health care for all, and to benchmark AI

models using secure and confidential, globally representative data sets [12].

The Need for Stakeholder Engagement

It is recognized that patient-centered data and engagement play a fundamental role in regulatory assessment of SaMD. The "patient-centered" approach referred to by the FDA addresses usability, equity, trust, and accountability. Engagement with both developers and end users occurred at a February 2020 Public Workshop on the Evolving Role of Artificial Intelligence in Radiological Imaging. At the latter event, The American College of Radiology (ACR) and Radiological Society of North America (RSNA) questioned [13] the ability of the FDA to ensure safety and effectiveness of continuous learning algorithms, without direct physician or expert oversight during each use. Familiar concerns relate to autonomous image interpretation independent of physician confirmation and oversight. If an algorithm ceases to function properly without radiologist oversight, a significant number of patients are at risk of incorrect screening before algorithm failure is recognized. It was noted that algorithm user manuals must have clear guidance regarding which equipment and protocols are supported, and deployment restricted to those settings studied during validation. Evaluation of real-world algorithm performance will reassure patients and health professionals of readiness for clinical use.

Conclusion

SaMD has great potential to improve health and health care at individual and system levels. To optimize on the benefits associated with SaMD, patient safety and effectiveness need to be aptly assessed for which 2 key steps are necessary. First, international standards and guiding principles addressing the uniqueness of SaMD with a continuous learning algorithm are required [14], outlining best practice oversight and reporting requirements. Aligned regulatory requirements, tailor-made for SaMD with a continuous learning algorithm, are essential, particularly to verify maintenance measures to keep in check modifications throughout the life cycle of SaMD. A special registry dedicated to these technologies may also be appropriate. Depending on the degree of risk to patients from a particular application of AI/ML SaMD, a degree of expert clinical oversight coupled with technology industry/developer assurance is likely to be required. Second, throughout the product life cycle, appropriate to the risk classification of the SaMD product, there needs to be continuous communication between regulators, developers, and SaMD end users to ensure vigilance and an accurate understanding of the technology. The latter will facilitate the adoption of state-of-the-art automation, optimizing clinical effectiveness and ensuring patient safety.

Authors' Contributions

JC was responsible for research concept, literature search, specialist engagement, initial draft, revisions, and final draft; JMG took care of specialist input, review, and editing; AD performed review and editing; PL performed review and editing; and AB was responsible for supervision, revisions, and final draft.

Conflicts of Interest

JC is employed by University College London (UCL) based at Perspectum Ltd through an Innovate UK grant: Knowledge Transfer Partnership (KTP). AB and PL receive research funding from the Perspectum/Innovate UK grant. AD and JMG are employees of Perspectum Ltd.

References

1. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health 2018 Dec;8(2):020303 [FREE Full text] [doi: 10.7189/jogh.08.020303] [Medline: 30405904]
2. Chen S, Banerjee A. Machine learning: a long way from implementation in cardiovascular disease. Heart 2020 Mar;106(5):318-320. [doi: 10.1136/heartjnl-2019-316030] [Medline: 31911505]
3. IMDRF SaMD Working Group. Software as a Medical Device (SaMD): Key Definitions. 2013. URL: http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions.pdf [accessed 2020-09-27]
4. US Food and Drug Administration. Artificial Intelligence and Machine Learning in Software as a Medical Device. White Oak, MD: US Food and Drug Administration; 2019. URL: https://tinyurl.com/49w63ey6 [accessed 2022-01-18]
5. Turpin R, Hoefer E, Lewelling J, Baird P. Machine Learning AI in Medical Devices: Adapting Regulatory Frameworks and Standards to Ensure Safety and Performance. 2020. URL: https://tinyurl.com/49w63ey6 [accessed 2022-01-18]
6. No authors listed. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020 Apr 01;369:m1312 [FREE Full text] [doi: 10.1136/bmj.m1312] [Medline: 32238345]
7. US Food and Drug Administration. Developing a Software Precertification Program: A Working Model. 2019. URL: https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM629276.pdf [accessed 2022-01-14]
8. Center for Devices and Radiological Health. Artificial Intelligence/Machine Learning (AI/ML)- Based Software as a Medical Device (SaMD) Action Plan. 2021 Jan. URL: https://tinyurl.com/49w63ey6 [accessed 2021-01-15]

9. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020 Sep;26(9):1364-1374 [FREE Full text] [doi: 10.1038/s41591-020-1034-x] [Medline: 32908283]
10. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 2020 Sep 09;26(9):1320-1324 [FREE Full text] [doi: 10.1038/s41591-020-1041-y] [Medline: 32908275]
11. Unsworth H, Dillon B, Collinson L, Powell H, Salmon M, Oladapo T, et al. The NICE Evidence Standards Framework for digital health and care technologies - Developing and maintaining an innovative evidence framework with global impact. Digit Health 2021 Jun 24;7:20552076211018617 [FREE Full text] [doi: 10.1177/20552076211018617] [Medline: 34249371]
12. Wiegand T, Krishnamurthy R, Kuglitsch M, Lee N, Pujari S, Salathé M, et al. WHO and ITU establish benchmarking process for artificial intelligence in health. Lancet 2019 Jul 06;394(10192):9-11. [doi: 10.1016/S0140-6736(19)30762-7] [Medline: 30935732]
13. Fleishon HB, Haffy BG. Public Workshop - Evolving Role of Artificial Intelligence in Radiological Imaging, Comments of the American College of Radiology. 2020 Jun 30. URL: https://tinyurl.com/2p8p76ym [accessed 2020-07-18]
14. Gilbert S, Fenech M, Hirsch M, Upadhyay S, Biasiucci A, Starlinger J. Algorithm Change Protocols in the Regulation of Adaptive Machine Learning-Based Medical Devices. J Med Internet Res 2021 Oct 26;23(10):e30545 [FREE Full text] [doi: 10.2196/30545] [Medline: 34697010]

Abbreviations

ACR: American College of Radiology
AI: artificial intelligence
CONSORT-AI: Consolidated Standards of Reporting Trials–Artificial Intelligence
FDA: the US Food and Drug Administration
FG-AI4H: Focus Group on Artificial Intelligence for Health
IEEE: Institute of Electrical and Electronic Engineers
IMDRF: International Medical Device Regulators Federation
MHRA: the UK Medicines and Healthcare products Regulatory Agency
MI-CLAIM: Minimum Information About Clinical Artificial Intelligence Modeling
ML: machine learning
NICE: National Institute for Health and Care Excellence
RSNA: Radiological Society of North America
SaMD: software as a medical device
SPIRIT-AI: Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence
TRIPOD-AI: The Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis-Artificial Intelligence

Edited by C Lovis; submitted 04.10.21; peer-reviewed by M Burns, A Tomar; comments to author 14.11.21; accepted 04.12.21; published 27.01.22.

Please cite as:
Carolan JE, McGonigle J, Dennis A, Lorgelly P, Banerjee A
Technology-Enabled, Evidence-Driven, and Patient-Centered: The Way Forward for Regulating Software as a Medical Device
JMIR Med Inform 2022;10(1):e34038
URL: https://medinform.jmir.org/2022/1/e34038
doi:10.2196/34038
PMID:35084352

Original Paper

Development of a Pipeline for Adverse Drug Reaction Identification in Clinical Notes: Word Embedding Models and String Matching

Klaske R Siegersma[1,2], MSc; Maxime Evers[1], MSc; Sophie H Bots[1], PhD; Floor Groepenhoff[1,3], MD, PhD; Yolande Appelman[2], MD, PhD; Leonard Hofstra[2,4], MD, PhD; Igor I Tulevski[4], MD, PhD; G Aernout Somsen[4], MD, PhD; Hester M den Ruijter[1], PhD; Marco Spruit[5,6*], PhD; N Charlotte Onland-Moret[7*], PhD

[1]Laboratory of Experimental Cardiology, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[2]Department of Cardiology, Amsterdam University Medical Centers, VU University Medical Center, Amsterdam, Netherlands

[3]Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[4]Cardiology Centers of the Netherlands, Utrecht, Netherlands

[5]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden University, Leiden, Netherlands

[6]Leiden Institute of Advanced Computer Science, Leiden University, Leiden, Netherlands

[7]Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[*]these authors contributed equally

Corresponding Author:
N Charlotte Onland-Moret, PhD
Department of Epidemiology
Julius Center for Health Sciences and Primary Care
University Medical Center Utrecht, Utrecht University
Universiteitsweg 100
Utrecht, 3584 CG
Netherlands
Phone: 31 887569610
Email: N.C.Onland@umcutrecht.nl

Abstract

Background: Knowledge about adverse drug reactions (ADRs) in the population is limited because of underreporting, which hampers surveillance and assessment of drug safety. Therefore, gathering accurate information that can be retrieved from clinical notes about the incidence of ADRs is of great relevance. However, manual labeling of these notes is time-consuming, and automatization can improve the use of free-text clinical notes for the identification of ADRs. Furthermore, tools for language processing in languages other than English are not widely available.

Objective: The aim of this study is to design and evaluate a method for automatic extraction of medication and Adverse Drug Reaction Identification in Clinical Notes (ADRIN).

Methods: Dutch free-text clinical notes (N=277,398) and medication registrations (N=499,435) from the Cardiology Centers of the Netherlands database were used. All clinical notes were used to develop word embedding models. Vector representations of word embedding models and string matching with a medical dictionary (Medical Dictionary for Regulatory Activities [MedDRA]) were used for identification of ADRs and medication in a test set of clinical notes that were manually labeled. Several settings, including search area and punctuation, could be adjusted in the prototype to evaluate the optimal version of the prototype.

Results: The ADRIN method was evaluated using a test set of 988 clinical notes written on the stop date of a drug. Multiple versions of the prototype were evaluated for a variety of tasks. Binary classification of ADR presence achieved the highest accuracy of 0.84. Reduced search area and inclusion of punctuation improved performance, whereas incorporation of the MedDRA did not improve the performance of the pipeline.

Conclusions: The ADRIN method and prototype are effective in recognizing ADRs in Dutch clinical notes from cardiac diagnostic screening centers. Surprisingly, incorporation of the MedDRA did not result in improved identification on top of word embedding models. The implementation of the ADRIN tool may help increase the identification of ADRs, resulting in better care and saving substantial health care costs.

(JMIR Med Inform 2022;10(1):e31063) doi:10.2196/31063

KEYWORDS

adverse drug reactions; word embeddings; clinical notes

Introduction

Background

Literature shows that adverse drug events (ADEs) and, more specifically, adverse drug reactions (ADRs) are structurally underreported [1]. Clinical trials may underreport or miss ADRs for various reasons, such as a follow-up that is usually too short to catch long-term effects [2]. In addition, the study population may be healthier or otherwise different from the target population in regular care [3]. As a result, the ADR risk of clinically relevant subgroups such as women and older adults remains unknown [4], which places a societal and economic burden on our health care system. The prevalence of hospital admissions associated with ADRs is reported to be as high as 5.3% and estimated to be twice as high in the older adult population [5]. In the United States alone, ADRs are estimated to generate US $30 billion in unnecessary costs [6]. Efforts have been made to structurally collect information on ADRs both on a national (eg, Lareb in the Netherlands) and international (EudraVigilance [7]) level; however, these pharmacovigilance databases do not include relevant patient characteristics and information about prescription rates.

Regular care data extracted from electronic health records can help in postmarketing surveillance of medication. ADRs are usually not reported in the electronic health record in a structured way, but the clinical notes made during consultations between patients and their physicians may hold relevant information when patients experience an ADR. However, these notes are often stored as free text and thus cannot be easily analyzed [8]. Methods that extract ADRs from these free-text fields are needed to access the full potential of these data.

Natural language processing (NLP) techniques can aid in the differentiation of relevant features from idle free text and prepare free text for research purposes [9,10]. One of the widespread topics in NLP is the use of word embeddings—a vector representation of a text, often established through evaluation of the word's context. The use of word embeddings for the evaluation of clinical free text for research purposes is increasing [11]. Research has shown that training word embedding models on a domain-specific data set generates better results than training on a general data set [12,13]. As a result, applications of word embedding models are studied in a wide range of topics within the health care domain (eg, evaluation of radiology reports [14], identification of ICD-10 codes [15], and identification of ADEs in English electronic health records [16]) and can potentially be a solution to extract ADRs from Dutch clinical notes.

Objectives

The objective of this research is to design a method for the identification of ADRs in clinical notes from a regular care database (Adverse Drug Reaction Identification in Clinical Notes [ADRIN]) using unlabeled data and word embeddings. Although the demonstrations in this study have been done with Dutch clinical notes from the cardiovascular domain, the method has been developed in a way that enables generalization not only to other languages but also to other research questions to mine text in clinical notes.

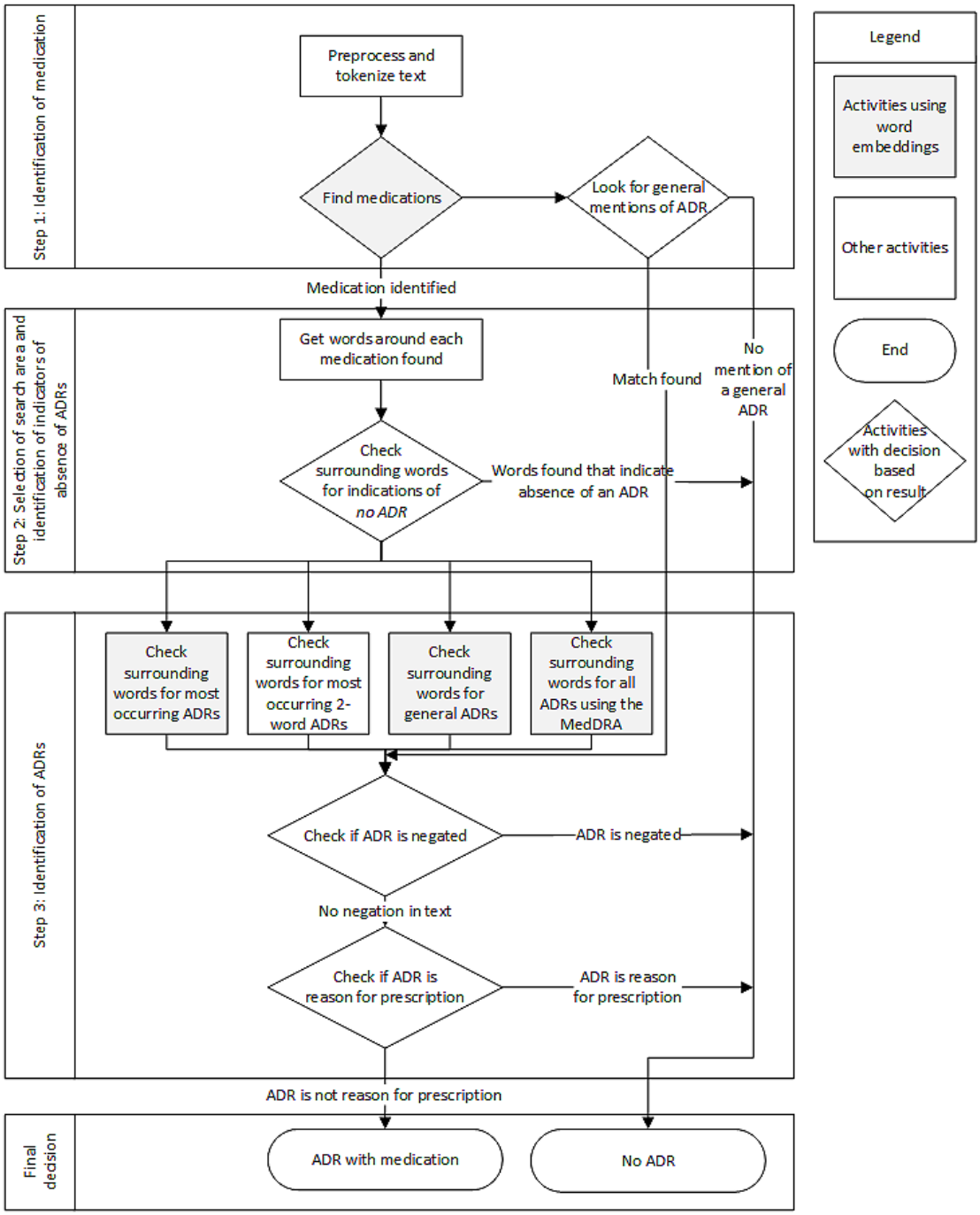
Methods

Overview

The ADRIN method is based on the implementation of a medical taxonomy to enhance standardized terminology (the Medical Dictionary for Regulatory Activities [MedDRA]) [17] and on word embeddings trained on a large database of medical free text. In addition, a prototype was developed and evaluated on labeled Dutch clinical notes to determine the performance of this method. Figure 1 shows the general workflow of the ADRIN method.

This study focused on the identification of ADRs and the corresponding medications. We assumed that patients were compliant with their medication regimen. We defined an ADR as any unwanted event that led to the discontinuation of the prescribed medication. In the following description, clinical notes are defined as the free text written down in the electronic health record by the physician after a patient's consultation.

Figure 1. Overview of the different steps in the Adverse Drug Reaction Identification in Clinical Notes method. ADR: adverse drug reaction; MedDRA: Medical Dictionary for Regulatory Activities.



Data Set

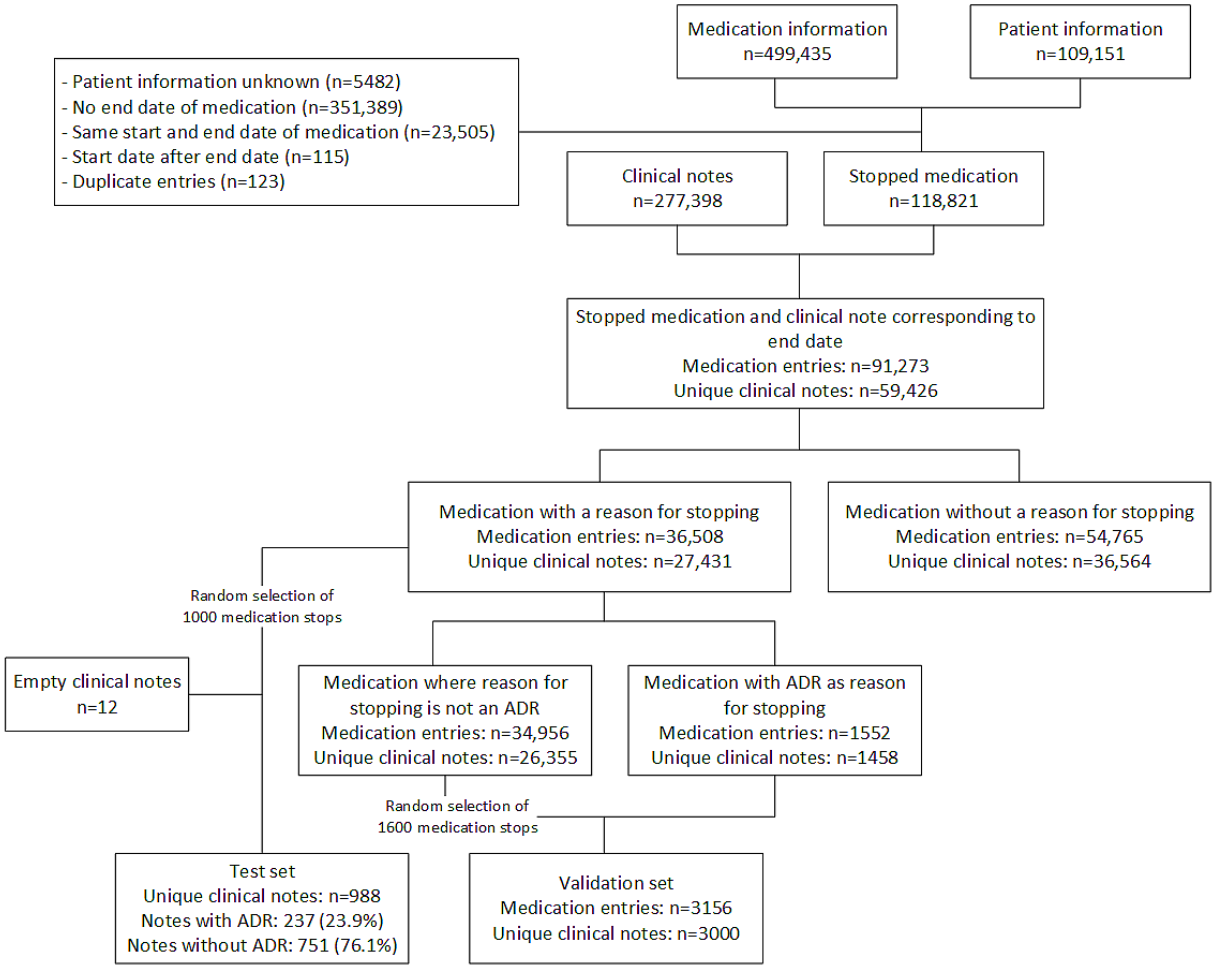
The Cardiology Centers of the Netherlands database is a large regular care database from 13 diagnostic cardiac screening centers. In short, this database consists of 109,151 patients who visited one of the outpatient cardiac screening centers between 2007 and 2018 and includes patient characteristics and information about diagnostic tests [18].

In total, there were 277,398 clinical notes in the database and 499,435 medication prescriptions. Clinical notes were deidentified using DEDUCE [19]. Medication prescriptions contain information about the prescribed medication, start date and end date (if the medication was discontinued at some point), and reason for discontinuation in free text.

Figure 2 describes the selection of discontinued medication entries from the database. The selected prescriptions were merged with the clinical notes. This resulted in 91,273 discontinued medication entries for which a clinical note was available on the end date of the medication. In cases where multiple prescriptions from the same patient were stopped on the same day (19,992/91,273, 21.9%), the same clinical note was used for all prescriptions. The reason for discontinuation

was reported in 40% (36,508/91,273) of the medication prescriptions. From these 91,273 medication entries, we randomly selected 1000 (1.1%) medication entries and corresponding clinical notes as a test set. However, in 1.2% (12/1000) of the cases, the clinical note was empty, resulting in a test set of 988 clinical notes.

Figure 2. Flowchart of selection of clinical notes and corresponding adverse drug reaction and medication. ADR: adverse drug reaction.



The validation set was obtained from discontinued medication entries and consisted of all medication stops with an ADR reported as a reason for discontinuation and a random selection of 1600 medication stops that were not ADR-related. The latter selection was made because we expected that the clinical notes corresponding to these medication stops might also contain information on possible ADRs. Thus, this selection made it more likely that medication and ADRs would be identified when compared with a random selection of all clinical notes (Figure 2). These 2 selections of medication stops were merged with the corresponding clinical notes and resulted in a data set of 3000 unique clinical notes as there were some notes linked to medication stops that reported ADRs as well as medication stops that did not report an ADR.

The Medical Research Ethics Committee of the University Medical Center Utrecht declared that research within the Cardiology Centers of the Netherlands database does not fall under the Dutch Medical Research Involving Human Subjects Act (proposal number 17/359).

Labeling

In total, 2 researchers (KRS and ME) independently labeled all clinical notes in the test set. Clinical notes containing ADR information were labeled as positive. When a note was labeled positively, all words in the text describing the medication and ADR combinations were extracted. Discrepancies in labeling between the 2 researchers were discussed, and interobserver variability was evaluated. Furthermore, a validation data set of 3000 unique clinical notes was labeled by one of the researchers (either KRS or ME). These notes were used for identification of thresholds for the word embedding models and for intermediate, qualitative, and direct feedback.

Preprocessing Clinical Notes

Before applying word embedding models to the clinical notes, the text underwent multiple preprocessing steps. First, all text was converted to lowercase and unidecoded. Second, the clinical notes were tokenized with a regular expression tokenizer set to greedy tokenization for every word in the presented text. Third, all numerical tokens were converted into their written form (number normalization [20]). It is assumed that this results in

numbers being more closely related in vector space (ie, *16* and *18* vs *sixteen* and *eighteen*). Doses were removed from the text using regular expressions. The doses were removed to reduce the similarity between frequently prescribed doses and specific medications. This would otherwise contaminate the word embedding models used for identification of medication. Finally, for each token, a check was performed to determine if the token was in the unigram word embedding model. If this was not the case, the word was removed from the list of tokens. An example of a text going through this process is presented in Multimedia Appendix 1, Figure S1. The text was preprocessed using Python version 3.7.9 (Python Software Foundation [21]) using the nltk package (version 3.5) [22].

Word Embedding Models

For the automatic identification of ADRs from the text, word embedding models were developed. In total, 2 Word2Vec models imported from the Python Gensim package (version 3.8.0) [23] were trained on the complete set of 277,398 clinical notes [24]. A unigram model was developed using vectors for single words. This model included all words and derived vectors that occurred more than once in the complete set of clinical notes. The second model used a combination of single words, bigrams, and the derived vectors (bigram model). For the development of this model, words that occurred together >5 times were represented as a vector. Stop words imported from the nltk package [22] were removed from the text. A skipgram approach was used.

The Word2Vec settings were a vector size of 200 dimensions, a window of 5 words around the main word, and 5 iterations of learning. Word embedding models were qualitatively evaluated through inspection of the similarity among words [25].

Identification of Medication and ADRs

A list of search words was created for both medication and ADRs. The medication search list was based on different groups of cardiovascular medications (Multimedia Appendix 2, Table S1). For ADR identification, the most frequently reported ADRs (Multimedia Appendix 2, Table S2) in the discontinued medication entries were considered. From these ADRs, a list of search words for ADR recognition was compiled (Multimedia Appendix 2, Table S1).

Word embeddings were used for evaluation of the clinical note. First, the cosine similarity between each word in the clinical note and the search words for medication was calculated. A medication was identified if the cosine similarity was above a predefined threshold (Multimedia Appendix 2, Table S1). If no medication was found in the text, a second search was performed to identify a mention of ADRs using more general search words such as *adverse drug reaction*. If these search words were also not identified in the text, the clinical note was automatically labeled as not containing an ADR (Figure 1, step 1).

Second, after identification of a medication, the clinical note was searched for ADRs using a predefined search area around the identified medication (Figure 1, step 2). This search area was restricted to prevent an increasing number of false positives and could be adjusted if it seemed too strict or too wide. This was one of the settings adjusted during the evaluation of the pipeline.

After this, the area was checked for *non-ADR keywords*. These words occurred immediately before or after the medication and indicated a medication change or extension, such as *increase* and *double*. Therefore, these words did not indicate the presence of an ADR. List comparison was used, in which the tokenized form of the clinical note was compared with a list of words that pointed toward a medication change not likely because of an ADR (Multimedia Appendix 2, Table S3).

The final step in the search for ADRs was the actual identification (Figure 1, step 3). In total, 2 sequential approaches were developed for this purpose. The first approach included the application of the MedDRA. A selection of the lower-level MedDRA terms (Lowest Level Terms) [17] was checked with text retrieval and string matching in the defined search area around the medication. Inclusion or exclusion of the MedDRA was one of the settings adjusted during the evaluation of the pipeline.

The second approach for identification of ADRs was the use of unigram and bigram word embedding models. For each word in the search area, the cosine similarity with the search words for ADRs was computed (Multimedia Appendix 2, Table S1). If this similarity was above the predefined threshold, the word was identified as an ADR. Threshold-setting was performed using a grid search. Visual inspection of the graphical representation of the number of correct matches for a specific word (Multimedia Appendix 1, Figure S2) and evaluation of the included words after inspection of the list of most similar words resulted in the setting of the thresholds. For example, in the case of a specific medication, the threshold was set such that spelling mistakes and closely related medications were selected but not words that were related to a significant other medication group or words that did not describe medication but a certain disease or condition. For this analysis, the validation data set was used. This is explained in more detail in Multimedia Appendix 1.

Pipeline Versions and Tasks

The pipeline was developed to execute four different tasks: (1) a binary classification of whether the clinical note contained an ADR (Figure 3A and Figure 4A), (2) the extraction of the medication that causes an ADR (Figure 3B and Figure 4B), (3) the extraction of the ADR individually (Figure 3C and Figure 4C), and (4) the exact extraction of the medication and corresponding ADR (Figure 3D and Figure 4D).

Figure 3. Performance of different experimental versions of the pipeline with the inclusion of the MedDRA on the different tasks (A: binary evaluation, B: medication identification, C: ADR identification, D: medication and ADR + adverse drug reaction identification). ADR: adverse drug reaction; MedDRA: Medical Dictionary for Regulatory Activities; NPV: negative predictive value; PPV: positive predictive value.

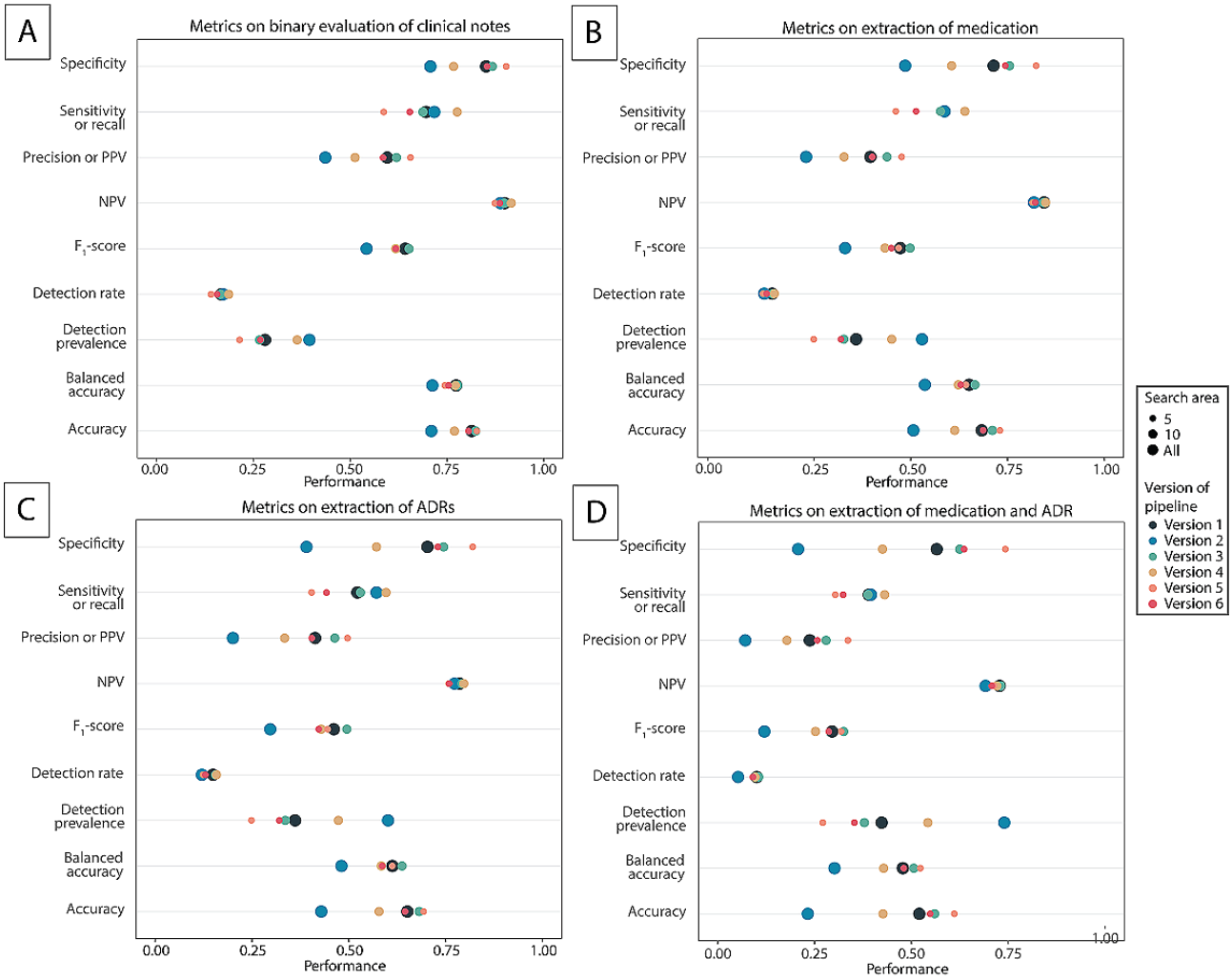
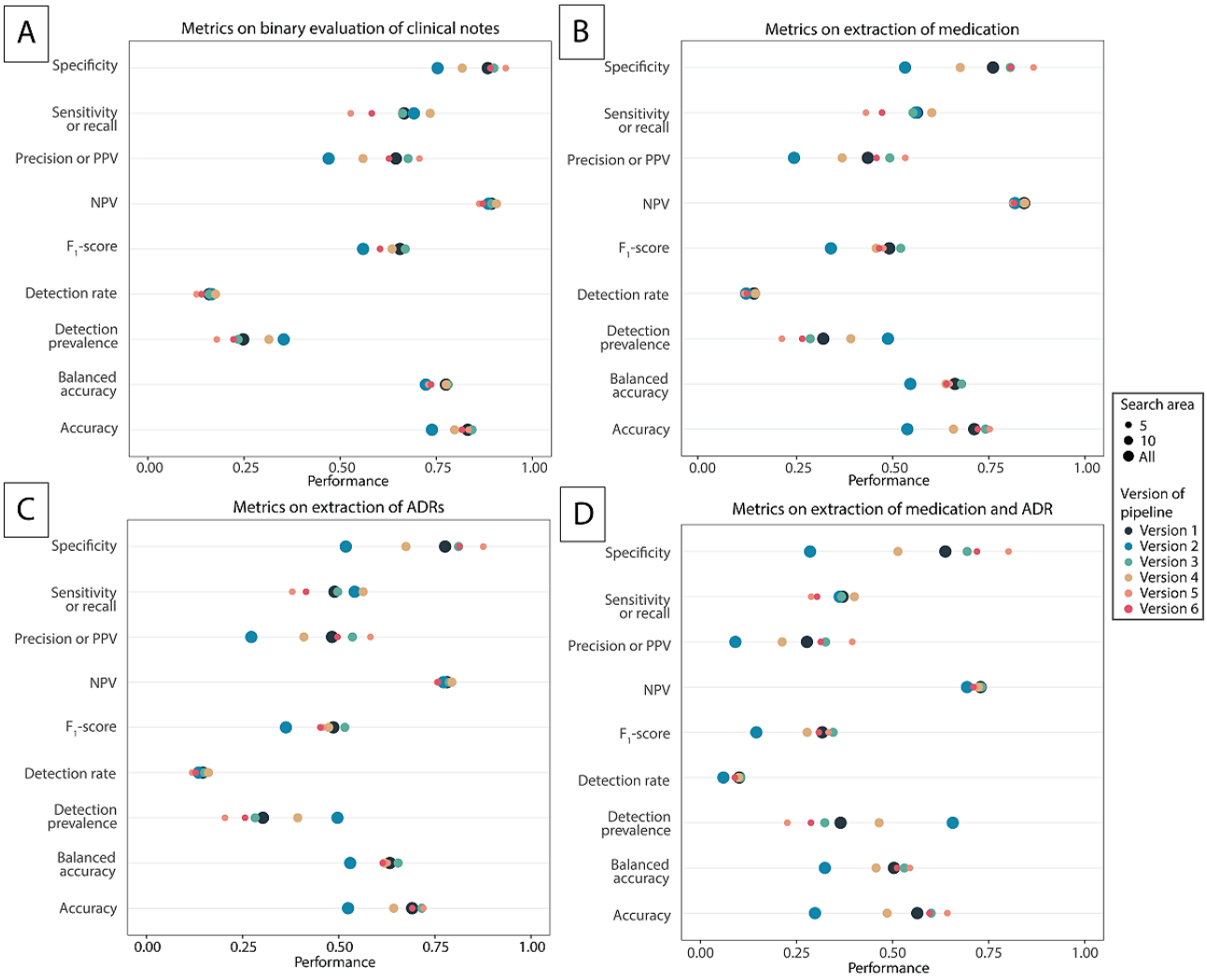


Figure 4. Performance of different experimental versions of the pipeline without the use of the MedDRA on the different tasks (A: binary evaluation, B: medication identification, C: ADR identification, D: medication and ADR + adverse drug reaction identification). ADR: adverse drug reaction; MedDRA: Medical Dictionary for Regulatory Activities; NPV: negative predictive value; PPV: positive predictive value.



Multiple settings were changed during the analysis to evaluate the performance of the predefined tasks of different experimental designs of the pipeline: inclusion or exclusion of the MedDRA for ADR identification, inclusion or neglect of punctuation for demarcation of the search area, and size of the search area. Table 1 provides an overview of the different settings evaluated in the versions of the pipeline. Analysis of the pipeline was performed using Python version 3.7.9 [21].

Table 1. Settings of the pipeline features of the different computational experiments.

Version	Words in search area	Considering punctuation	Version without MedDRA[a]
1A	All	Yes	1B
2A	All	No	2B
3A	10	Yes	3B
4A	10	No	4B
5A	5	Yes	5B
6A	5	No	6B

[a]MedDRA: Medical Dictionary for Regulatory Activities.

Performance Metrics

The pipeline was evaluated on the test set of 988 labeled clinical notes. Different metrics were calculated to assess the performance of different versions of the pipeline. The metrics that were calculated included accuracy and balanced accuracy, sensitivity, specificity, precision or positive predictive value, negative predictive value, recall, F_1 score, detection rate, and detection prevalence. An elaborate overview of the performance metrics and the evaluation process can be found in Multimedia Appendix 3, Table S1 and Tables S2-S6, respectively. The outcome was evaluated using the R programming language version 4.0.2 (R Foundation for Statistical Computing [26]) and RStudio version 1.3.1093 (RStudio Team [27]). The caret package was used for evaluation (version 6.0-86) [28].

Results

Data Set

The information on the complete data set for word embedding models, validation set, and test set is described in Table 2. The characteristics of the included free text are the informal writing style, use of abbreviations, and relatively short text length. Multimedia Appendix 3 contains 4 different translated examples of clinical notes, as shown in Multimedia Appendix 3, Table S2.

Table 2. Characteristics of selected clinical notes for development of the word embedding models, validation set, and test set.

Variable	Word embedding models	Validation set	Test set
Language	Dutch	Dutch	Dutch
Number of unique records	277,398	3000	988
Unique patients	108,940	2707	955
Number of unique tokens	96,086	9297	5464
Number of tokens per record, mean (SD)	54 (44)	53 (44)	53 (48)
Number of tokens per record, median (IQR)	43 (26-70)	42 (25-67)	41 (24-66)
Individuals of the female sex, n (%)	56,527 (51.89)	1320 (49.07)	459 (48.06)

Word Embedding Models

Several search terms of the prototype were independently reviewed in the word embedding models to evaluate the performance of the word embedding models. Table 3 lists a selection of these keywords and the 5 most similar words. It was noted that, if the search word was a specific group of medications (eg, β-blockers), other groups of medications were also identified (eg, *diltiazem* in the case of the search word β-*blocker*). As the identified word was used for the analysis and not the search word, this had no consequences for the analysis.

Free text from clinical notes was used in the training of the word embedding models. These are domain-specific data, which can improve the embedding of domain-specific words. An illustrative example is the word embedding of *red*. In our word embedding models trained specifically on medical text, *red* was closely associated with *itching*, *swollen*, *irritated*, and *colourings*, whereas, in word embeddings on general text, *red* would be associated with other colors.

Table 3. Selection of results from the word embedding models, adverse drug reaction, and medication search words, and a selection of the most relevant similar words where spelling mistakes are excluded. Similarity is based on the cosine similarity.

Keyword	Most similar words in Dutch (English, cosine similarity)
Pijn op de borst (chest pain)	*Druk op de borst* (chest pressure, 0.80), *kramp op de borst* (chest cramping, 0.70), *pijn in de armen* (pain in the arms, 0.68), and *retrosternale pijn* (retrosternal pain, 0.67)
Verminderde conditie (decreased condition)	*Afname conditie* (decreasing stamina, 0.63), *conditieverlies* (loss of condition, 0.63), *verminderde inspanningstolerantie* (decreased exercise tolerance, 0.62), and *overmating transpireren* (excessive sweating, 0.62)
Oedeem (edema)	*Perifeer* (peripheral edema, 0.81), *enkeloedeem* (ankle edema, 0.80), *pitting* (pitting edema, 0.80), and *enkels* (ankle edema, 0.75)
Hoesten (coughing)	*Sputum* (sputum, 0.75), *slijm* (mucus, 0.71), *hoestklachten* (coughing complaints, 0.70), and *kuchen* (to cough, 0.70)
Duizelig (dizziness)	*Zweterig* (sweaty, 0.73), *misselijk* (nauseous, 0.71), *zweverig* (floaty, 0.70), and *draaierig* (dizzy, 0.69)
Statine (statin)	*Simvastatine* (simvastatin, 0.80), *pravastatine* (pravastatin, 0.76), *crestor* (rosuvastatin, 0.75), and *atorvastatine* (atorvastatin, 0.74)
Betablokker (β-blocker)	Metoprolol (0.74), atenolol (0.71), diltiazem (0.66), and bisoprolol (0.65)
Antistolling	*Acenocoumarol* (acenocoumarin, 0.80), *anticoagulantia* (anticoagulants, 0.78), *NOAC* (novel oral anticoagulant, 0.77), and *fenprocoumon* (phenprocoumon, 0.74)
Amlodipine	Nifedipine (0.85), lisinopril (0.82), barnidipine (0.81), and enalapril (0.79)

Interobserver Variability

A test set (n=988 clinical notes) was manually labeled by 2 independent researchers (KRS and ME) and used for the evaluation of the pipeline. During this process, 91.9% (908/988) of the clinical notes were identically labeled. This resulted in an interobserver variability of 91% for the binary presence of an ADR. Regarding the literal extraction of the ADR and the medication, there were 21.8% (215/988) of instances where the result differed among the researchers. This was mostly due to a difference in taking adjectives or adverbs into account or a different interpretation of the clinical note. As the pipeline was trained on 1-word and 2-word ADRs, it was decided that these words would not be considered.

Manual labeling of the 988 clinical notes in the test set resulted in 23.9% (237/988) notes that were binary classified as containing an ADR. In the notes, 286 medication names (task 2) and 364 individual ADRs (task 3) were mentioned. These notes contained a total of 392 combinations of triggered ADRs (task 4) and corresponding medications.

Evaluation of the Pipeline

Figures 3 and 4 show the performance of the pipeline on the different metrics and for the different tasks. Multimedia Appendix 2, Table S4 shows the values for true and false negatives and true and false positives per version and per task. The task for binary classification achieved the highest accuracy, varying from 0.70 to 0.84 (Figure 3A). However, as this was the easiest task, the accuracy of the pipeline on the exact extraction of medication and ADR together was much lower, varying from 0.23 to 0.64 (Figure 3D).

If we look at the specific settings of the different pipelines, the results show that the addition of the MedDRA to the pipeline did not lead to an increase in the performance of the pipeline (Figures 4A-4D). Overall, the inclusion of punctuation led to a better performance than transcending sentences (versions 1, 3, and 5), and a search area of 5 words seemed to lead to the best results overall (versions 5 and 6).

The negative predictive value—the chance that no ADR was present when the pipeline did not produce an ADR—was approximately the same per task (0.69-0.91) for all versions of the pipeline. However, the positive predictive value (ie, the chance that, when the pipeline reported an ADR, it was in fact reported in the clinical notes) varied much more per version (Figures 3 and 4) and varied between 0.071 and 0.71. This could be explained by the proportion of false negatives. The proportion of false negatives did not vary much per version of the pipeline for a given task. However, the proportion of false positives had much more variety, caused by a change in the search area and the inclusion or exclusion of punctuation, which led to more ADRs found with a specific medication.

The optimal version of the pipeline depends on the task for which the pipeline is used. If the task is to select notes based on whether they contain ADRs, the results of the binary classification task (task 1) are most relevant. For this task, version 3B (ie, no MedDRA used, search area of 10 words, and considering punctuation) generated the highest accuracy (0.84) and F_1 score (0.67). In this case, 8.1% (80/988) of notes were classified as false negatives, indicating that 8.1% (80/988) of notes would not be selected when looking for ADRs. The most optimal version based on accuracy for identification of medication, ADRs, and ADRs and medication combined was version 5B, with an accuracy for the different tasks of 0.75, 0.72, and 0.64, respectively. Version 3B was the optimal version when emphasis was on the F_1 score, with scores of 0.52, 0.52, and 0.35 for identification of medication, ADRs, and medication and ADRs combined, respectively.

During the evaluation of the notes in the test set, the prototype incorporating the MedDRA required approximately 70 minutes to generate an outcome for all notes, whereas the versions without the MedDRA took approximately 14 seconds.

Discussion

Principal Findings

In this study, the ADRIN method and a corresponding prototype were developed. The method was evaluated on a subset of clinical notes. Different versions of the prototype led to differing results on the various tasks. The optimal version of the pipeline depends on the task and the trade-off being made—Is it more valuable to find as many medication and ADR combinations as possible or to find fewer ADRs but also make fewer mistakes? If the goal is the former, a larger search area is better. However, even with the entire note as the search area, at least 8% of all medication and ADR combinations were missed. When one wants to be more accurate, a smaller search area is preferred, and punctuation should be considered. This reduces the number of false positives generated, which results in increased accuracy and F_1 score.

Surprisingly, the versions incorporating the MedDRA performed worse on most tasks than the same versions without the MedDRA. The negative effect of the MedDRA on the performance was due to the large increase in false positives it generated. This was caused by string matching with the MedDRA, leading to more identifications than the specific set of frequently occurring ADRs defined by the predefined search words. Incorporation of the MedDRA could lead to an improved uptake of rare ADRs, but this was not evaluated in more detail. Furthermore, misspelled ADRs were not recognized by the MedDRA search, creating added value for the incorporation of word embedding models. Moreover, implementation of the MedDRA in the prototype significantly increased execution time, a significant attribute if real-time evaluation of clinical notes is required.

Illustrative of the underreporting of ADRs is that, in 60% (54,765/91,273) of the discontinued medication entries, no reason was reported for ending the medication in the registration of a patient's medication. However, 61.5% (36,564/59,426) of clinical notes were matched to these medication entries, which illustrates the potential additional value of clinical notes in unraveling ADRs in this data set.

When we put these results in light of the ongoing developments of ADR extraction from clinical notes, we see that the performance of our pipeline is similar to that of other presented pipelines. First, most publications have focused on the automatic

extraction of ADRs, ADEs, or adverse events [29-32], whereas our study identified the combination of medication and triggered ADR. Another publication that identified both ADR and medication showed increased performance, with F_1 scores for drug, ADR, and combination of drug and ADR of 0.930, 0.887, and 0.534, respectively [33], versus the performance of 0.52, 0.51, and 0.34, respectively, that we showed. When comparing methodologies, our method predominantly relies on internal information and similarity from word embeddings, whereas Tang et al [33] use external reference sources for the development of their dictionaries, which is the case in most studies. The use of word embeddings increases the identification of spelling mistakes in medication and ADRs, brand names, and synonyms. However, in our methodology, there were also an increased number of false positives.

Thus, word embedding models can be used for the identification of spelling mistakes and brand names of medications. However, for the identification of synonyms, the use case must be critically evaluated. It was shown that words that indicated what was done with a specific prescription (eg, *to lower* and *to increase*) were considered similar by the word embedding models. Therefore, it is not suitable to use word embedding models for identification of *non-ADR* keywords, which was solved with string matching in the ADRIN method. The use of domain-specific word embedding models is not new or limited to ADR identification but is increasingly used in the evaluation of clinical notes (eg, in ICD-10 classification [15] and anonymization [34]).

Second, publications on identification of ADRs in the English language are numerous, using different methods such as General Architecture for Text Engineering NLP [35], trigger words [30], or trigger phrases [31]. Regarding foreign languages, the field is maturing. Methods developed for the English language can, in some cases, be transferred to other languages. However, the effort that must be put into this depends on the complexity of the task and the level of text interpretation [36]. For example, a study of Danish clinical notes obtained better performance (recall of 0.75 in [32] vs 0.59 in this study) for sole ADR identification. This study missed approximately one-fourth of all possible ADRs, whereas our optimal performance missed approximately 40%. However, this pipeline included manual dictionary selection and more rule-based filters in the model [32].

We chose to use the presence of a mention of medication in the clinical note as the starting point for identification of an ADR. However, this might result in experienced ADRs being missed. The performance of the pipeline might benefit from the removal of the identification of medication and, for example, coupling with structured medication prescriptions to obtain information about medication use. However, the end user should be aware that this might also increase the number of false positives as the presence of an ADR is no longer limited by the presence of medication.

Limitations that were identified during the evaluation of the method and prototype are primarily related to missed ADRs from the clinical free text even when the entire clinical note was used for analysis. This problem can be solved by lowering the identifying threshold, but this would also lead to a potentially large increase in false positives. The use of machine and deep learning models can improve the performance of the ADRIN method. However, a large data set of labeled clinical notes is required to train machine and deep learning models, which was unavailable during the development of this model.

An overall limitation of the prototype is the direct translatability to other languages. The word embedding models were specifically trained on Dutch clinical notes. Search terms for word embedding functions must be translated into the new language to implement this method in clinical notes in a different language. Moreover, word embedding models must be trained with notes in the specific language before applying the developed method. Therefore, a large number of clinical free-text notes are required. Because of ethical and privacy constraints, this can be hard to acquire. However, it is technically possible to test and validate the ADRIN method in other languages through translation of search words and negations and after training word embedding models with the specific language.

Conclusions

In conclusion, the ADRIN method and prototype are effective in recognizing ADRs in Dutch clinical notes. Surprisingly, incorporation of the MedDRA did not result in improved identification on top of word embedding models. However, not all versions of the prototype were equally accurate. Different parameter settings can be chosen for the prototype to optimize the task of the model. In a future stage, incorporation of a pipeline in an electronic health record environment can lead to automatic identification and registration of ADRs. This saves the physician’s precious time and decreases the previously mentioned underreporting of ADRs in clinical care, increasing our knowledge about ADRs, which might ultimately benefit the patient.

Acknowledgments

This study was funded by the Dutch Heart Foundation (CVON-AI 2018B017).

Authors' Contributions

KRS, ME, SHB, HMdR, MS, and NCOM designed the study. KRS and ME designed the analysis plan. ME coded the pipeline, and KRS improved and optimized the pipeline. ME and KRS labeled the clinical notes, and FG provided critical evaluation of the labeling. MS and NCOM supervised the project. All authors critically edited the manuscript, approved the final work, and agree to be accountable for the accuracy and integrity of the work.

Conflicts of Interest

LH, GAS, and IIT are employed by the Cardiology Centers of the Netherlands.

Multimedia Appendix 1
Supplementary methods: text preprocessing and threshold setting for word embedding models.
[DOCX File , 162 KB - medinform_v10i1e31063_app1.docx]

Multimedia Appendix 2
Overview of model settings and results.
[DOCX File , 26 KB - medinform_v10i1e31063_app2.docx]

Multimedia Appendix 3
Evaluation of the pipeline.
[DOCX File , 19 KB - medinform_v10i1e31063_app3.docx]

References

1. Hazell L, Shakir SA. Under-reporting of adverse drug reactions : a systematic review. Drug Saf 2006;29(5):385-396. [doi: 10.2165/00002018-200629050-00003] [Medline: 16689555]
2. Seruga B, Templeton AJ, Badillo FE, Ocana A, Amir E, Tannock IF. Under-reporting of harm in clinical trials. Lancet Oncol 2016 May;17(5):209-219. [doi: 10.1016/S1470-2045(16)00152-2] [Medline: 27301048]
3. Leening MJ, Heeringa J, Deckers JW, Franco OH, Hofman A, Witteman JC, et al. Healthy volunteer effect and cardiovascular risk. Epidemiology 2014 May;25(3):470-471. [doi: 10.1097/EDE.0000000000000091] [Medline: 24713887]
4. de Vries ST, Denig P, Ekhart C, Burgers JS, Kleefstra N, Mol PG, et al. Sex differences in adverse drug reactions reported to the National Pharmacovigilance Centre in the Netherlands: An explorative observational study. Br J Clin Pharmacol 2019 Jul;85(7):1507-1515 [FREE Full text] [doi: 10.1111/bcp.13923] [Medline: 30941789]
5. Kongkaew C, Noyce PR, Ashcroft DM. Hospital admissions associated with adverse drug reactions: a systematic review of prospective observational studies. Ann Pharmacother 2008 Jul;42(7):1017-1025. [doi: 10.1345/aph.1L037] [Medline: 18594048]
6. Sultana J, Cutroneo P, Trifirò G. Clinical and economic burden of adverse drug reactions. J Pharmacol Pharmacother 2013 Dec;4(Suppl 1):73-77 [FREE Full text] [doi: 10.4103/0976-500X.120957] [Medline: 24347988]
7. Postigo R, Brosch S, Slattery J, van Haren A, Dogné J, Kurz X, et al. EudraVigilance medicines safety database: publicly accessible data for research and public health protection. Drug Saf 2018 Jul;41(7):665-675 [FREE Full text] [doi: 10.1007/s40264-018-0647-1] [Medline: 29520645]
8. Murdoch TB, Detsky AS. The inevitable application of big data to health care. J Am Med Assoc 2013 Apr 3;309(13):1351-1352. [doi: 10.1001/jama.2013.393] [Medline: 23549579]
9. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
10. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. J Allergy Clin Immunol 2020 Feb;145(2):463-469 [FREE Full text] [doi: 10.1016/j.jaci.2019.12.897] [Medline: 31883846]
11. Khattak FK, Jeblee S, Pou-Prom C, Abdalla M, Meaney C, Rudzicz F. A survey of word embeddings for clinical text. J Biomed Inform 2019;100S:100057 [FREE Full text] [doi: 10.1016/j.yjbinx.2019.100057] [Medline: 34384583]
12. Zhao M, Masino A, Yang C. A framework for developing and evaluating word embeddings of drug-named entity. In: Proceedings of the BioNLP 2018 workshop. 2018 Presented at: BioNLP 2018 workshop; July 2018; Melbourne, Australia p. 156-160. [doi: 10.18653/v1/w18-2319]
13. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform 2018 Sep 11:12-20. [doi: 10.1016/j.jbi.2018.09.008] [Medline: 30217670]
14. Banerjee I, Madhavan S, Goldman R, Rubin D. Intelligent word embeddings of free-text radiology reports. arXiv. 2017. URL: https://arxiv.org/abs/1711.06968 [accessed 2022-01-04]
15. Sammani A, Bagheri A, van der Heijden PG, Te Riele AS, Baas AF, Oosters CA, et al. Automatic multilabel detection of ICD10 codes in Dutch cardiology discharge letters using neural networks. NPJ Digit Med 2021 Feb 26;4(1):37. [doi: 10.1038/s41746-021-00404-9] [Medline: 33637859]
16. Dai H, Su C, Wu C. Adverse drug event and medication extraction in electronic health records via a cascading architecture with different sequence labeling models and word embeddings. J Am Med Inform Assoc 2020 Jan 01;27(1):47-55 [FREE Full text] [doi: 10.1093/jamia/ocz120] [Medline: 31334805]
17. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf 1999 Feb;20(2):109-117. [doi: 10.2165/00002018-199920020-00002] [Medline: 10082069]

18. Bots SH, Siegersma KR, Onland-Moret NC, Asselbergs FW, Somsen GA, Tulevski II, et al. Routine clinical care data from thirteen cardiac outpatient clinics: design of the Cardiology Centers of the Netherlands (CCN) database. BMC Cardiovasc Disord 2021 Jun 10;21(1):287 [FREE Full text] [doi: 10.1186/s12872-021-02020-7] [Medline: 34112101]
19. Menger V, Scheepers F, van Wijk LM, Spruit M. DEDUCE: a pattern matching method for automatic de-identification of Dutch medical text. Telemat Informat 2018 Jul;35(4):727-736. [doi: 10.1016/j.tele.2017.08.002]
20. Sproat R, Black AW, Chen S, Kumar S, Ostendorf M, Richards C. Normalization of non-standard words. Comput Speech Lang 2001 Jul;15(3):287-333. [doi: 10.1006/csla.2001.0169]
21. Python Language Reference, Version 3.7.9. URL: https://www.python.org/ [accessed 2022-01-11]
22. Bird S. NLTK: The Natural Language Toolkit. In: Proceedings of the COLING/ACL on Interactive Presentation Sessions. 2006 Presented at: COLING/ACL on Interactive Presentation Sessions; July 17 - 18, 2006; Sydney Australia p. 69-72. [doi: 10.3115/1225403.1225421]
23. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. In: Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks. 2010 Presented at: LREC 2010 workshop New Challenges for NLP Frameworks; 2010; University of Malta, Valletta, Malta p. 46-50 URL: https://is.muni.cz/publication/884893/en/Software-Framework-for-Topic-Modelling-with-Large-Corpora/Rehurek-Sojka
24. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of the 1st International Conference on Learning Representations, ICLR 2013. 2013 Presented at: 1st International Conference on Learning Representations, ICLR 2013; May 2-4, 2013; Scottsdale, Arizona, USA URL: https://arxiv.org/abs/1301.3781
25. Wang B, Wang A, Chen F, Wang Y, Kuo CJ. Evaluating word embedding models: methods and experimental results. APSIPA Transactions on Signal and Information Processing 2019 Jul 08;8:1-13. [doi: 10.1017/atsip.2019.12]
26. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. URL: https://www.r-project.org/ [accessed 2022-01-11]
27. RStudio Team. RStudio: integrated development environment for R. RStudio, PBC, Boston, MA. 2021. URL: https://www.rstudio.com/ [accessed 2022-01-11]
28. Kuhn M. The caret Package. 2009. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.2466&rep=rep1&type=pdf [accessed 2022-01-05]
29. Honigman B, Lee J, Rothschild J, Light P, Pulling RM, Yu T, et al. Using computerized data to identify adverse drug events in outpatients. J Am Med Inform Assoc 2001;8(3):254-266 [FREE Full text] [doi: 10.1136/jamia.2001.0080254] [Medline: 11320070]
30. Murff HJ, Forster AJ, Peterson JF, Fiskio JM, Heiman HL, Bates DW. Electronically screening discharge summaries for adverse medical events. J Am Med Inform Assoc 2003;10(4):339-350 [FREE Full text] [doi: 10.1197/jamia.M1201] [Medline: 12668691]
31. Cantor MN, Feldman HJ, Triola MM. Using trigger phrases to detect adverse drug reactions in ambulatory care notes. Qual Saf Health Care 2007 Apr;16(2):132-134 [FREE Full text] [doi: 10.1136/qshc.2006.020073] [Medline: 17403760]
32. Eriksson R, Jensen PB, Frankild S, Jensen LJ, Brunak S. Dictionary construction and identification of possible adverse drug events in Danish clinical narrative text. J Am Med Inform Assoc 2013;20(5):947-953 [FREE Full text] [doi: 10.1136/amiajnl-2013-001708] [Medline: 23703825]
33. Tang Y, Yang J, Ang PS, Dorajoo SR, Foo B, Soh S, et al. Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer. Int J Med Inform 2019 Aug;128:62-70. [doi: 10.1016/j.ijmedinf.2019.04.017] [Medline: 31160013]
34. Abdalla M, Abdalla M, Rudzicz F, Hirst G. Using word embeddings to improve the privacy of clinical notes. J Am Med Inform Assoc 2020 Jun 01;27(6):901-907 [FREE Full text] [doi: 10.1093/jamia/ocaa038] [Medline: 32388549]
35. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of adverse drug events from free text electronic patient records and information in a large mental health case register. PLoS One 2015;10(8):e0134208 [FREE Full text] [doi: 10.1371/journal.pone.0134208] [Medline: 26273830]
36. Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical natural language processing in languages other than English: opportunities and challenges. J Biomed Semantics 2018 Mar 30;9(1):12 [FREE Full text] [doi: 10.1186/s13326-018-0179-8] [Medline: 29602312]

Abbreviations

ADE: adverse drug event
ADR: adverse drug reaction
ADRIN: Adverse Drug Reaction Identification in Clinical Notes
MedDRA: Medical Dictionary for Regulatory Activities
NLP: natural language processing

Edited by C Lovis; submitted 10.06.21; peer-reviewed by KM Kuo, JP Goldman; comments to author 23.09.21; revised version received 02.11.21; accepted 14.11.21; published 25.01.22.

Please cite as:
Siegersma KR, Evers M, Bots SH, Groepenhoff F, Appelman Y, Hofstra L, Tulevski II, Somsen GA, den Ruijter HM, Spruit M, Onland-Moret NC
Development of a Pipeline for Adverse Drug Reaction Identification in Clinical Notes: Word Embedding Models and String Matching
JMIR Med Inform 2022;10(1):e31063
URL: https://medinform.jmir.org/2022/1/e31063
doi:10.2196/31063
PMID:35076407

Original Paper

Validity and Reliability of the Korean Version of the Health Information Technology Usability Evaluation Scale: Psychometric Evaluation

Jisan Lee[1,2], RN, PhD; Rebecca Schnall[3,4], MPH, PhD

[1]Department of Nursing Science, College of Life & Health Sciences, Hoseo University, Asan, Republic of Korea

[2]The Research Institute for Basic Sciences, Hoseo University, Asan, Republic of Korea

[3]School of Nursing, Columbia University, New York, NY, United States

[4]Department of Population and Family Health, Mailman School of Public Health, Columbia University, New York, NY, United States

Corresponding Author:
Rebecca Schnall, MPH, PhD
School of Nursing
Columbia University
560 West 168th Street
New York, NY, 10032
United States
Phone: 1 2123426886
Email: rb897@columbia.edu

Abstract

Background: Rigorous development of mobile technologies requires the use of validated instruments to evaluate the usability of these tools, which has become more relevant with the expansion of these technologies. Although various usability evaluation tools have been developed, there are relatively few simple evaluation instruments that have been validated across diseases and languages in mobile health (mHealth) information technology for use in multiple diseases.

Objective: The purpose of this study is to validate the Korean version of the Health Information Technology Usability Evaluation Scale (Korean Health-ITUES) and assess its applicability for different health conditions.

Methods: To develop the Korean Health-ITUES, we used a validation process involving the following 3 steps: (1) customization of the Health-ITUES for menstrual symptoms, (2) translation of the Health-ITUES from English into Korean, and (3) examination of the reliability and validity of the instrument. The translation process adhered to the World Health Organization (WHO) guidelines for translation and back-translation, expert review, and reconciliation.

Results: The Korean Health-ITUES showed reliable internal consistency with Cronbach α=.951; meanwhile, factor loadings of the 20 items in the 4 subscales ranged from 0.416 to 0.892.

Conclusions: The Health-ITUES demonstrated reliability and validity for its use in assessing mHealth apps' usability in young Korean women with menstrual discomfort. Given the strong psychometric properties of this tool in Korean and English and across 2 different health conditions, the Health-ITUES is a valid and reliable instrument for assessing the usability of mHealth apps. The Health-ITUES is also a valid instrument for evaluating mHealth technologies, which are widely used by patients to self-manage their health and by providers to improve health care delivery.

(JMIR Med Inform 2022;10(1):e28621) doi:10.2196/28621

KEYWORDS

mobile application; menstruation; survey; questionnaire; translations; medical informatics; app; validity; reliability; usability; assessment; mHealth; evaluation

Introduction

Background

In the past decade, 1 of the most challenging components of technology development has been to ensure the usability of tools to ensure their quality in use [1,2]. To support use of the technology, usability must be assessed during the development process [3]. Usability is the measure of the quality of a user's experience when interacting with a system—whether a website, mobile technology, or any user-operated device [4]. In other words, usability refers to how well users can navigate a system to achieve their goals and how satisfied they are with the process. A successful system needs to work for its users, and it needs to work well. However, many mHealth technologies have been made available to the public, with insufficient attention devoted to their design, development, and evaluation [5]. Technologies produced with poor design and inadequate consideration of the needs of their intended users will be difficult to learn, misused, or underutilized and will ultimately fail to accomplish their objectives [6]. For this reason, usability has been widely recognized as a critical consideration in evaluating the efficacy of technologies [7].

Usability is especially critical for mobile technology, which is widely used in health care [8-10]. In fact, there were more than 800,000 mobile health (mHealth) apps in Apple App Store and Google Play Store in 2020. There is continued growth, with about 200 mHealth apps added each day, with some focusing on healthy eating, physical activity, and improved mental health [11-13]. To date, several studies have evaluated the effect of mHealth apps on disease management and prevention [14-17].

Research to evaluate the usability of mHealth apps has been conducted for various health states, such as chronic obstructive pulmonary disease, HIV, obesity, depression, anxiety, dysmenorrhea, and premenstrual syndrome [18-26]. However, to ensure the rigor of mHealth technologies, it is necessary to create tools to evaluate the quality and usability of mHealth apps [27].

The quality of mHealth apps requires understanding the context of their use and ensuring the apps' usability [28]. Therefore, it is important to develop mHealth tools using rigorous usability evaluation tools. Although there are several mobile app assessment tools, most have a large number of items or have only been validated for a single disease or in a single language [29,30]. Given these limitations, there is a need for instruments that are validated across languages and diseases. This is especially true in South Korea, which has 1 of the highest rates (94%) of smartphone use in the world. This high penetration of smartphones has enabled the rapid integration of mHealth apps [31,32]. However, despite the high usage of mHealth apps in South Korea, there is a dearth of availability of simple instruments to assess their usability.

The Study

This study sought to translate the Health Information Technology Usability Evaluation Scale (Health-ITUES) from English into Korean and validate its use. The Health-ITUES is a customized questionnaire comprising 20 items and a modified version of the Health Information Technology Usability Evaluation Model (Health-ITUEM) [33]. The English version of the Health-ITUES has been previously validated through exploratory factor analysis (EFA) and confirmatory factor analysis after use by nurses [34] and community-dwelling adults with HIV [35]; however, it has not been translated into other languages and validated. This study translated the Health-ITUES into Korean and validated it in a sample of 244 women who experienced menstrual-related symptoms and used a menstrual tracking app called PINKDIARY.

Methods

Sample/Participants

This study was approved by the institutional review board of the Catholic University of Pusan (CUPIRB-2019-003) before the commencement of study activities. Inclusion criteria for this study were unmarried women >20 and <39 years of age who were previously or are currently using the menstrual-tracking app PINKDIARY for more than a month. The app records the highest usage rate in Korea, and as of October 2021, it was ranked sixth in the health and fitness category in Apple App Store. This app is the official app of the Korean Association of Obstetricians and Gynecologists. PINKDIARY is used to track menstruation and premenstrual symptoms. Features of the app include symptom records, doctor consultations, an online community, and a shopping mall for menstrual items (eg, pads, tampons, and menstrual cups). Marriage and age were also inclusion criteria, as these may have affected the participants' usage of or experience with using the app.

The sample size was set to ensure a minimum number of participants based on the number of items in the instrument. Nunnally [36] recommended a minimum participant ratio of 10 participants:1 survey item. In this study, the target sample size was between 200 and 250 after multiplying the number of questions (20) by 10. This estimate was based on an anticipated attrition rate of about 20%. In the past, the dropout rate in app-related studies was about 20%-50%. Since this study was not an intervention study, the dropout rate was estimated at 20% [37].

Recruitment was conducted through the KakaoTalk (Kakao Corp) messenger and other online communities (eg, Everytime). Potential participants sent a screenshot of the PINKDIARY app to the researcher's messenger to authenticate their use of the app. The online consent form and questionnaire were developed in Survey Monkey, and the link was sent to participants on the KakaoTalk messenger. After filling in the consent form and questionnaire, a 2000 won (about US $2) online coffee coupon was sent to the participants as a token of appreciation for their time.

Step 1: Modification of the Health-ITUES

We customized the Health-ITUES, which was previously validated in a sample of persons with HIV, for women with menstrual discomfort [25]. In this study, menstrual discomfort was defined as primary dysmenorrhea and premenstrual syndrome (PMS), which are the most common menstrual discomfort symptoms among women of reproductive age

[38,39]. The modified version was reviewed by the senior author of this manuscript, RS.

Step 2: Korean Health-ITUES

The translation and back-translation followed WHO guidelines [40].

Forward Translation

To align with WHO guidelines, all items had to remain unchanged from their original meaning when translated and had to be translated into English by 2 or more translators. In this study, 2 Korea-born nurses who had lived in the United States for more than 5 years and received doctoral degrees in the United States translated the Health-ITUES from Korean into English. The translators then independently translated it from English into Korean. Following the independent translation, the 2 translators discussed the findings during 3 separate meetings until a consensus was reached.

Expert Review

The translated Health-ITUES was reviewed and reconfirmed by a bilingual (English/Korean) physician with dual training in medical informatics. The physician was fluent in both English and Korean. He checked whether the translated version applied to Korean-speaking people and revised the items with expressions or cultural differences that could cause different meanings to be conveyed. After the review process, the translation of the Health-ITUES from Korean into English was considered complete.

Back-Translation

Back-translation was conducted by a professional translator from an official translation company and a nurse who lived in the United States for over 10 years and was currently enrolled in a PhD program in the United States. The 2 translators translated the tool back into English and focused on culture and concepts rather than word differences, as was done during the translation process. Inconsistencies were reviewed by the first author to produce a back-translated version of the Health-ITUES.

Expert Review

The bilingual physician who participated in the first expert review after forward translation reviewed the back-translated version again. In this step, we focused on whether the Korean words before translation and the Korean words that were translated back had the same cultural and conceptual meanings, rather than focusing on whether they were completely identical.

Original Author's Review

After completing the translation and back-translation, the senior author (RS) reviewed the English version of the Health-ITUES and confirmed the content.

Step 3: Reliability and Validity

Pilot Test

According to WHO guidelines, the minimum number of pilot test respondents is 10 and should represent males and females of all ages and socioeconomic groups. However, in our study, due to time constraints and the fact that we reached data saturation after interviewing 5 women, we limited our sample size to only 5 respondents for this component part of the study [41]. Young women who majored in nursing and had previously used the menstrual app (for a minimum of 1 month to a maximum of 5 years) completed the survey items. Following completion of the survey, they provided feedback about the questionnaire's items through an in-depth interview. During the interview, we asked the participants to justify their responses and whether they encountered any words in the Health-ITUES that were difficult to understand.

Reliability and Validity Test

Internal consistency reliability was measured using Cronbach's alpha and Pearson's correlation coefficient. Construct validity was analyzed using exploratory factor analysis. Exploratory factor analysis was used to confirm the predicted factor loadings based on the original instrument [42].

Questionnaire

The questionnaire comprised the items described in Table 1.

Table 2 shows each item of the Health-ITUES and the Korean Health-ITUES customized for this study (see Multimedia Appendix 1 for the Korean Health-ITUES).

Table 1. The questionnaire's items.

Categories	Items (N=49), n (%)	Item description	Scale
General characteristics	5 (10)	• Sex, marital status, age, residential area, job	N/A[a]
Smartphone experience and menstrual app usage	11 (22)	• Smartphone type, which features they used most often, how long, how often they used the menstrual app, etc	N/A
Dysmenorrhea	2 (4)	• Pain on the first and second days of the menstrual period [23]	• Visual analog scale
PMS[b]	11 (22) Subscale: emotion (4 items), water congestion (4 items), pain (2 items), and appetite (1 item)	• The changes that participants experience before menstruation (an appetite item added to the shortened Premenstrual Assessment Form) [43]	• 6-point Likert-type scale with responses ranging from 1 (not at all) to 6 (very severe change). • The higher the score, the more severe the symptoms.
Korean Health-ITUES[c]	20 (42) Subscale: impact (3 items), perceived usefulness (9 items), perceived ease of use (5 items), and user control (3 items)	• Impact: high expectations for system impact and perceived usefulness as well as performance assessment of tasks through system usage • User-system interactions measured	• 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). • The higher the scale value, the higher the usability of the technology. • The overall Korean Health-ITUES score is the average of all items with the same weight for each item.

[a]N/A: not applicable.

[b]PMS: premenstrual syndrome.

[c]Health-ITUES: Health Information Technology Usability Evaluation Scale.

Table 2. Health-ITUES[a] and Korean Health-ITUES.

Health-ITUES [24]		Korean Health-ITUES	
Number	Item	Number	Item
Impact			
1	I think Mobile Video Information Provider (mVIP)[b] would be a positive addition for persons with HIV.	1	I think PINKDIARY[c] would provide positive health outcomes for women with menstrual discomfort.
2	I think mVIP would improve the quality of life of persons with HIV.	2	I think PINKDIARY would improve the quality of life of women with menstrual discomfort.
3	mVIP is an important part of meeting my information needs related to symptom self-management.	3	PINKDIARY helps to meet the information needs for the self-management of my menstrual-related symptoms.
Perceived usefulness			
1	Using mVIP makes it easier to self-manage my HIV-related symptoms.	1	Using PINKDIARY makes self-managing my menstrual-related symptoms easy.
2	Using mVIP enables me to self-manage my HIV-related symptoms more quickly.	2	Using PINKDIARY allows me to manage my menstrual-related symptoms more quickly.
3	Using mVIP makes it more likely that I can self-manage my HIV-related symptoms.	3	Using PINKDIARY makes self-managing my menstrual-related symptoms better.
4	Using mVIP is useful for self-management for HIV-related symptoms.	4	Using PINKDIARY is useful for the self-management of my menstrual-related symptoms.
5	I think mVIP presents a more equitable process for self-management of HIV-related symptoms.	5	I think PINKDIARY provides a more equitable process for the self-management of my menstrual-related symptoms. (Health equity refers to health equality, which means ensuring the optimal level of health for all people, regardless of income and educational level.)
6	I am satisfied with mVIP for self-management of HIV-related symptoms.	6	I am satisfied with PINKDIARY for the self-management of my menstrual-related symptoms.
7	I self-manage my HIV-related symptoms in a timely manner because of mVIP.	7	I can self-manage my menstrual-related symptoms in a timely manner thanks to PINKDIARY.
8	Using mVIP increases my ability to self-manage my HIV-related symptoms.	8	Using PINKDIARY enhances my ability to self-manage my menstrual-related symptoms.
9	I am able to self-manage my HIV-related symptoms whenever I use mVIP.	9	I can self-manage my menstrual-related symptoms when I use PINKDIARY.
Perceived ease of use			
1	I am comfortable with my ability to use mVIP.	1	I am satisfied with my ability to use PINKDIARY.
2	Learning to operate mVIP is easy for me.	2	It is easy for me to learn how to operate PINKDIARY.
3	It is easy for me to become skillful at using mVIP.	3	It was easy for me to become skillful in using PINKDIARY.
4	I find mVIP easy to use.	4	I find PINKDIARY easy to use.
5	I can always remember how to log on to and use mVIP.	5	I always remember how to log on to and use PINKDIARY.
User control			
1	mVIP gives error messages that clearly tell me how to fix problems.	1	PINKDIARY provides error messages that clearly explain how to solve problems with PINKDIARY.
2	Whenever I make a mistake using mVIP, I easily and quickly recover.	2	I can recover quickly and easily whenever I make a mistake while using PINKDIARY.
3	The information provided by mVIP (eg, online help, on-screen messages, and other documentations) is clear.	3	The information provided by PINKDIARY (eg, online help, screen messages, and other documents) is clear.

[a]Health-ITUES: Health Information Technology Usability Evaluation Scale.

[b]mVIP: HIV self-management app.

[c]PINKDIARY: menstrual app.

Statistical Analysis

Descriptive statistics were used to analyze the demographic characteristics of the study participants. Reliability and validity were analyzed using Cronbach α, Pearson correlation coefficient, and EFA. All analyses used IBM SPSS Statistics 24.0. Two-sided *P* values of <0.05 were considered statistically significant.

Ethical Considerations

All study participants provided informed consent, and the study design was approved by the appropriate ethics review board.

Results

Pilot Test

Five women completed the draft version of the survey and provided the following feedback about the Health-ITUES:

> *Overall, I think the questionnaire is easy to understand. However, in the case of items that need multiple responses, I would like for such questions to be clearly marked as "requiring multiple responses."* [Interviewee A]
>
> *I like this online questionnaire because there are appropriate app icon figures with items. However, I need further explanation about Item 8: I cannot understand what "equitable process" means.* [Interviewee B]
>
> *I did not have any problem with understanding the items in the questionnaire. However, I went to the next page without answering a few questions. If the respondent does not answer all the items, that is, if even one item has not been answered, please set the questionnaire such that it does not move to the next page.* [Interviewee C]

As a result of this feedback, item 8 was changed. More specifically, further explanation was added to clarify the term "equitable process" as follows: "Health equity refers to health equality, which means ensuring the optimal level of health for all people, regardless of income and educational level."

Demographic Characteristics

A total of 244 unmarried female participants completed the Health-ITUES. The participants' ages ranged from 20 to 36 years (mean=22.45, SD 3.039). Meanwhile, 127 (51.2%) participants majored in a health-related topic, and 117 (47.0%) participants majored in non-health-related topics for their university degrees. Participants resided in the following cities and provinces: Seoul (n=76, 30.5%), Gyeonggi-do (n=61, 24.6%), Chungcheong-do (n=52, 21.7%), Gyeongsang-do (n=48, 19.4%), and Else (7, 2.8%).

Internal Consistency Reliability

Internal consistency reliability was measured by Cronbach α, and the results are presented in Table 3. All items showed good Cronbach α values (>0.8) ranging from 0.83 to 0.94 (Table 3). All values were less than 0.95, which demonstrates that there was no redundancy among the items [44]. Internal scale correlation between items ranged from 0.45 to 0.71, indicating moderate to strong correlations. Notably, perceived usefulness was more highly correlated with impact than other subscales.

Table 3. Descriptive statistics, internal scale consistency score, and internal scale correlation for Korean Health-ITUES[a] subscales (N=244).

Subscale	Mean±SD	Impact		Perceived usefulness		Perceived ease of use		User control	
		Cronbach α	r	Cronbach α	r	Cronbach α	r	Cronbach α	r
Impact	10.87±2.52	0.84	—[b]	—	—	—	—	—	—
Perceived usefulness	34.51±6.55	—	0.707[c]	0.94	—	—	—	—	—
Perceived ease of use	19.60±4.13	—	0.510[c]	—	0.660[c]	0.91	—	—	—
User control	10.16±2.43		0.446[c]		0.584[c]	—	0.647[c]	0.83	—

[a]Health-ITUES: Health Information Technology Usability Evaluation Scale.

[b]Not applicable.

[c]*P*<.05.

Construct Validity

Exploratory Factor Analysis

EFA was performed to assess the construct validity of the Korean Health-ITUES items to extract potential factors. Results are reported in Table 4. The Kaiser-Meyer-Olkin measure of sampling adequacy (MSA) was 0.942, indicating that the data in this study were suitable for factor analysis. In addition, as a result of the Bartlett sphericity test, the correlation between the Korean Health-ITUES variables was recognized based on the significance level of .05 with x^2=3929.635 and *P*<.01. Thus, 4 subfactors were extracted.

Table 4. Principal axis factoring with varimax rotation.

Item	Commonality	Component			
		Perceived usefulness	Perceived ease of use	Impact	User control
usefulness8	0.805	*0.831*[a]	0.208	0.149	0.223
usefulness9	0.764	*0.811*	0.197	0.209	0.154
usefulness7	0.677	*0.747*	0.212	0.114	0.249
usefulness3	0.694	*0.735*	0.222	0.306	0.105
usefulness4	0.659	*0.722*	0.143	0.316	0.137
usefulness2	0.663	*0.700*	0.175	0.318	0.201
usefulness1	0.728	*0.687*	0.234	0.401	0.200
usefulness6	0.701	*0.678*	0.323	0.314	0.196
easeofuse1	0.715	0.574	*0.537*	0.183	0.251
usefulness5	0.539	*0.416*	0.162	0.410	0.414
easeofuse2	0.888	0.220	*0.892*	0.161	0.136
easeofuse3	0.901	0.251	*0.863*	0.197	0.233
easeofuse4	0.904	0.314	*0.856*	0.171	0.209
easeofuse5	0.597	0.177	*0.579*	0.080	0.474
impact1	0.789	0.303	0.124	*0.798*	0.065
impact2	0.748	0.309	0.119	*0.798*	0.205
impact3	0.708	0.374	0.284	*0.687*	0.126
control1	0.805	0.204	0.115	0.096	*0.861*
control2	0.765	0.214	0.429	0.062	*0.729*
control3	0.692	0.266	0.255	0.288	*0.688*

[a]Italics indicate the number of items corresponding to the component.

Discussion

Principal Findings

A major challenge to technology development is ensuring the usability of the tools [1-3]. However, many mHealth tools are currently available with little attention to their usability [5]. Technologies produced with poor design and inadequate consideration of the needs of their intended users are often difficult to use, and the consumers often cannot accomplish their goals, as a result [6]. Therefore, usability has been widely recognized as an important factor in the development of technology [7].

To address the need for understanding the usability of technology, various studies have been conducted [29,30]. However, each assessment tool could not be easily used, because it had too many questions, took a long time to answer due to difficult questions, was developed for specific users, or was developed only in English [23,29,30,45]. Therefore, the Health-ITUES, a simple and verified tool for multiple populations, was chosen for translation and validation. Because the Health-ITUEM, which is the basis of the Health-ITUES, was developed based on several usability models with strong reliability and validity, including the technology acceptance model (TAM) and ISO9241-11 [46-48]. Additionally, the Health-ITUES was validated in a sample of nurses and community-dwelling adults with HIV [33-35].

This study sought to not only translate but also verify the Health-ITUES considering the Korean context. In step 1, the authors modified the Health-ITUES that was used to evaluate an app for adults with HIV to an app for women suffering from menstrual discomfort. In Korea, there are fewer HIV-infected people compared to the United States. Additionally, because of negative views regarding homosexual contact, HIV-infected people are more likely to not disclose themselves [49]. As of October 2021, when searching for HIV/AIDS-related apps in Apple App Store and Google Play Store, it was difficult to find mHealth for patients with HIV in Korean. To obtain sufficient app users for the validation of the translated tool, the Health-ITUES was modified to enable usability evaluation of the menstrual app PINKDIARY for managing dysmenorrhea experienced by 75% of domestic women [38,50].

In step 2, the Health-ITUES was translated based on WHO guidelines [40]. These guidelines have been used to translate various instruments into many languages, such as Japanese and Arabic, and are not limited to Korean [51-53]. The process of translation and adaptation of instruments were as follows: forward translation, back-translation, and cognitive testing. In this study, 3 experts who majored in nursing or medical informatics and 1 professional English/Korean translator

participated in the forward translation and back-translation, respectively. The 2 translators independently translated and reconciled any discrepancies after the forward translation following WHO guidelines. This study was further strengthened by the review of the Korean Health-ITUES by 1 of the authors, who validated the Health-ITUES for mHealth technology. This process ensured that the original meanings of the Health-ITUES items were retained.

Another strength of this reliability and validity is our study sample size. The sample size exceeded the number of items, 20 × 10, and the participants live in various regions in Korea. Their field of work (major) is also not biased, so it can be said that geographical biases are small.

Moreover, in step 3, internal scale correlation ranged from 0.45 to 0.71, which indicates moderate to strong correlations. Notably, perceived usefulness was more highly correlated with impact than the other factors. The results mirrored the findings from the validation study using the Health-ITUES in a sample of adults with HIV [34]. In EFA, in the case of values for *usefulness*, 5 items ("I think PINKDIARY provides a more equitable process for the self-management of my menstrual-related symptoms.") were included as 3 components because the values were so similar: 0.416, 0.410, and 0.414. As a result, we decided to keep this item with the first component *perceived usefulness* with which it is most closely conceptually aligned [54].

Limitations

One limitation of this study is that when recruiting survey participants, only the experience of using the menstruation-related mHealth app was checked and no restrictions were placed on the past period of use. Future research should recruit participants by suggesting clear past usage periods based on objective evidence. However, this study successfully translated the Health-ITUES from English into Korean and validated it, and this instrument can be used to evaluate the usability of mHealth apps. These findings will contribute to the systematic evaluation of the rapidly growing field of mHealth apps.

Conclusion

The Health-ITUES demonstrated reliability and validity for use in assessing mHealth apps' usability in young Korean women with menstrual discomfort. Given the strong psychometric properties of this tool in Korean and English and across 2 different health conditions, the Health-ITUES is a strong instrument for evaluating the usability of mHealth apps.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea, funded by the Ministry of Education (NRF-2018R1C1B5030802, NRF-2021R1C1C1004505). RS was supported by the National Institute of Nursing Research of the National Institutes of Health (award no. K24NR018621). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

All authors had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors were involved in the study concept and design. JL performed data acquisition and statistical analysis of the data, drafted the manuscript, and provided administrative, technical, and material support. RS performed data interpretation, critical revision of the manuscript for important intellectual content, and study supervision and gave the final approval of the version to be submitted. JL and RS obtained funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Korean version of the Health-ITUES. Health-ITUES: Health Information Technology Usability Evaluation Scale.
[PNG File , 70 KB - medinform_v10i1e28621_app1.png]

References

1. Folmer E, Bosch J. Architecting for usability: a survey. J Syst Softw 2004 Feb;70(1-2):61-78. [doi: 10.1016/s0164-1212(02)00159-0]
2. Bengtsson PO, Lassing NH, Bosch J, van Vliet H. Analyzing Software Architectures for Modifiability. Ronneby, Sweden; 2000. URL: https://www.diva-portal.org/smash/get/diva2:838104/FULLTEXT01.pdf [accessed 2021-03-08]
3. Holzinger A. Usability engineering methods for software developers. Commun ACM 2005 Jan;48(1):71-74. [doi: 10.1145/1039539.1039541]
4. Your Guide for Developing Usable and Useful Web Sites. URL: http://www.usability.gov/ [accessed 2022-01-03]
5. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012 May 02;17 Suppl 1(sup1):5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]

6. Maguire M. Methods to support human-centred design. Int J Hum-Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
7. Shackel B. Usability: context, framework, definition, design and evaluation. Interact Comput 2009 Dec;21(5-6):339-346. [doi: 10.1016/j.intcom.2009.04.007]
8. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
9. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec 13;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
10. Huzooree G, Kumar Khedo K, Joonas N. Pervasive mobile healthcare systems for chronic disease monitoring. Health Inform J 2019 Jun 02;25(2):267-291 [FREE Full text] [doi: 10.1177/1460458217704250] [Medline: 28464728]
11. Clement J. Number of mHealth Apps Available at Google Play from 1st Quarter 2015 to 2nd Quarter 2020. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2021-12-28]
12. Aitken B. The Growing Value of Digital Health. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2022-01-03]
13. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
14. Kuhns LM, Garofalo R, Hidalgo M, Hirshfield S, Pearson C, Bruce J, et al. A randomized controlled efficacy trial of an mHealth HIV prevention intervention for sexual minority young men: MyPEEPS mobile study protocol. BMC Public Health 2020 Jan 15;20(1):65 [FREE Full text] [doi: 10.1186/s12889-020-8180-4] [Medline: 31941475]
15. van der Merwe MN, Mosca R, Swanepoel DW, Glascoe FP, van der Linde J. Early detection of developmental delays in vulnerable children by community care workers using an mHealth tool. Early Child Dev Care 2018 Jun 04;189(5):855-866. [doi: 10.1080/03004430.2018.1480481]
16. Palmer M, Barnard S, Perel P, Free C. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. Cochrane Database Syst Rev 2018 Jun 22;6:CD012675 [FREE Full text] [doi: 10.1002/14651858.CD012675.pub2] [Medline: 29932455]
17. Ravn Jakobsen P, Hermann AP, Søndergaard J, Wiil UK, Clemensen J. Help at hand: women's experiences of using a mobile health application upon diagnosis of asymptomatic osteoporosis. SAGE Open Med 2018 Oct 29;6:2050312118807617 [FREE Full text] [doi: 10.1177/2050312118807617] [Medline: 30397471]
18. Beauchemin M, Gradilla M, Baik D, Cho H, Schnall R. A multi-step usability evaluation of a self-management app to support medication adherence in persons living with HIV. Int J Med Inform 2019 Feb;122:37-44 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.11.012] [Medline: 30623782]
19. Boer L, Bischoff E, van der Heijden M, Lucas P, Akkermans R, Vercoulen J, et al. A smart mobile health tool versus a paper action plan to support self-management of chronic obstructive pulmonary disease exacerbations: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 09;7(10):e14408 [FREE Full text] [doi: 10.2196/14408] [Medline: 31599729]
20. Cho H, Powell D, Pichon A, Kuhns LM, Garofalo R, Schnall R. Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation. Int J Med Inform 2019 Sep;129:366-373 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.07.010] [Medline: 31445278]
21. Grossman JT, Frumkin MR, Rodebaugh TL, Lenze EJ. mHealth assessment and intervention of depression and anxiety in older adults. Harv Rev Psychiatry 2020 Apr 20;28(3):203-214. [doi: 10.1097/hrp.0000000000000255]
22. Lee J, Kim J. Development and efficacy testing of a social network-based competitive application for weight loss. Telemed J E Health 2016 May;22(5):410-418. [doi: 10.1089/tmj.2015.0067] [Medline: 26540485]
23. Lee J, Kim J. Can menstrual health apps selected based on users' needs change health-related factors? A double-blind randomized controlled trial. J Am Med Inform Assoc 2019 Jul 01;26(7):655-666 [FREE Full text] [doi: 10.1093/jamia/ocz019] [Medline: 30946478]
24. Levy J. Of mobiles and menses: researching period tracking apps and issues of response-ability. Stud Home Comm Sci 2018 Jan 08;11(2):108-115. [doi: 10.1080/09737189.2017.1420400]
25. Moglia ML. Evaluation of smartphone menstrual cycle tracking applications using an adapted applications scoring system. Obstet Gynecol 2016;127(6):1153-1160. [doi: 10.1097/aog.0000000000001444]
26. Pray W. Nonprescription Product Therapeutics. Philadelphia, PA: Lippincott Williams & Wilkins; 2006.
27. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
28. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]

29. Faieta J, Hand BN, Schmeler M, Onate J, Digiovine C. Health app review tool: matching mobile apps to Alzheimer's populations (HART Match). J Rehabil Assist Technol Eng 2020 Sep 23;7:2055668320938604 [FREE Full text] [doi: 10.1177/2055668320938604] [Medline: 33014412]
30. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
31. Intralink. Digital Health South Korea Market Intelligence Report. 2019 Jun 01. URL: https://www.intralinkgroup.com/getmedia/3153c79b-463d-47c7-84e6-56848c98aab7/Intralink-ReportLife [accessed 2022-01-03]
32. Silver L. Smartphone Ownership is Growing Rapidly Around the World, but Not Always Equally. 2019 Feb. URL: https://tinyurl.com/ywpub5p4 [accessed 2022-01-03]
33. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
34. Schnall R, Cho H, Liu J. Health Information Technology Usability Evaluation Scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
35. Yen P, Sousa KH, Bakken S. Examining construct and predictive validity of the Health-IT Usability Evaluation Scale: confirmatory factor analysis and structural equation modeling results. J Am Med Inform Assoc 2014 Oct 01;21(e2):e241-e248 [FREE Full text] [doi: 10.1136/amiajnl-2013-001811] [Medline: 24567081]
36. Nunnally J. Psychometric Theory, 2nd edition. New York, NY: McGraw-Hill; 1978.
37. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
38. Quick F, Mohammad-Alizadeh-Charandabi S, Mirghafourvand M. Primary dysmenorrhea with and without premenstrual syndrome: variation in quality of life over menstrual phases. Qual Life Res 2019 Jan 12;28(1):99-107. [doi: 10.1007/s11136-018-1999-9] [Medline: 30209721]
39. Bilir E, Yıldız Ş, Yakın K, Ata B. The impact of dysmenorrhea and premenstrual syndrome on academic performance of college students, and their willingness to seek help. Turk J Obstet Gynecol 2020 Sep 2;17(3):196-201 [FREE Full text] [doi: 10.4274/tjod.galenos.2020.97266] [Medline: 33072424]
40. World Health Organization. Process of Translation and Adaptation of Instruments. URL: https://www.who.int/substance_abuse/research_tools/translation/en/ [accessed 2021-12-28]
41. Faulkner S, Trotter S. Data saturation. In: The International Encyclopedia of Communication Research Methods. Hoboken, NJ: John Wiley & Sons; 2017.
42. Plichta Kellar S, Kelvin E. Munro's Statistical Methods for Health Care Research, 6th Edition. Philadelphia, PA: Lippincott Williams & Wilkins; 2005:1.
43. Allen SS, McBride CM, Pirie PL. The shortened premenstrual assessment form. J Reprod Med 1991 Nov;36(11):769-772. [Medline: 1765953]
44. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
45. Lee J, Kim J. Method of app selection for healthcare providers based on consumer needs. Comput Inform Nurs 2018;36(1):45-54. [doi: 10.1097/cin.0000000000000399]
46. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
47. ISO. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs). URL: https://www.iso.org/standard/16883.html [accessed 2022-01-03]
48. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
49. Kim JM, Kim NJ, Choi JY, Chin BS. History of acquired immune deficiency syndrome in Korea. Infect Chemother 2020 Jun;52(2):234-244 [FREE Full text] [doi: 10.3947/ic.2020.52.2.234] [Medline: 32618149]
50. Lee EH, Kim JI, Kim HW, Lee HK, Lee SH, Kang NM, et al. Dysmenorrhea and menstrual attitudes in adult women. Korean J Women Health Nurs 2003;9(2):105. [doi: 10.4069/kjwhn.2003.9.2.105]
51. Cho E, Choi M, Kim E, Yoo IY, Lee N. Construct validity and reliability of the Korean version of the practice environment scale of nursing work index for Korean nurses. J Korean Acad Nurs 2011 Jun;41(3):325-332. [doi: 10.4040/jkan.2011.41.3.325] [Medline: 21804341]
52. Toma G, Guetterman TC, Yaqub T, Talaat N, Fetters MD. A systematic approach for accurate translation of instruments: experience with translating the Connor–Davidson Resilience Scale into Arabic. Meth Innov 2017 Dec 21;10(3):205979911774140. [doi: 10.1177/2059799117741406]
53. Kurosawa S, Matsushima M, Fujinuma Y, Hayashi D, Noro I, Kanaya T, et al. Two principal components, coping and independence, comprise patient enablement in Japan: cross sectional study in Tohoku area. Tohoku J Exp Med 2012 Jun;227(2):97-104 [FREE Full text] [doi: 10.1620/tjem.227.97] [Medline: 22688526]

54. Hair JF, Black WC, Babin B, Anderson RE. Multivariate Data Analysis: A Global Perspective, 7th Edition. Upper Saddle River, NJ: Prentice Hall; 2009.

Abbreviations

EFA: exploratory factor analysis
Health-ITUEM: Health Information Technology Usability Evaluation Model
Health-ITUES: Health Information Technology Usability Evaluation Scale
mHealth: mobile health
MSA: measure of sampling adequacy
PMS: premenstrual syndrome
TAM: technology acceptance model
WHO: World Health Organization

Edited by G Eysenbach; submitted 08.03.21; peer-reviewed by D Yoon, J Mitchell, H Ranjani; comments to author 30.03.21; revised version received 28.04.21; accepted 30.11.21; published 24.01.22.

Please cite as:
Lee J, Schnall R
Validity and Reliability of the Korean Version of the Health Information Technology Usability Evaluation Scale: Psychometric Evaluation
JMIR Med Inform 2022;10(1):e28621
URL: https://medinform.jmir.org/2022/1/e28621
doi:10.2196/28621
PMID:35072630

Original Paper

Real-world Health Data and Precision for the Diagnosis of Acute Kidney Injury, Acute-on-Chronic Kidney Disease, and Chronic Kidney Disease: Observational Study

Karen Triep[1*], Dr med; Alexander Benedikt Leichtle[2*], Prof Dr; Martin Meister[3*], MA; Georg Martin Fiedler[4*], Prof Dr; Olga Endrich[2*], Dr med

[1]Medical Directorate, Medizincontrolling, Inselspital, University Hospital Bern, Insel Gruppe, Bern, Switzerland
[2]Insel Data Science Center, Inselspital, University Hospital Bern, Insel Gruppe, Bern, Switzerland
[3]Directorate of Technology and Innovation, Inselspital, University Hospital Bern, Insel Gruppe, Bern, Switzerland
[4]University Institute of Clinical Chemistry, Inselspital, University Hospital Bern, Insel Gruppe, Bern, Switzerland
[*]all authors contributed equally

Corresponding Author:
Karen Triep, Dr med
Medical Directorate, Medizincontrolling
Inselspital, University Hospital Bern
Insel Gruppe
S96 C223
Schwarztorstrasse 96
Bern, 3010
Switzerland
Phone: 41 31 632 51 96
Email: karen.triep@insel.ch

Abstract

Background: The criteria for the diagnosis of kidney disease outlined in the Kidney Disease: Improving Global Outcomes guidelines are based on a patient's current, historical, and baseline data. The diagnosis of acute kidney injury, chronic kidney disease, and acute-on-chronic kidney disease requires previous measurements of creatinine, back-calculation, and the interpretation of several laboratory values over a certain period. Diagnoses may be hindered by unclear definitions of the individual creatinine baseline and rough ranges of normal values that are set without adjusting for age, ethnicity, comorbidities, and treatment. The classification of correct diagnoses and sufficient staging improves coding, data quality, reimbursement, the choice of therapeutic approach, and a patient's outcome.

Objective: In this study, we aim to apply a data-driven approach to assign diagnoses of acute, chronic, and acute-on-chronic kidney diseases with the help of a complex rule engine.

Methods: Real-time and retrospective data from the hospital's clinical data warehouse of inpatient and outpatient cases treated between 2014 and 2019 were used. Delta serum creatinine, baseline values, and admission and discharge data were analyzed. A Kidney Disease: Improving Global Outcomes–based SQL algorithm applied specific diagnosis-based International Classification of Diseases (ICD) codes to inpatient stays. Text mining on discharge documentation was also conducted to measure the effects on diagnosis.

Results: We show that this approach yielded an increased number of diagnoses (4491 cases in 2014 vs 11,124 cases of ICD-coded kidney disease and injury in 2019) and higher precision in documentation and coding. The percentage of unspecific ICD N19-coded diagnoses of N19 codes generated dropped from 19.71% (1544/7833) in 2016 to 4.38% (416/9501) in 2019. The percentage of specific ICD N18-coded diagnoses of N19 codes generated increased from 50.1% (3924/7833) in 2016 to 62.04% (5894/9501) in 2019.

Conclusions: Our data-driven method supports the process and reliability of diagnosis and staging and improves the quality of documentation and data. Measuring patient outcomes will be the next step in this project.

(JMIR Med Inform 2022;10(1):e31356) doi:10.2196/31356

KEYWORDS

acute kidney injury; chronic kidney disease; acute-on-chronic; real-world health data; clinical decision support; KDIGO; ICD coding

Introduction

Background

Many definitions of diagnoses are rule-based and contain complex algorithms. This applies in particular to the diagnoses of kidney injury and kidney disease (KD). For example, the diagnosis of acute kidney injury (AKI) stage 3 according to the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines is defined as follows: an increase in serum creatinine (SCr) from under 4 mg/dL (353.6 µmol/L) to over 4 mg/dL within 7 days or an increase of SCr by 200% or more within 7 days. The increase and decrease have to be considered as follows: the gradient of increase versus the absolute increase, the increase versus decrease, and the highest stage. Moreover, the SCr baseline calculation has to be conducted: the lowest value during hospitalization or the arithmetic mean of all outpatient measurements before the index admission.

With the increasing availability of health data, automatic deducing of complex diagnoses has become possible. Correctly assigning diagnoses requires high precision, validity and reliability, the varying interrater reliability of diagnosis and the International Classification of Diseases (ICD) coding, affecting accuracy [1-4]. Interrater reliability shows insufficient values for certain diagnoses when comparing ICD codes or patients' records of the diagnoses of AKI and chronic KD (CKD) [5,6].

The global burden of KDs is high. Using a modification of the original glomerular filtration rate (GFR) estimating equation (the Chronic Kidney Disease Epidemiology Collaboration [CKD-EPI] equation), it was discovered that 11.6% of the adult residents in the United States have CKD stages 1-4, and its prevalence has increased over the past decade. Similar figures have been reported in several other countries [7-10]. In 2009, the US Renal Data System estimated that depending on the estimating equations used, the prevalence of CKD had increased by 20%-25% over the preceding decade [11].

The diagnoses of AKI, CKD, and acute-on-chronic KD are highly relevant as a comorbidity, intercurrent disease, or complication [12,13]. Inpatients with KD and kidney injury show a higher mortality and the staging implies an impact on outcomes [14,15]. The 2012 KDIGO Clinical Practice Guideline for AKI [16] and the Clinical Practice Guideline for the Evaluation and Management of CKD [17] offer guidelines containing definitions and classifications; ongoing areas of controversies and limitations of the evidence are also discussed in these documents. The definitions of AKI and CKD require a complex analysis of a patient's recent and historical laboratory values, a time-consuming process impeded by missing values and prone to errors if conducted manually. Misclassification impairs the choice of therapeutic approach, outcomes, high-quality documentation, data validity, and reimbursement. Moreover, an unclear definition of the individual creatinine baseline level and the approximate ranges of normal values without adjusting for age, ethnicity, comorbidities, and treatment aggravate the difficulties of diagnosis [18-27].

Clinical decision support systems can provide a systematic and objective way to enhance complex reasoning related to differential diagnostics. They can facilitate the process of diagnosis, contributing to its reliability [28-33]. Accumulating health data enables the providers to access relevant information for timely diagnosis, supporting effective management throughout care [10,34]. In recent times, national health systems, such as the National Health Service (NHS), have started supporting more advanced approaches for detecting patients with kidney injuries [35].

In Switzerland, since 2017, based on the official coding rules, AKI and CKD have been coded according to the KDIGO classification. However, documentation of the exact staging is often missing in the discharge documentation in many cases.

At our hospital (quaternary care university level), KD shows a rising relevance because the prevalence of patients with a GFR of <60 ml/min measured has been increasing during the recent years. Moreover, the ICD diagnoses of KD are relevant for reimbursement. Nevertheless, many inpatient cases with a GFR of <60 ml/min were not ICD-coded for any KD, and a clinical decision support has not yet been implemented.

Objectives

This study aims to evaluate a novel data-driven method to assign highly specific diagnoses of AKI and CKD by extracting historical and real-time data from the hospital's data warehouse. We hypothesize that by using a data-driven approach of diagnosis on routinely collected laboratory values, we can improve the detection and precision of the diagnosis and staging of AKI and CKD.

Methods

Study Population and Setting

Administrative and laboratory data of all inpatient and outpatient cases were used (Inselspital University Hospital Bern, 2014-2019; all Insel Gruppe, Bern, 2016-2019 with 200,000 inpatient and outpatient cases per year, of which, approximately 62,000 inpatient cases had ICD-coded diagnoses). Data from 2014 to 2016 were used for benchmarking purposes as a baseline at the start of the study (2017). Test data sets of cases from 2016 were used to evaluate the accuracy of the algorithm. The data for measuring the impact were selected from 2017 to 2019.

Definition of AKI

According to the KDIGO Clinical Practice Guidelines (2012) for AKI [16], we defined and staged AKI as follows (plasma creatinine instead of SCr):

- Stage 3: increase of SCr from under 4 mg/dL (353.6 µmol/L) to over 4 mg/dL within 7 days
- Stage 3: increase of SCr by 200% or more within 7 days

- Stage 2: increase of SCr by 100%-200% within 7 days
- Stage 1: increase of SCr by 50%-100% within 7 days
- Stage 1: increase of SCr by 0.3 mg/dL (26.52 µmol/L) within 48 hours

The decrease in SCr to baseline levels after starting the in-hospital measurement was interpreted as suggested by the KDIGO guidelines. With several positive findings, the gradient of the increase versus the absolute increase, the increase versus decrease, and the highest stage were prioritized for applying the specific stage. All available SCr measurements, along with date and time stamps were used. Inpatients with no available SCr measurements were classified as not having AKI.

Oliguria

Oliguria is still a controversial diagnostic criterion with regard to definition and practice of measurement, especially outside the intensive care setting. The hourly urine output data required to determine oliguria within any 6-, 12-, or 24-hour window is not reliably captured in the non–intensive care unit setting [25]. Therefore, we did not include it in the AKI definition of this study, which was consistent with the NHS England National Patient Safety Alert [35,36].

Baseline Definition

The baseline estimation was not specifically defined by the KDIGO guidelines; however, several methods were compared for baseline estimation [16,19,21,24,25,37].

We defined the baseline value for AKI as either the lowest value during hospitalization or the arithmetic mean of all outpatient SCr measurements 90 days before the index admission, if available, and took the lowest values for diagnosis. Either one may reasonably reflect the patient's premorbid baseline. Using the values at admission was considered; however, although the values may be the lowest for community-acquired AKI, they may be missing. Different approaches were not compared in this study.

Definition of CKD

In this study, a possible CKD was defined according to the KDIGO Clinical Practice Guideline for the Evaluation and Management (2012) of CKD [17], that is, a decreased GFR of <60 mL/min/1.73 m^2, an albumin creatinine ratio (ACR) of >30 mg/g (>3 mg/mmol), or a history of kidney transplantation (estimated according to the CKD-EPI equation). GFR categories were assigned as follows:

- Stage 5: all values under 15 mL/min/1.73 m^2 for >91 days
- Stage 4: all values under 30 mL/min/1.73 m^2 for >91 days
- Stage 3: all values under 60 mL/min/1.73 m^2 for >91 days
- Stage 2: all values under 90 mL/min/1.73 m^2 for >91 days

Definition of ACR

In addition, according to the KDIGO criteria we integrated albuminuria into the model as a marker of kidney damage, related to mortality and kidney outcome in CKD [17,37], using the ACR values, as follows:

- Severely increased: SCr (µmol/L)/albumin (g/L) >30 mg/g
- Moderately increased: SCr (µmol/L)/albumin (g/L) between 3 and 30 mg/g
- Normal to mildly increased: SCr (µmol/L)/albumin (g/L) <30 mg/g

Values from one sample or values measured within an interval of 30 days were considered.

Architecture and Algorithm

A complex dataflow was established to make all the required variables available for calculation. First, an SQL-based algorithm processed the data warehouse's data (rule engine and HL7 [Health Level Seven International] messages) and detected the potential cases of KD. All available SCr measurements with date and time stamps were used. The patient identification number was defined as the primary key but was only used as a linkage code for administrative and clinical or laboratory data. Patient- and case-related laboratory and administrative historical and real-time data had to be extracted from the source systems, merged, and computed for diagnosis and stage. Second, the output of the correlating ICD [38,39] code was connected by the detection date to the distinct date of the patient's inpatient case (the case ID linked to the entry and discharge dates related to the patient ID). Third, the test results were processed to the recipient and included a staging of AKI and CKD according to the abovementioned criteria. The architecture and dataflow of AKI and CKD are illustrated in Figure 1, and the architecture and dataflow of the retrospective calculation are shown in Figure 2.

Figure 1. Architecture and dataflow. AKI: acute kidney injury; CKD: chronic kidney disease; ICD: International Classification of Diseases.

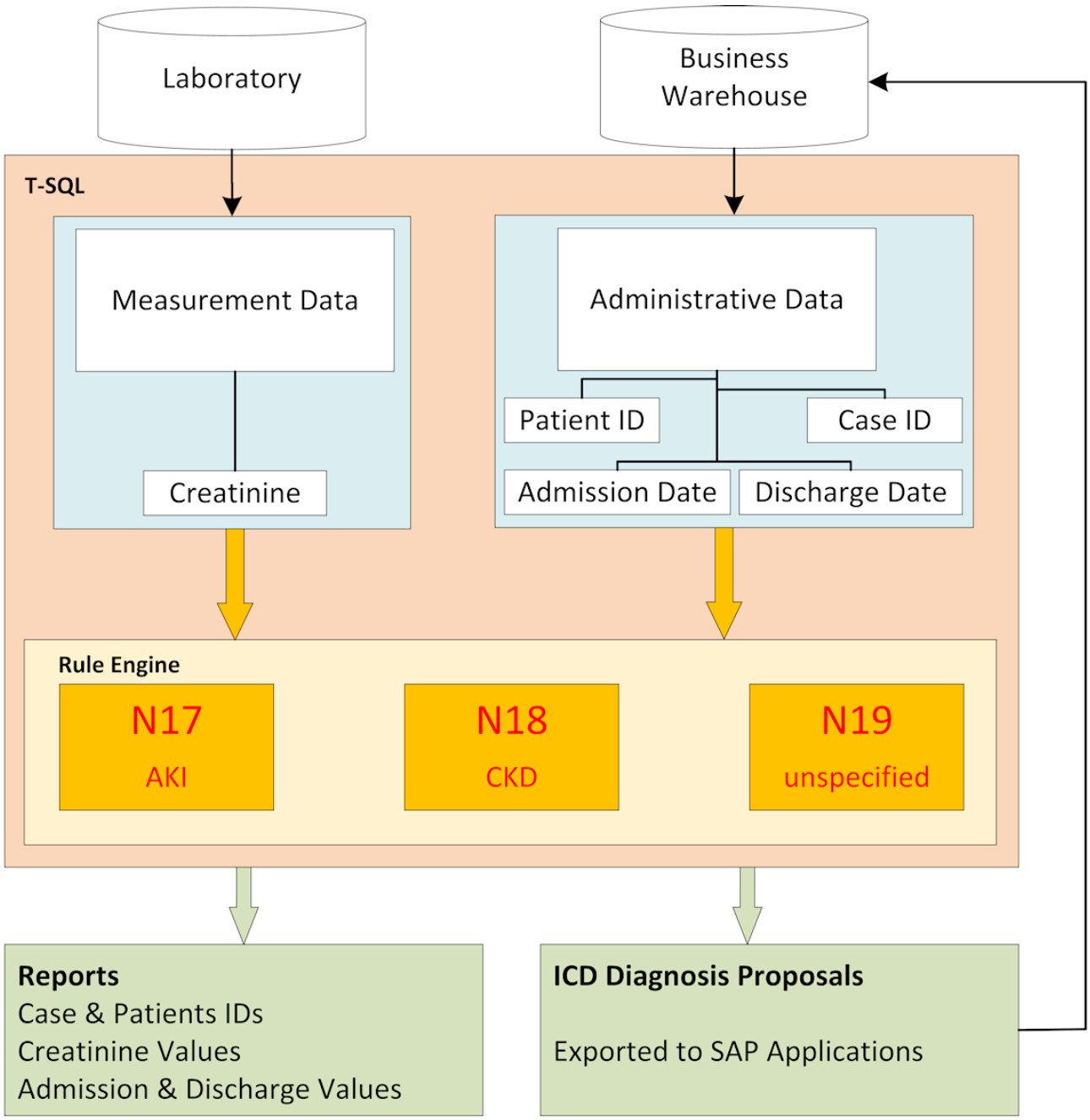
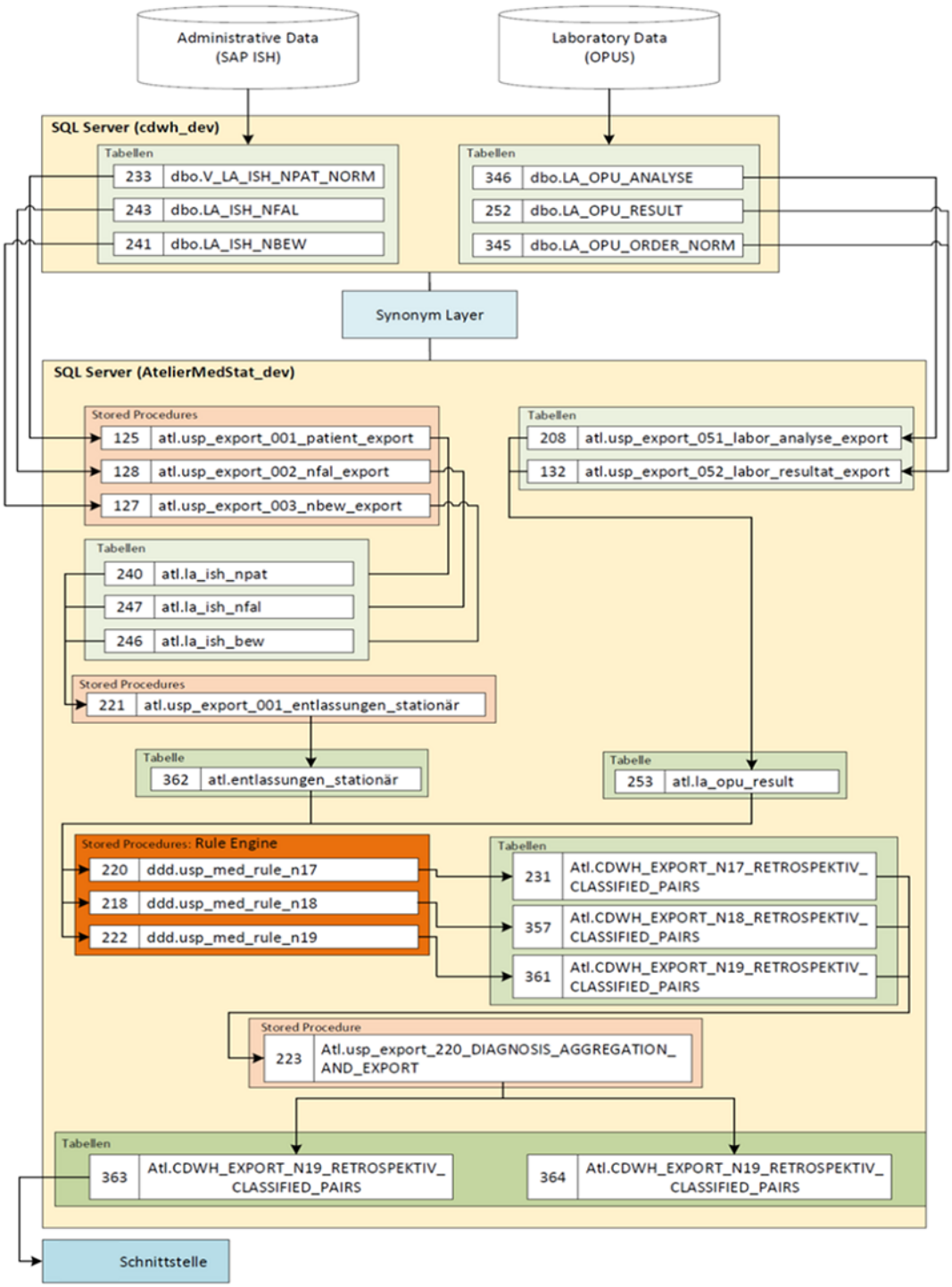


Figure 2. Architecture for retrospective analysis. OPUS: laboratory information system by OSM Group; SAP ISH: Systems Applications and Products in Data Processing Industry Solution Healthcare.

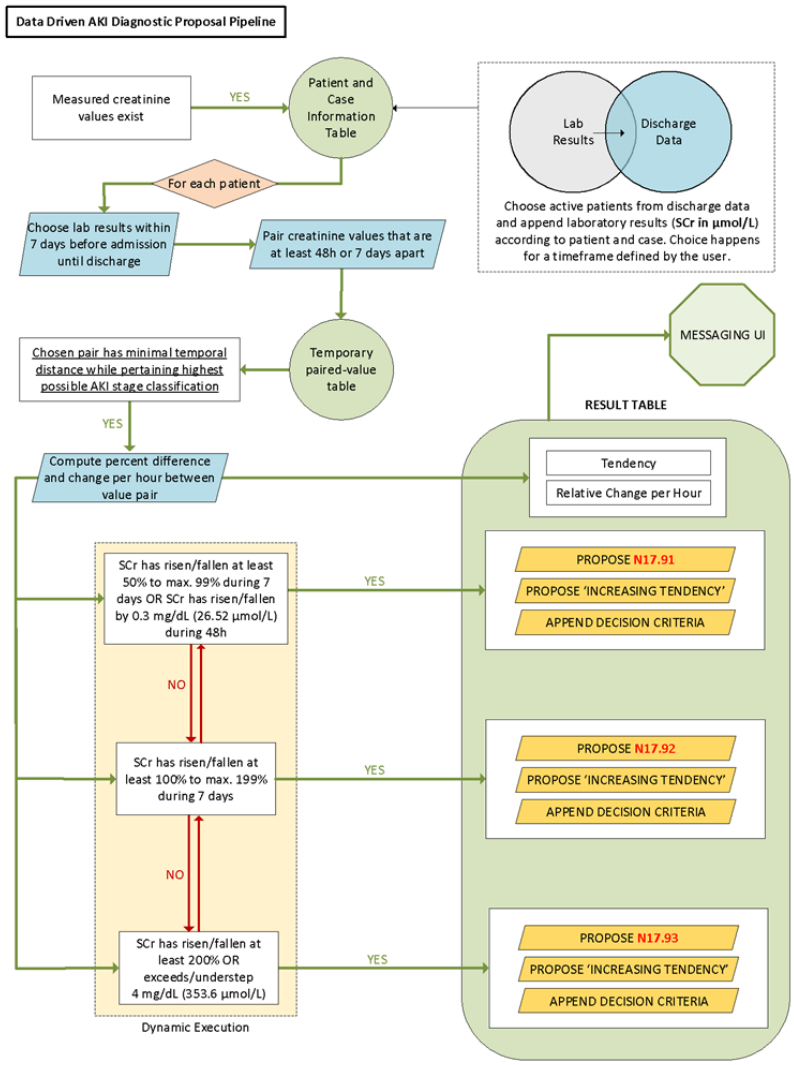


The steps of computation were as follows: (1) for AKI, selecting inpatients hospitalized during a specified period, selecting laboratory values (SCr) 7 days before admission until discharge, mapping values from 48 hours to 7 days apart, and classifying values according to the ICD standard [38,39]; and (2) for CKD, selecting inpatients' or outpatients' laboratory values (estimated GFR [eGFR]), mapping the values of eGFR at least three months apart, calculating the mean minimum and maximum values of each period and the difference of the mapped values in hours, and classifying values according to the ICD standard.

The output for AKI was defined as the highest stage of diagnosis with the shortest period between the mapped values. The relevant result for CKD was defined as the highest stage of diagnosis with the longest period between mapped values. All patients with fulfilled criteria during the previous year for the specific diagnosis and with values positively corresponding to

diagnosis during the last 3 months were detected. The algorithm for AKI is presented in Figure 3 and for CKD, in Multimedia Appendix 1. The algorithm was tested on testing data sets by technicians and clinicians. The algorithm was technically adjusted until all tested cases showed correct diagnoses and stages according to the formal definitions provided.

Figure 3. Algorithm of diagnosis of AKI. AKI: acute kidney injury; SCr: serum creatinine; UI: user interface.



Text Mining

A text mining pipeline was implemented using Apache Solr (The Apache Software Foundation) to compare the results of the algorithms with those of the reports. In this process, all relevant reports were loaded into a Solr collection and searched for terms, such as *AKI*, *CKD*, *eGFR*, *KDIGO*, *creatinine*, and *renal failure* (German translation: *kreatinin*, *niereninsuffizienz*, and *nierenversagen*), and the exact KDIGO staging (eg, G1A1 [GFR ≥90 ml/min per 1.73 m^2 ACR <30 mg/g (<3mg/mmol)]).

For terms with a tilde character ("~"), a fuzzy search algorithm was applied to ensure that not only one spelling of a term was found. The Damerau-Levenshtein distance algorithm was used for this purpose. A separate CSV file was generated for each search term, including the case ID, date of report generation, and report type. Each row corresponded to a finding of the respective search term.

The CSV files were then loaded into the database containing the algorithm results and other case data. Using a transact-SQL script, the results from all sources were then aggregated at the case level. On this basis, the cases could be filtered and evaluated for constellations of interest.

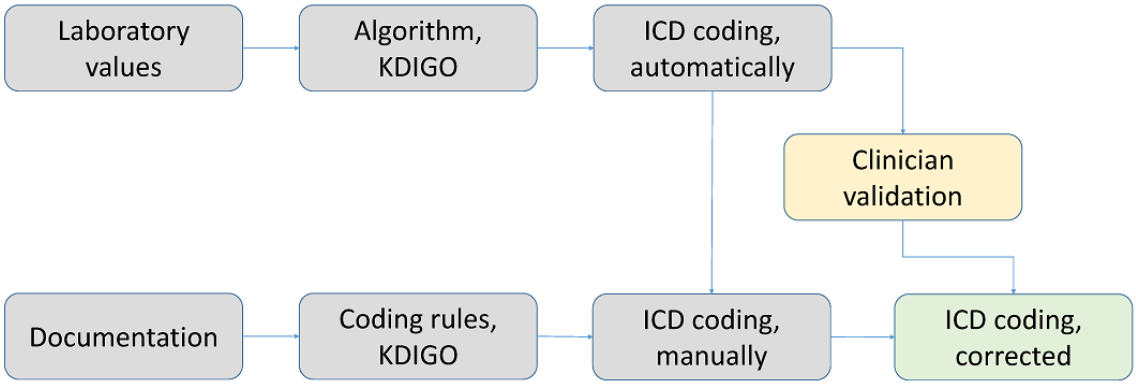
Process of Diagnosis

Being aware of the purely arithmetic method of diagnosis implemented in 2017, clinical judgment was integrated into the process, especially to verify the chronic diagnoses and distinguish between AKI and unstable CKD [40,41]. Therefore, the real-time information and retrospectively detected diagnoses were compared with the documentation in the patients' health records using ICD diagnoses coded manually (comparison of automatically generated and manually coded ICD codes) and text mining (diagnoses and stages). Differences were then analyzed by the clinicians. Mostly, the cases with singular ICD codes relevant for reimbursement were analyzed and validated. After rejecting a certain fraction of the automatically generated diagnoses because of the clinician's judgment or lack of documentation, the corresponding codes were deleted. Only validated diagnoses were retained in the database. The effect

of the validation was monitored by analyzing the mutation of diagnoses from generated to coded ICD codes (log file of all ICD code mutations).

During the period of the project, the process of diagnosis was supported by documentation templates, instruction, and close communication with the clinicians. The data processed were not used to set up an alerting system but were validated retrospectively after the patients' discharge. The diagnosis process is illustrated in Figure 4.

Figure 4. Process of diagnosis. ICD: International Classification of Diseases; KDIGO: Kidney Disease: Improving Global Outcomes.



Catalogs

Bound by the Swiss regulations, the following catalogs were applied [38,39]: for the discharge year 2014, the International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification (ICD-10-GM) 2012; from 2015 to 2016, the ICD-10-GM 2014; from 2017 to 2018, the ICD-10-GM 2016; and for 2019, the ICD-10-GM 2018. Because of a mutation in codes for AKI (ICD N17-) from 2014 to 2016, some analyses could only be conducted for the data from 2017 onward (Multimedia Appendix 2); ICD-10-GM codes catalogs from 2012-2018, effective in Switzerland from 2014 to 2020.

Reimbursement

To measure the effect after the successful implementation of the algorithm we planned a simulation of Swiss Diagnosis Related Groups, Inpatient Tariff (SwissDRG) income of 6 months' coding (inpatient cases from February 1, 2020, to July 31, 2020) in 2020 with and without grouping the automatically calculated ICD diagnoses; the SwissDRG web-based batch grouper version 9.0 2020/2020 was used.

Analysis and Software

The automatically generated, previously coded, rejected, and validated ICD diagnoses were compared per code category and per specific code. The prevalence of the codes was calculated for all inpatient cases and for inpatient cases with coded KD (all diagnoses). The proportion of specific medical information (text, laboratory values, reference to KDIGO classification, and formal KDIGO staging) documented in the corresponding discharge letters was calculated.

The following software were used during analyses: Medical coding software Systems Applications and Products in Data Processing Industry Solution Healthcare (SAP IS-H), Medical Coding Tool ID Diacos, Clinical Data Phoenix CGM, Business Data Ware House SAP BW, Microsoft Excel 2010, R developing software (R version 3.5.0 2018-4-23), RStudio version 1.1.453, and RStudio Team (2016) as well as RStudio: Integrated Development for R (RStudio, Inc) and ggplot2 version 3.1.0.

Ethics

The ethics committee of the Canton Bern approved this study (BASEC-Req-2018-01184).

Results

General Remarks

The method applied in this study to assign the specific diagnoses and exact stages of AKI and CKD produced highly reliable results. Moreover, the process of communicating and verifying the diagnoses improved the validity in the medical context of the individual patient. Diagnoses and stages could be displayed in near to real time and retrospective calculations could be conducted for the previous 6 years. As the algorithm considered acute and chronic diseases, this project is one of the few to integrate the diagnosis of acute-on-chronic KD. The specific diagnoses documentation and the exact staging in the patients' discharge letters could be improved.

Overview

An increasing prevalence of inpatient cases with a measured eGFR of <60 ml/min can be shown for the discharge years 2014-2019 (from 4362/42,703, 10.21% cases in 2014 to 12,519/66,958, 18.69% cases in 2019). The proportion of ICD-coded inpatient cases with ICD codes for any KD diagnosis in the ICD categories N17-/N18-/N19- during the same period increased for all inpatients (from 4491/42,703, 10.52% cases in 2014 to 11,124/66,958, 16.62% cases in 2019) and for the group of cases with an eGFR of <60 ml/min (from 2167/4362, 49.48% cases in 2014 to 7596/12,519, 60.68% in 2019). The proportion of coded cases of KD with an eGFR of <60 ml/min was 49.68% (2167/4362) in 2014 and 45.09% (5005/11,100)

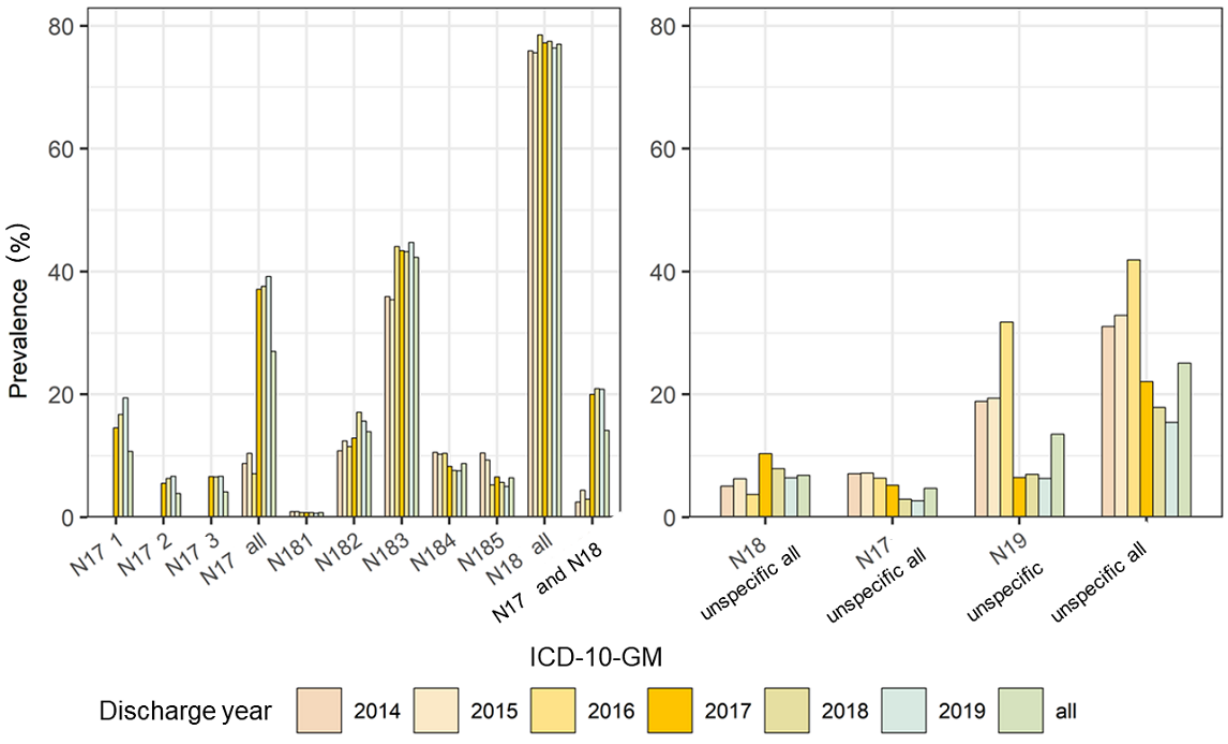
in 2016 and dropped to 60.68% (7596/12,519) in 2019. Between 2014 and 2019, the prevalence of all KD-coded cases increased from 10.52% (4491/42,703) cases in 2014 to 16.61% 11,124/66,958) cases in 2019. The main increase in the prevalence of coded cases of KD was observed between 2017 and 2019, after project initiation as shown in Table 1.

Table 1. Prevalence of cases with estimated glomerular filtration rates (eGFRs) of <60 ml/min and kidney injury (KI) coding (all International Classification of Diseases [ICD] codes N17-/N18-/N19-).

Distribution	Year of discharge						
	2014	2015	2016	2017	2018	2019	Total
Inpatient cases, N	42,703	45,138	64,478	65,146	66,038	66,958	350,461
Inpatient cases with measured eGFR, n (% of inpatient cases total)	20,610 (48.26)	37,326 (82.69)	40,917 (63.46)	40,109 (61.57)	41,552 (6292)	46,800 (69.89)	227,314 (64.86)
Inpatient cases with an eGFR of <60 ml/min, n (% of inpatient cases total)	4362 (10.21)	8786 (19.47)	11,100 (17.22)	10,695 (16.42)	10,570 (16.01)	12,519 (18.67)	58,032 (16.65)
Any KI-coded (ICD N17-/N18-/N19-) cases, n (%)	4491 (10.51)	4786 (10.6)	8422 (13.06)	8512 (13.06)	10,165 (15.39)	11,124 (16.61)	47,500 (13.55)
Any KI-coded (ICD N17-/N18-/N19-) inpatient cases with an eGFR of <60 ml/min, n (%)	2167 (49.68)	4029 (45.86)	5005 (45.09)	5031 (47.04)	5983 (56.6)	7596 (60.68)	29,811 (51.37)

Figure 5 highlights an increase in cases with specifically coded diagnosis (ICD codes N17-/N18- with staging) and acute-on-chronic KD (ICD codes N17-/N18-) in 2014-2019 of all KD-coded cases, for example, acute-on-chronic KD cases increased from 0.26% (111/42,703) cases in 2014 and 2.62% (1706/65,146) cases in 2017 to 3.47% (2320/66,958) cases in 2019 (Multimedia Appendices 3 and 4).

Figure 5. Proportion of specific ICD KD codes in all cases coded with any ICD code for KD (KD-coded cases). A: data for specific codes; B: data for unspecific codes. ICD: International Classification of Diseases; ICD-10-GM: International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification; KD: kidney disease.



Correspondingly, unspecified diagnoses (ICD N19) decreased from 2014 to 2019 (see all prevalence data in Multimedia Appendices 3 and 4 and Figure 5). A sharp decrease can be observed among 2016, 2017 (onset of the project), and 2019 in the proportion of all unspecified diagnoses (all N17-, N18-, and N19-, without staging) of all KD-coded cases, that is, 41.91% (3530/8422) cases in 2016, 22.1% (1881/8512) cases in 2017, and 15.46% (1720/11,124) cases in 2019 (Multimedia Appendices 3 and 4 and Figure 5).

Moreover, the mutation of unspecified diagnoses (ie, ICD N19-) to more precise coding (ICD N17- for AKI and N18-for CKD, including stages) during the process of diagnosis of individual cases can be demonstrated, for example, the conversion of ICD N19- to more specific codes, rising in 2017. Tables 2 and 3 illustrate the impact on the prevalence of ICD N19-.

Table 2. Impact of manual validation—conversion of unspecific N19 codes to specific codes and the rejection of any International Classification of Diseases coding.

Discharge year	N19 codes generated, N	N17 codes from the N19 codes generated, n (%)	N18 codes from the N19 codes generated, n (%)	N19 codes from the N19 codes generated, n (%)
2014	2967	188 (6.34)	1705 (57.47)	357 (12.03)
2015	6279	475 (7.56)	3126 (49.78)	684 (10.89)
2016	7833	460 (5.87)	3924 (50.10)	1544 (19.71)
2017	7605	2027 (26.65)	3831 (50.37)	301 (3.96)
2018	7560	2420 (32.01)	4571 (60.46)	467 (6.18)
2019	9501	3204 (33.72)	5894 (62.04)	416 (4.38)

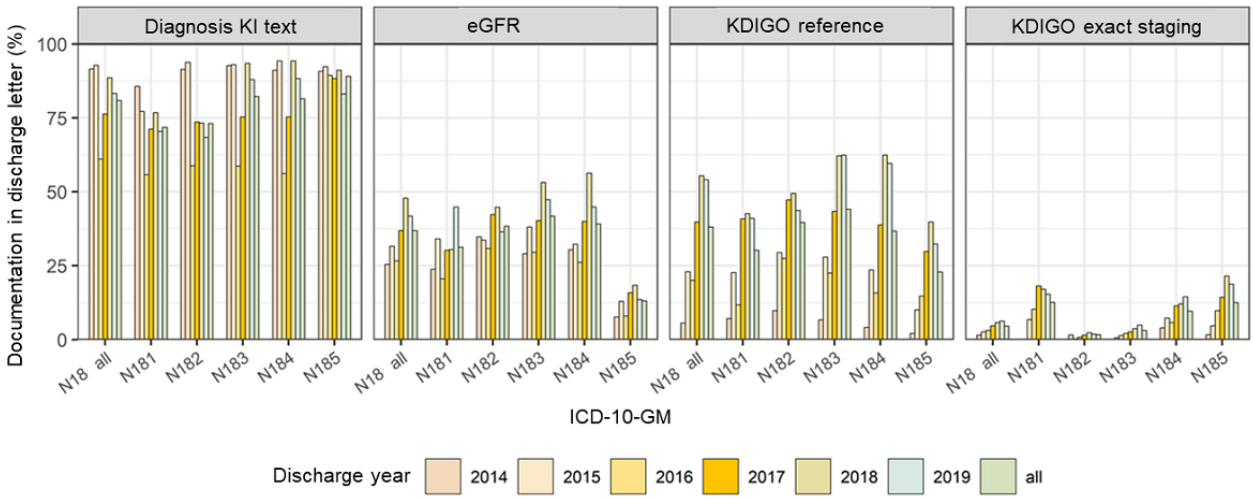
Table 3. Prevalence of cases with estimated glomerular filtration rates (eGFR) of <60 ml/min and unspecified kidney injury (KI) coding (International Classification of Diseases [ICD] N19) compared with those from 2016.

Year of discharge	Cases with an eGFR of <60 ml/min, N	KI-coded (ICD N19-) cases with an eGFR of <60 ml/min, n (%)
2016	11,100	1544 (13.91)
2017[a]	10,695	301 (2.81)
2018	10,570	467 (4.42)
2019	12,519	416 (3.32)

[a]Start of the project.

Regarding discharge documentation, we observed an increase in the proportion of documented diagnoses for some KD code categories but mostly an increase in references to the KDIGO classification mentioning eGFR and SCr. Concerning the cases with coded CKD, the correct KDIGO staging could be detected more often, with all ICD N18- coded cases being 1.5% in 2014, 4.7% in 2017, and 6.3% in 2019 (Figure 6 and Multimedia Appendix 5).

Figure 6. Proportion of positive text mining results in discharge letters of International Classification of Diseases N18–coded cases (chronic kidney disease). eGFR: estimated glomerular filtration rate; ICD-10-GM: International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification; KDIGO: Kidney Disease: Improving Global Outcomes; KI: kidney injury.



Regarding the diagnosis of acute-on-chronic KD, a drop in documentation of the textual diagnosis could be observed at the onset of the project. Nevertheless, the SCr, eGFR, and KDIGO references were documented more often (Figure 7 and Multimedia Appendix 6). The results for AKI are shown in Figure 8 and Multimedia Appendix 7.

Figure 7. Proportion of positive text mining results in discharge letters of International Classification of Diseases N17/N18–coded cases (acute-on-chronic kidney disease). eGFR: estimated glomerular filtration rate; ICD-10-GM: International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification; KDIGO: Kidney Disease: Improving Global Outcomes.

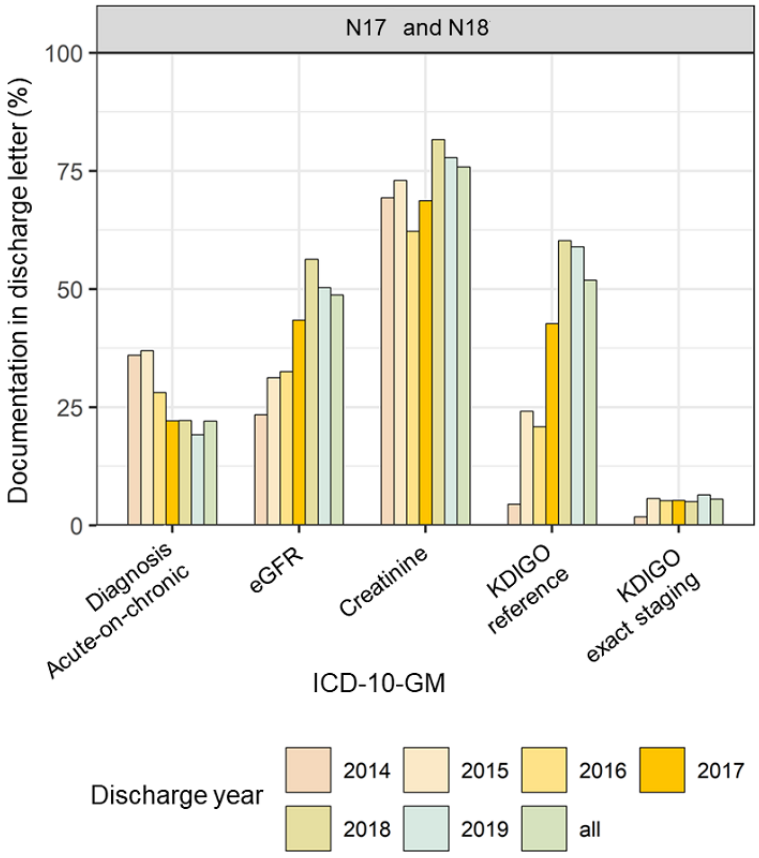
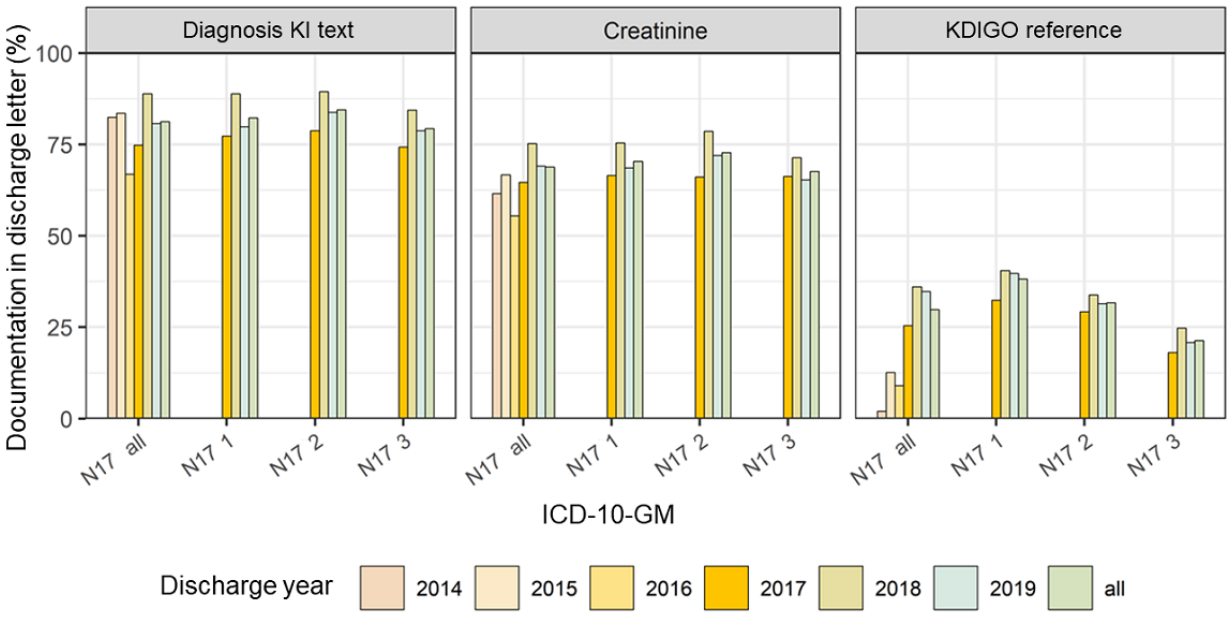


Figure 8. Proportion of positive text mining results in the discharge letters of International Classification of Diseases N17–coded cases (acute kidney injury). ICD-10-GM: International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification; KI: kidney injury; KDIGO: Kidney Disease: Improving Global Outcomes.



The effect on SwissDRG income after successful implementation of this approach accounted for a case-mix difference of 198.87 points analyzing the relevant inpatient cases (5877) from February 2020 to July 2020. Multiplied by the current standard base rate (CHF 10,800 [Swiss francs]; US $11,800), this results in CHF 2,147,753 (US $2,337,700) for this period (Table 4).

Table 4. Delta income of Swiss Diagnosis Related Groups, Inpatient Tariff from February 2020 to July 2020 owing to the automatization of the International Classification of Diseases (ICD) of kidney disease (KD).

Characteristics	Values
Period for inpatient cases	02/01/2020 to 07/31/2020
Cases, N	28,314
Case diagnoses[a], n (%)	5876 (20.75)
CM[b] with ICD, n	14,340.08
CM without ICD, n	14,141.21
Delta CM, n	198.87
Delta Swiss francs[c], CHF	2,147,752.80[d]

[a]Any ICD diagnosis of KD automatically generated and validated afterward.

[b]CM: casemix.

[c]Standard base rate USD $11,800.

[d]USD $2,337,700.

Discussion

Principal Findings

After introducing the algorithm to apply AKI or CKD diagnoses, we observed an increase in the number of ICD-coded diagnoses and a shift toward higher precision in the applied stages of the diseases. Correspondingly, the number of unspecifically coded diagnoses (ICD N19-) dropped. Moreover, the documentation also improved (the correct KDIGO staging of CKD for all ICD N18- coded cases was 1.5% in 2014 and 6.3% in 2019).

Strengths of the Project

Most studies concerning an algorithm to apply AKI or CKD diagnoses and stages consider only one diagnosis, either AKI or CKD [28,29,31]. As our project combines the 2 diagnostic criteria formulated by the KDIGO for both AKI and CKD, it improves the validity of diagnosis and enables the clinicians to easily recognize acute-on-chronic KD.

By referring to the same data set when testing for both AKI and CKD diagnoses, consistency could be improved.

This project established a link to the acknowledged impact on health at discharge by involving a defined process of validation that is conducted by text mining and communication with the clinician for retrospectively defining the exact diagnosis and staging. The process of validation of automatically generated diagnoses resulted in a decrease of unspecific diagnoses both in coding and documentation and therefore had a practical impact on the clinician's work and on the SwissDRG income.

Many projects conducted so far have been limited to outpatients, causing a bias when calculating the overall prevalence. In particular, severe stages associated with underlying morbidities treated in inpatient care might not be recognized [10,18,24,34,42,43]. The availability of inpatient and outpatient data from the previous 6 years stored in the Insel Data Platform for all patients offered the advantage of calculating the diagnoses and the disease stages for the inpatients.

Limitations

The validation of ICD diagnoses was strongly aimed at ICD codes with an impact on reimbursement, resulting in a bias toward validating cases with potentially higher income. Consecutively, the exact staging of diagnoses was limited to this group. Furthermore, the KDIGO classification of CKD grade 1 could not be considered for technical reasons (no limiting value of eGFR defined by KDIGO). Data on inpatient cases of all Insel Gruppe sites are available only for the years 2017-2019. Therefore, we could neither benchmark outpatient cases nor compare the data with the data of previous years. The study was limited to the description of the impact of the automatization of diagnosis; it was not designed to compare methods concerning the impact of the ACR or to determine the baseline SCr.

Algorithm and Precision of Diagnosis

As the data required contains only laboratory results, the time stamp of the taken samples, and a patient and case identifier, the algorithms for both AKI and CKD that are presented here are transferable and ready to use.

Moreover, the process of diagnosis is facilitated and staging as a time-consuming back-calculation can be automated instead. As the algorithm applies criteria of both diagnoses separately at a specific point of time for the same case, cases with the calculated diagnosis of acute-on-chronic KD can be easily extracted to evaluate an unstable CKD versus AKI in the clinical context.

A weakness of the algorithm caused by the classifications themselves lies in the definition of CKD grade 1 according to KDIGO and ICD N18.1. As no upper value of eGFR is set, the formal testing of the data produces no sensible results. The

diagnosis of CKD stage 1 can be defined only with an effective diagnosis." [17,38].

The results displayed show the impact on (1) the increasing number of diagnoses and (2) the increasing precision in staging, documentation, and ICD coding. The higher validity and precision of diagnosis will not only improve the quality of documentation and data but also specific and timely treatment when integrated into a decision support system [30,32,34]. As the findings are translated into ICD-10 codes within the algorithm and the data of diagnosis and stage are stored, as encoded by ICD, the algorithm and the data extracted support international benchmarking and quality control by standardized diagnoses.

Baseline

The absence of a shared approach to baseline SCr definitions [22,24,42,44] and an inter- and intraindividual and technical variability has resulted in a variability among centers regarding the interpretations for diagnosis and classification. The use of inpatient creatinine measurements as surrogates for baseline function resulted in misclassification, and the use of a minimum SCr value as a baseline inflated disease incidence [44]. In contrast to an imputed or minimum SCr value, use of the admission SCr value as a baseline resulted in nearly 50% reduction in the reported incidence of AKI compared with that of using a known outpatient baseline value [12,24,25]. This decrease is perhaps best explained by the missed diagnosis of community-acquired AKI that improves during hospitalization. The higher mortality rates observed when using this baseline reflect the bias of using this method, which is only sensitive to AKI that continues to worsen during hospitalization. Because of this lack of joint approach to baseline SCr definitions and lack of other markers, we specified the following to reflect the patients' premorbid state: either baseline SCr is the lowest value during hospitalization or the arithmetic mean of all outpatient SCr measurements 90 days before the index admission (relying on the lowest value of both methods for diagnosis) is the lowest value.

The comparability of studies concerning AKI, including this project, might be impaired regarding the baseline definition, a weakness that can only to be resolved by additional consensus criteria to better characterize preadmission AKI and by specifying a standard method to incorporate previously known baseline data. Being aware of the potential inflation of diagnosis when using the lowest inpatient SCr [44], the approach would ensure a higher sensitivity regarding the clinician's awareness.

Electronic Health Records

Integrating data and computer-based entries into electronic health records may support precision, standardization, and decision regarding patients' health care and lead to a more specific, valid, reliable, and consistent database. Greater data integration may also provide information not only for timely treatment but also disease registries and clinical trials [15,28,30,31,33,34]. Automated decision support based on arithmetic algorithms may be too rigid. Therefore, with the experience gained from this project, we favor an integrated solution that closes the loop between an automated alert and clinicians' validation. As demonstrated in this study, part of the validation can be automated by text mining to minimize the workload [30]. However, many cases require manual validation. Lack of documentation as seen in the inpatient cases of 2016 (fusion of data and documentation of all Insel Gruppe sites) compromises automatization.

Lessons Learned and Future Work

The project will be an important achievement for inpatient and outpatient care, especially with chronic diseases, such as CKD and acute-on-chronic KD and its complex algorithm. As the prevalence of KD is underestimated [7,8,11], the higher validity and precision of diagnosis will not only improve the reliability of documentation and data but will also improve treatment and reimbursement. This will result in the efficiency and quality of the diagnosis process, a higher reliability, and a highly standardized database.

The difficulty of defining the right baseline for AKI could not finally be solved. Missing values before admission should be addressed and anticipated. Clinicians and other medical experts should be closely involved in the process of setting up requirements and validating diagnoses.

This project introduced an end-to-end approach to clinical decision support at the Insel Gruppe hospitals.

Acknowledgments

The authors would like to thank Matthias Kämpf for supporting the project and for his technical and conceptual expertise; Reto Baumgartner, Nenad Torbica, and Roger Mathis for their technical support; team Medizincontrolling Insel Gruppe Bern for professional support; and Ben von Huenerbein for reviewing the manuscript.

This interdisciplinary initiative involved the university hospital's cardiovascular clinic, the Directorate of Technology and Innovation, the Medical Directorate, and the Data Science Center IDSC (Insel Data Science Center) of the Insel University Hospital Bern.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Algorithm for all diagnoses.

[PPTX File , 75 KB - medinform_v10i1e31356_app1.pptx]

Multimedia Appendix 2
International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification (ICD-10-GM) code catalogs from 2012 to 2018 that were effective in Switzerland from 2014 to 2020.
[DOCX File , 14 KB - medinform_v10i1e31356_app2.docx]

Multimedia Appendix 3
Counts of cases with N17 coded.
[DOCX File , 15 KB - medinform_v10i1e31356_app3.docx]

Multimedia Appendix 4
Counts of cases with N18, N19, acute-on-chronic, unspecified kidney injury, and kidney disease (KI/KD) coded.
[DOCX File , 14 KB - medinform_v10i1e31356_app4.docx]

Multimedia Appendix 5
Proportion of International Classification of Diseases (ICD)–coded cases N18- with documentation in the discharge letter.
[DOCX File , 16 KB - medinform_v10i1e31356_app5.docx]

Multimedia Appendix 6
Proportion of International Classification of Diseases (ICD)–coded cases N17- and 18- acute-on-chronic with documentation in discharge letter.
[DOCX File , 13 KB - medinform_v10i1e31356_app6.docx]

Multimedia Appendix 7
Proportion of International Classification of Diseases (ICD)–coded cases N17- with documentation in the discharge letter.
[DOCX File , 14 KB - medinform_v10i1e31356_app7.docx]

References

1. Entezarjou A, Bonamy AE, Benjaminsson S, Herman P, Midlöv P. Human- versus machine learning-based triage using digitalized patient histories in primary care: comparative study. JMIR Med Inform 2020 Sep 03;8(9):e18930 [FREE Full text] [doi: 10.2196/18930] [Medline: 32880578]
2. Kalatharan V, McArthur E, Nash DM, Welk B, Sarma S, Garg AX, et al. Diagnostic accuracy of administrative codes for autosomal dominant polycystic kidney disease in clinic patients with cystic kidney disease. Clin Kidney J 2021 Feb;14(2):612-616 [FREE Full text] [doi: 10.1093/ckj/sfz184] [Medline: 33623686]
3. Pellathy T, Saul M, Clermont G, Dubrawski AW, Pinsky MR, Hravnak M. Accuracy of identifying hospital acquired venous thromboembolism by administrative coding: implications for big data and machine learning research. J Clin Monit Comput 2021 Feb 08;10 (forthcoming). [doi: 10.1007/s10877-021-00664-6] [Medline: 33558981]
4. Cerullo E, Quinn TJ, McCleery J, Vounzoulaki E, Cooper NJ, Sutton AJ. Interrater agreement in dementia diagnosis: a systematic review and meta-analysis. Int J Geriatr Psychiatry 2021 Aug;36(8):1127-1147. [doi: 10.1002/gps.5499] [Medline: 33942363]
5. Waikar SS, Wald R, Chertow GM, Curhan GC, Winkelmayer WC, Liangos O, et al. Validity of International Classification of Diseases, Ninth Revision, clinical modification codes for acute renal failure. J Am Soc Nephrol 2006 Jun;17(6):1688-1694 [FREE Full text] [doi: 10.1681/ASN.2006010073] [Medline: 16641149]
6. Tomlinson LA, Riding AM, Payne RA, Abel GA, Tomson CR, Wilkinson IB, et al. The accuracy of diagnostic coding for acute kidney injury in England - a single centre study. BMC Nephrol 2013 Mar 13;14:58 [FREE Full text] [doi: 10.1186/1471-2369-14-58] [Medline: 23496869]
7. Mills KT, Xu Y, Zhang W, Bundy JD, Chen C, Kelly TN, et al. A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. Kidney Int 2015 Nov;88(5):950-957 [FREE Full text] [doi: 10.1038/ki.2015.230] [Medline: 26221752]
8. Bikbov B, Purcell C, Levey A, Smith M, Abdoli A, Abebe M, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2020 Feb;395(10225):709-733. [doi: 10.1016/S0140-6736(20)30045-3]
9. Hoste EA, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 2015 Aug;41(8):1411-1423. [doi: 10.1007/s00134-015-3934-7] [Medline: 26162677]

10. Sawhney S, Fraser SD. Epidemiology of AKI: utilizing large databases to determine the burden of AKI. Adv Chronic Kidney Dis 2017 Jul;24(4):194-204 [FREE Full text] [doi: 10.1053/j.ackd.2017.05.001] [Medline: 28778358]
11. Saran R, Robinson B, Abbott KC, Agodoa LY, Bragg-Gresham J, Balkrishnan R, et al. US Renal Data System 2018 annual data report: epidemiology of kidney disease in the United States. Am J Kidney Dis 2019 Mar;73(3 Suppl 1):A7-A8. [doi: 10.1053/j.ajkd.2019.01.001] [Medline: 30798791]
12. Coca SG, Yusuf B, Shlipak MG, Garg AX, Parikh CR. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis 2009 Jun;53(6):961-973 [FREE Full text] [doi: 10.1053/j.ajkd.2008.11.034] [Medline: 19346042]
13. Bello AK, Alrukhaimi M, Ashuntantang GE, Basnet S, Rotter RC, Douthat WG, et al. Complications of chronic kidney disease: current state, knowledge gaps, and strategy for action. Kidney Int Suppl (2011) 2017 Oct;7(2):122-129 [FREE Full text] [doi: 10.1016/j.kisu.2017.07.007] [Medline: 30675426]
14. Indicator details: all cause mortality rates by eGFR category/stage. Centers for Disease Control and Prevention. URL: https://nccd.cdc.gov/CKD/detail.aspx?Qnum=Q185#refreshPosition [accessed 2021-12-26]
15. Liu KD, Forni LG, Heung M, Wu V, Kellum JA, Mehta RL, Acute Disease Quality Initiative Investigators. Quality of care for acute kidney disease: current knowledge gaps and future directions. Kidney Int Rep 2020 Oct;5(10):1634-1642 [FREE Full text] [doi: 10.1016/j.ekir.2020.07.031] [Medline: 33102955]
16. Kellum JA, Lameire N, Aspelin P, Barsoum RS, Burdmann EA, Goldstein SL, et al. Kidney disease: improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Supplements 2012 Mar;2(1):1-138 [FREE Full text]
17. Eknoyan G, Lameire N. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Official J Int Soc Nephrol 2013 Jan;3(1) [FREE Full text]
18. Zappitelli M, Parikh CR, Akcan-Arikan A, Washburn KK, Moffett BS, Goldstein SL. Ascertainment and epidemiology of acute kidney injury varies with definition interpretation. Clin J Am Soc Nephrol 2008 Jul;3(4):948-954 [FREE Full text] [doi: 10.2215/CJN.05431207] [Medline: 18417742]
19. Siew ED, Matheny ME, Ikizler TA, Lewis JB, Miller RA, Waitman LR, et al. Commonly used surrogates for baseline renal function affect the classification and prognosis of acute kidney injury. Kidney Int 2010 Mar;77(6):536-542 [FREE Full text] [doi: 10.1038/ki.2009.479] [Medline: 20042998]
20. Bernier-Jean A, Beaubien-Souligny W, Goupil R, Madore F, Paquette F, Troyanov S, et al. Diagnosis and outcomes of acute kidney injury using surrogate and imputation methods for missing preadmission creatinine values. BMC Nephrol 2017 Apr 28;18(1):141 [FREE Full text] [doi: 10.1186/s12882-017-0552-3] [Medline: 28454562]
21. Liu KD, Hsu C, Yang J, Tan TC, Zheng S, Ordonez JD, et al. Acute kidney injury ascertainment is affected by the use of first inpatient versus outpatient baseline serum creatinine. Kidney Int Rep 2018 Jan;3(1):211-215 [FREE Full text] [doi: 10.1016/j.ekir.2017.08.011] [Medline: 29340333]
22. Siew ED, Matheny ME. Choice of reference serum creatinine in defining acute kidney injury. Nephron 2015;131(2):107-112 [FREE Full text] [doi: 10.1159/000439144] [Medline: 26332325]
23. Levey A, Inker L. Assessment of glomerular filtration rate in health and disease: a state of the art review. Clin Pharmacol Ther 2017 Sep;102(3):405-419. [doi: 10.1002/cpt.729] [Medline: 28474735]
24. Lafrance J, Miller DR. Defining acute kidney injury in database studies: the effects of varying the baseline kidney function assessment period and considering CKD status. Am J Kidney Dis 2010 Oct;56(4):651-660. [doi: 10.1053/j.ajkd.2010.05.011] [Medline: 20673605]
25. De Rosa S, Samoni S, Ronco C. Creatinine-based definitions: from baseline creatinine to serum creatinine adjustment in intensive care. Crit Care 2016 Mar 15;20:69 [FREE Full text] [doi: 10.1186/s13054-016-1218-4] [Medline: 26983854]
26. Björk J, Nyman U, Larsson A, Delanaye P, Pottel H. Estimation of the glomerular filtration rate in children and young adults by means of the CKD-EPI equation with age-adjusted creatinine values. Kidney Int 2021 Apr;99(4):940-947 [FREE Full text] [doi: 10.1016/j.kint.2020.10.017] [Medline: 33157151]
27. Hsu C, Yang W, Parikh RV, Anderson AH, Chen TK, Cohen DL, CRIC Study Investigators. Race, genetic ancestry, and estimating kidney function in CKD. N Engl J Med 2021 Nov 04;385(19):1750-1760. [doi: 10.1056/NEJMoa2103753] [Medline: 34554660]
28. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical decision support for in-hospital AKI. J Am Soc Nephrol 2018 Feb;29(2):654-660 [FREE Full text] [doi: 10.1681/ASN.2017070765] [Medline: 29097621]
29. Howarth M, Bhatt M, Benterud E, Wolska A, Minty E, Choi K, et al. Development and initial implementation of electronic clinical decision supports for recognition and management of hospital-acquired acute kidney injury. BMC Med Inform Decis Mak 2020 Nov 04;20(1):287 [FREE Full text] [doi: 10.1186/s12911-020-01303-x] [Medline: 33148237]
30. Hoste EA, Kashani K, Gibney N, Wilson FP, Ronco C, Goldstein SL, 15 ADQI Consensus Group. Impact of electronic-alerting of acute kidney injury: workgroup statements from the 15(th) ADQI Consensus Conference. Can J Kidney Health Dis 2016 Feb 26;3:10 [FREE Full text] [doi: 10.1186/s40697-016-0101-1] [Medline: 26925246]
31. James MT, Hobson CE, Darmon M, Mohan S, Hudson D, Goldstein SL, Acute Dialysis Quality Initiative (ADQI) Consensus Group. Applications for detection of acute kidney injury using electronic medical records and clinical information systems:

workgroup statements from the 15(th) ADQI Consensus Conference. Can J Kidney Health Dis 2016 Feb 26;3:9 [FREE Full text] [doi: 10.1186/s40697-016-0100-2] [Medline: 26925245]
32. Mohamadlou H, Lynn-Palevsky A, Barton C, Chettipally U, Shieh L, Calvert J, et al. Prediction of acute kidney injury with a machine learning algorithm using electronic health record data. Can J Kidney Health Dis 2018 Jun 8;5:2054358118776326 [FREE Full text] [doi: 10.1177/2054358118776326] [Medline: 30094049]
33. Holmes J, Rainer T, Geen J, Roberts G, May K, Wilson N, Welsh AKI Steering Group. Acute kidney injury in the era of the AKI E-Alert. Clin J Am Soc Nephrol 2016 Dec 07;11(12):2123-2131 [FREE Full text] [doi: 10.2215/CJN.05170516] [Medline: 27793961]
34. Siew ED, Basu RK, Wunsch H, Shaw AD, Goldstein SL, Ronco C, 15th ADQI Consensus Group. Optimizing administrative datasets to examine acute kidney injury in the era of big data: workgroup statement from the 15(th) ADQI Consensus Conference. Can J Kidney Health Dis 2016 Feb 26;3:12 [FREE Full text] [doi: 10.1186/s40697-016-0098-5] [Medline: 26925248]
35. Selby NM, Hill R, Fluck RJ, NHS England 'Think Kidneys' AKI Programme. Standardizing the early identification of acute kidney injury: the NHS England national patient safety alert. Nephron 2015;131(2):113-117 [FREE Full text] [doi: 10.1159/000439146] [Medline: 26351847]
36. Acute Kidney Injury (AKI) algorithm. NHS England. URL: https://www.england.nhs.uk/akiprogramme/aki-algorithm/ [accessed 2021-12-26]
37. Fiedler GM, Vogt B. [Laboratory tests for kidney disease]. Internist (Berl) 2019 May;60(5):485-501. [doi: 10.1007/s00108-019-0598-3] [Medline: 30997523]
38. Klassifikationen, Terminologien und Standards im Gesundheitswesen. Bundesinstitut für Arzneimittel und Medizinprodukt. URL: https://www.dimdi.de/dynamic/de/klassifikationen/ [accessed 2021-12-26]
39. Instrumente zur medizinischen Kodierung. Bundesamt für Statistik. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/nomenklaturen/medkk/instrumente-medizinische-kodierung.html [accessed 2021-12-26]
40. Rosa SE, Prowle J, Samoni S, Villa G, Ronco C. Acute kidney injury in patients with chronic kidney disease. In: Critical Care Nephrology (Third Edition). Amsterdam, Netherlands: Elsevier Inc; 2019.
41. Venkatachalam MA, Griffin KA, Lan R, Geng H, Saikumar P, Bidani AK. Acute kidney injury: a springboard for progression in chronic kidney disease. Am J Physiol Renal Physiol 2010 May;298(5):F1078-F1094 [FREE Full text] [doi: 10.1152/ajprenal.00017.2010] [Medline: 20200097]
42. Makris K, Spanou L. Acute kidney injury: diagnostic approaches and controversies. Clin Biochem Rev 2016 Dec;37(4):153-175 [FREE Full text] [Medline: 28167845]
43. Mehta R, Bihorac A, Selby NM, Quan H, Goldstein SL, Kellum JA, Acute Dialysis Quality Initiative (ADQI) Consensus Group. Establishing a continuum of acute kidney injury - tracing AKI using data source linkage and long-term follow-up: workgroup statements from the 15th ADQI consensus conference. Can J Kidney Health Dis 2016;3:13 [FREE Full text] [doi: 10.1186/s40697-016-0102-0] [Medline: 26925249]
44. Cooper DJ, Plewes K, Grigg MJ, Patel A, Rajahram GS, William T, et al. An evaluation of commonly used surrogate baseline creatinine values to classify AKI during acute infection. Kidney Int Rep 2021 Mar;6(3):645-656 [FREE Full text] [doi: 10.1016/j.ekir.2020.12.020] [Medline: 33732979]

Abbreviations

ACR: albumin creatinine ratio
AKI: acute kidney injury
CKD: chronic kidney disease
CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration
eGFR: estimated glomerular filtration rate
GFR: glomerular filtration rate
ICD: International Classification of Diseases
ICD-10-GM: International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification
KD: kidney disease
KDIGO: Kidney Disease: Improving Global Outcomes
NHS: National Health Service
SCr: serum creatinine
SwissDRG: Swiss Diagnosis Related Groups, Inpatient Tariff

Edited by C Lovis; submitted 20.06.21; peer-reviewed by T Ntalindwa, X Cheng, D Sengupta; comments to author 31.07.21; revised version received 14.10.21; accepted 14.11.21; published 25.01.22.

Please cite as:
Triep K, Leichtle AB, Meister M, Fiedler GM, Endrich O
Real-world Health Data and Precision for the Diagnosis of Acute Kidney Injury, Acute-on-Chronic Kidney Disease, and Chronic Kidney Disease: Observational Study
JMIR Med Inform 2022;10(1):e31356
URL: https://medinform.jmir.org/2022/1/e31356
doi:10.2196/31356
PMID:35076410

Original Paper

Patient Perspectives on the Digitization of Personal Health Information in the Emergency Department: Mixed Methods Study During the COVID-19 Pandemic

Sophia Ly[1], BSc; Ricky Tsang[1], BSc; Kendall Ho[2], MD

[1]Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

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

Corresponding Author:
Sophia Ly, BSc
Faculty of Medicine
University of British Columbia
3312-818 West 10th Avenue
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 6048220327
Email: sophia.ly@alumni.ubc.ca

Abstract

Background: Although the digitization of personal health information (PHI) has been shown to improve patient engagement in the primary care setting, patient perspectives on its impact in the emergency department (ED) are unknown.

Objective: The primary objective was to characterize the views of ED users in British Columbia, Canada, on the impacts of PHI digitization on ED care.

Methods: This was a mixed methods study consisting of an online survey followed by key informant interviews with a subset of survey respondents. ED users in British Columbia were asked about their ED experiences and attitudes toward PHI digitization in the ED.

Results: A total of 108 participants submitted survey responses between January and April 2020. Most survey respondents were interested in the use of electronic health records (79/105, 75%) and patient portals (91/107, 85%) in the ED and were amenable to sharing their ED PHI with ED staff (up to 90% in emergencies), family physicians (up to 91%), and family caregivers (up to 75%). In addition, 16 survey respondents provided key informant interviews in August 2020. Interviewees expected PHI digitization in the ED to enhance PHI access by health providers, patient-provider relationships, patient self-advocacy, and postdischarge care management, although some voiced concerns about patient privacy risk and limited access to digital technologies (eg, smart devices, internet connection). Many participants thought the COVID-19 pandemic could provide momentum for the digitization of health care.

Conclusions: Patients overwhelmingly support PHI digitization in the form of electronic health records and patient portals in the ED. The COVID-19 pandemic may represent a critical moment for the development and implementation of these tools.

(JMIR Med Inform 2022;10(1):e28981) doi:10.2196/28981

KEYWORDS

emergency medicine; digital health; health informatics; electronic health record; patient portal; patient-physician relationship; COVID-19

Introduction

Patient-centeredness, identified by the Institute of Medicine as one of six pillars of quality care, refers to care that is guided by patient preferences, needs, and values [1]. Although patient-centered approaches in the emergency department (ED) are associated with improved clinical outcomes and patient satisfaction [2-4], they can be challenging when high medical acuity, frequent care transitions, and an unpredictable environment compromise provider-patient communication and collaborative decision-making [5,6].

Personal health information (PHI) digitization is a potential strategy for improving provider-patient communication to support patient-centered care in the ED [6,7]. It encompasses a range of technologies that allow for the collection, analysis, and distribution of digital patient data [8]. These technologies can include electronic health records (EHRs) operated by health care providers as well as EHR-tethered portals for patients to access real-time PHI online.

There has been growing public interest in digital PHI tools. The percentage of Canadian physicians reporting that their patients used digital PHI technologies grew from 20.8% in 2017 to 44.7% by 2019 [9,10], when 74% of Canadian respondents expressed an interest in using patient portals [11]. The COVID-19 pandemic has further encouraged patients and providers to adopt digital health solutions in response to public health guidelines and social distancing requirements [12,13] and has precipitated calls for the widespread integration of digital tools in health care as our systems navigate beyond the COVID-19 crisis [12-14].

Although access to digital PHI has been shown to reduce anxiety, motivate lifestyle changes, and promote patient engagement in the primary care setting [15,16], patient attitudes toward digitization are not well characterized in the emergency setting where patient demographics, priorities, and care journeys may differ [5]. Nonetheless, most EDs in British Columbia (BC), Canada, now use some version of an EHR system that is integrated across the hospital departments within the local health authority and that feeds into CareConnect, a province-wide EHR platform viewable by physicians and other hospital-associated care providers [17]. Laboratory results—but not other EHR components, such as consult notes, imaging reports, and medication orders—are accessible by patients via an online portal [18].

There has been limited work examining the extent to which current digital PHI systems meet the needs of ED users or what opportunities there are to leverage PHI digitization to optimize care delivery in the ED setting. We therefore conducted a mixed methods study to explore the general perspectives of BC ED users on PHI digitization in emergency care.

Methods

Participant Recruitment

English-speaking adults aged >19 years who had received care in a BC ED within the last 5 years were invited to complete an online questionnaire via the University of British Columbia Digital Emergency Medicine social media channels, Vancouver Coastal Health Research Institute's REACH BC directory [19], regional patient networks that shared study details with members, and notices posted in the Vancouver General Hospital ED. Written consent was obtained from all participants.

Survey

The questionnaire was developed in consultation with 6 patients who have lived ED experiences and a working group of 15 clinicians and researchers brought together through a grant from the Michael Smith Foundation for Health Research in 2019. The questionnaire (Multimedia Appendix 1) included a combination of multiple-choice questions, Likert scales, and free-form text boxes. The eHealth Literacy Scale (eHEALS) [20] was included in the questionnaire to assess participants' digital health literacy. The questionnaire was administered online via Qualtrics and took approximately 20 minutes to complete. Participants were asked about their demographics, recent experiences in the ED, experiences with digital health technologies, preferences on the use of their digitized ED PHI, and the expected impacts of PHI digitization on the ED experience.

Key Informant Interviews

Survey participants who indicated that they wished to participate in future activities related to the study were invited by email to provide key informant interviews. Interviews (Multimedia Appendix 2) took place by phone or via the videoconferencing platform Zoom and lasted approximately 30 minutes. Participants were asked about their ED experiences and attitudes toward digital health technologies in the ED. Interviews were audio recorded and transcribed.

Data Analysis

Survey submissions with more than 20% of items missing were excluded from analysis. Quantitative responses were summarized with descriptive statistics (eg, mean, SD, frequency) and figures were generated using Google Sheets (Google LLC). Statistical tests were not performed as the purpose of our quantitative analysis was to provide a general picture of ED user characteristics and preferences rather than to make comparisons or to identify associations. Qualitative survey and interview responses were analyzed using a conventional content analysis approach wherein codes were defined a posteriori over the course of the analysis [21]. Coding was done independently in NVivo 12 (version 12.6.0; QSR International) by SL and RT, who met regularly to discuss thematic findings. Consensus was achieved for all codes.

Results

Participant Demographics

A total of 205 participants responded to the online survey between January and April 2020, of which 108 submissions had <20% of items missing and were included in the final analysis. Of these 108 participants, 16 provided key informant interviews in August 2020. Participant characteristics are summarized in Table 1. Participants were predominantly female (77/108, 71%) and Caucasian (83/108, 77%). Almost all participants reported daily internet (102/107, 95%) and smart device (106/108, 98%) access. Survey and interview participants were comparable in their ED and digital technology experiences, although interview participants reported higher levels of education and income.

Most participants resided within the Lower Mainland of British Columbia (67/108, 62%). In British Columbia, there are 5 geographic health authorities that manage health services in different parts of the province: Vancouver Coastal Health, Fraser Health, Vancouver Island Health, Interior Health, and Northern Health. The distribution of participants who received care from each health authority is also shown in Table 1.

Table 1. Participant demographics.

Demographics	Survey (N=108)[a]	Interview (N=16)[a]
Age (years), mean (SD; range)	47.1 (16.8; 19-84)	50.7 (15.9; 21-76)
Sex, n (%)		
Female	77 (71)	12 (75)
Male	24 (22)	2 (13)
Other/prefer not to answer	7 (7)	2 (13)
Ethnicity, n (%)[b]		
Caucasian	83 (77)	13 (81)
East Asian	11 (10)	2 (13)
Aboriginal	5 (5)	0 (0)
Latin American/Hispanic	3 (3)	0 (0)
South Asian	2 (2)	0 (0)
Other/prefer not to answer	22 (20)	5 (31)
Education, n (%)		
Some high school	1 (1)	0 (0)
High school diploma	15 (14)	1 (6)
Trade/technical training	23 (21)	1 (6)
Bachelor's degree	34 (31)	8 (50)
Graduate/professional degree	24 (22)	6 (38)
Prefer not to answer	11 (10)	0 (0)
Household income ($), n (%)		
<40,000	28 (26)	3 (19)
40,000-60,000	10 (9)	3 (19)
60,000-80,000	11 (10)	1 (6)
80,000-100,000	18 (17)	2 (13)
>100,000	20 (19)	5 (31)
Prefer not to answer	21 (19)	2 (13)
British Columbia health authority in which emergency department care was most recently accessed, n (%)		
Vancouver Coastal Health Authority	42 (39)	9 (56)
Fraser Health Authority	25 (23)	4 (25)
Vancouver Island Health Authority	19 (18)	1 (6)
Interior Health Authority	10 (9)	2 (13)
Northern Health Authority	5 (5)	0 (0)
Prefer not to answer	7 (6)	0 (0)
Chronic disease, n (%)		
Yes	66 (62)	9 (56)
No	31 (29)	5 (31)
I don't know	10 (9)	2 (13)
Number of emergency department visits in last 5 years, mean (SD; range)	3.4 (2.8; 1-15)	3.0 (2.5; 1-10)
Emergency department visits with altered level of consciousness, n (%)	13 (12)	2 (13)
Emergency department visits with life-threatening medical circumstances, n (%)	32 (30)	5 (31)
Internet use, n (%)		

Demographics	Survey (N=108)[a]	Interview (N=16)[a]
Daily	102 (95)	16 (100)
Weekly	3 (3)	0 (0)
Monthly	0 (0)	0 (0)
Less than once per month	2 (2)	0 (0)
Computer, tablet, or smartphone use, n (%)		
Daily	106 (98)	16 (100)
Weekly	1 (1)	0 (0)
Monthly	0 (0)	0 (0)
Less than once per month	1 (1)	0 (0)
Past use of digital health technologies, n (%)	87 (81)	14 (88)
eHealth Literacy Scale, mean (SD; range)	33.0 (7.4; 8-40)	32.3 (7.8; 16-40)

[a]Total number of responses may not equal total number of participants as responses were not required for all questions.

[b]Percentages may sum to greater than 100% as participants were able to select multiple responses.

Survey

Figure 1 summarizes participant attitudes toward EHRs in BC EDs. Survey respondents generally supported EHR implementation, with 75% (79/105) in favor, 7% (7/105) against, and 18% undecided (19/105). Respondents expected EHR use to improve their understanding of their medical condition (64/108, 59%), their overall quality of care (59/108, 55%), their relationship with ED staff (50/108, 46%), and their say in care (48/108, 44%). In contrast, 1%-8% (1/108 to 9/108) of respondents expected EHRs to worsen care across these domains. Respondents were generally willing to disclose different components of their EHR to ED staff (68/108, 64% to 90/108, 83% of participants in nonemergencies and 86/108, 80% to 97/108, 90% in emergencies). They were more willing to provide access to their family physicians (83/108, 86% to 98/108, 91% in both nonemergencies and emergencies) and less willing to provide access to designated family/friend caregivers (26/108, 24% to 57/108, 53% in nonemergencies and 57/108, 53% to 81/108, 75% in emergencies). In addition, 73% (79/108) were willing to share deidentified health data with researchers.

When asked about other potential impacts of EHRs in the ED, participants stated that they may provide ED staff with more timely access to relevant PHI (17 respondents) and allow patients to review clinician comments, promoting accountability (2 respondents). In addition, 16 respondents voiced concerns that EHRs increase the risk of unauthorized PHI disclosure, with 5 respondents stating that this was a definite barrier to their support for PHI digitization.

Figure 1. Patient perspectives on ED EHRs. (A) Percentage of respondents who support implementation of EHRs in the ED (N=105). (B) Perceived impacts of EHRs on satisfaction with ED care (N=108). (C) Preferences for ED EHR information disclosure in nonemergency and emergency situations (N=108). "General medical information" refers to test results, diagnoses, and medications. "Sensitive health information" refers to details about sexual health, mental health, and domestic violence. ED: emergency department; EHR: electronic health record; HCP: health care provider.

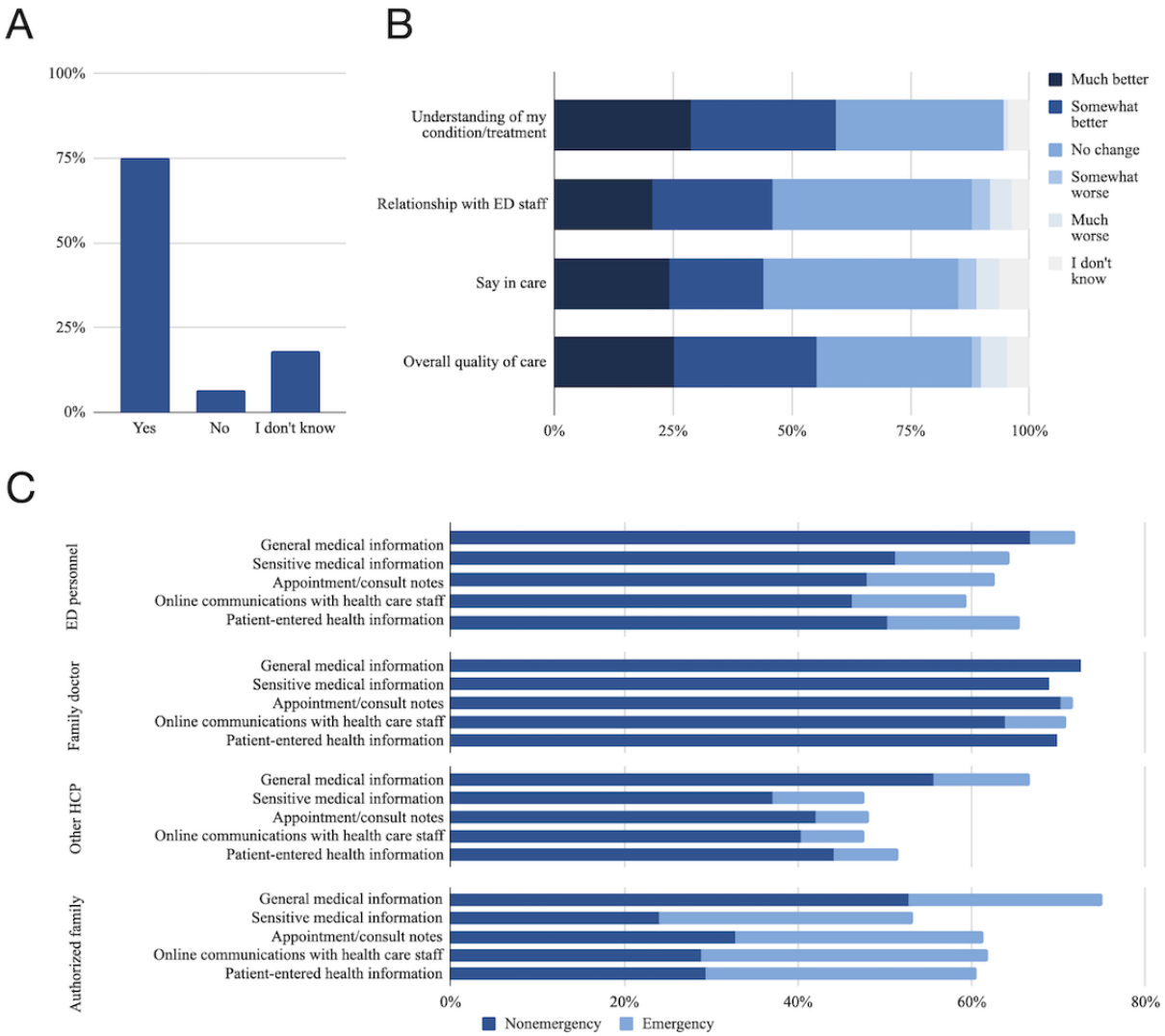
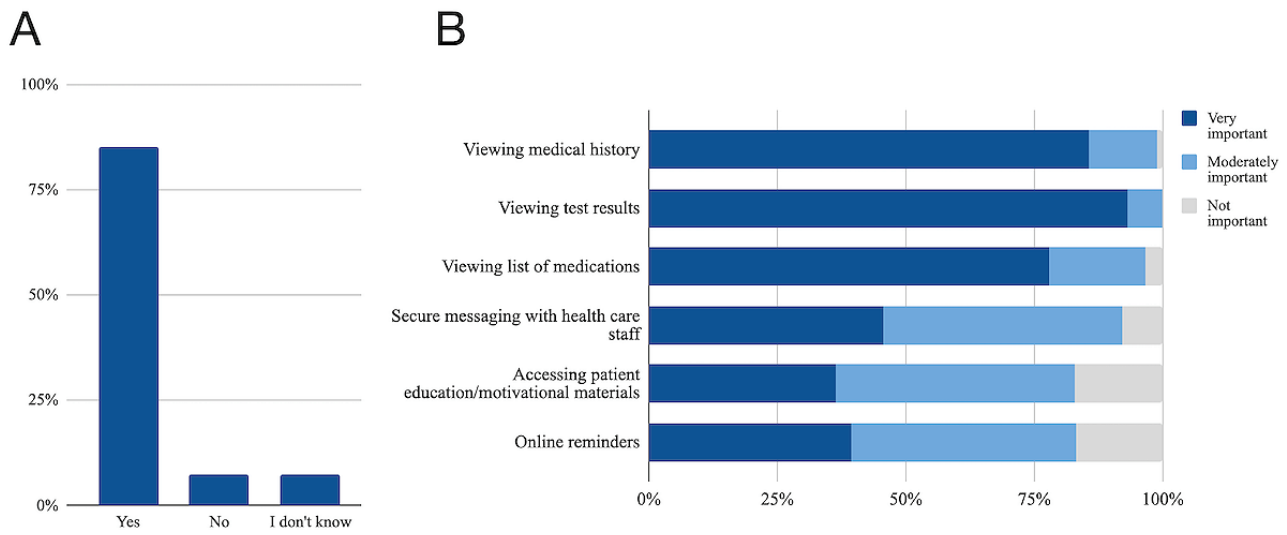


Figure 2 summarizes participants' views on ED patient portals. Overall, 85% (91/107) of survey respondents were interested in using a portal to access their ED EHR. Of those respondents, 73% (66/91) reported that they would use it in hospital and 100% (91/91) postdischarge. Patient-prioritized features included the ability to view personal medical histories, test results, and medications, which were rated as "very important" by 77% (70/91) to 85% (83/91) of respondents. Some respondents also rated as "very important" the ability to securely message ED staff (41/91, 45%), access patient education or motivational materials (32/91, 35%), and access online reminders (35/91, 38%).

Figure 2. Patient perspectives on ED patient portals. (A) Percentage of respondents interested in using a patient portal to access digitized PHI in their own ED EHR (N=107). (B) Patient-prioritized features for an ED portal (N=91, corresponding to the participants who indicated that they were interested in using a portal to access their ED EHR). ED: emergency department; EHR: electronic health record; PHI: personal health information.



When asked about other potential impacts of patient portals in the ED, participants stated that they would help them to learn about their ED journey (3 respondents), follow discharge instructions (2 respondents), and share information about their visit with community care providers (4 respondents). Participants stated that barriers to portal use include medical incapacity in the ED (6 respondents); limited access to smart devices, internet, or electrical outlets in the ED (5 respondents); limited access to smart devices or the internet in the community (7 respondents); and a challenging user interface (15 respondents).

Key Informant Interviews

Key informant interviews were conducted to clarify how participants expected PHI digitization to impact ED care. A total of 62 survey participants expressed an interest in being interviewed, of which 16 were ultimately recruited (4 declined, 42 did not respond to follow-up). Of the 16 interviewees, 7 had work experience in health care.

ED Access to PHI During Emergencies

Multiple factors may limit ED access to past medical information: patients may be unable to share PHI due to medical incapacity or emotional stress (7 interviewees), collateral may be incomplete (1 interviewee), and patients may not be trusted to provide accurate information concerning controversial diagnoses (eg, Ehlers-Danlos syndrome) without documentation (2 interviewees).

Several participants expected PHI digitization to enhance history-taking by facilitating ED access to data stored in an EHR integrated between EDs and other health services (9 interviewees). One interviewee expressed surprise upon learning that BC EDs did not already have access to her family physician's electronic records:

> *When I realized that the hospital didn't have my health history digitally when they did that intake a couple years ago, I was like, oh my gosh. People think that their health is saved more digitally at their doctor's office and in the hospital than it actually is.*

Improved ED access to patient medical histories was expected to increase the efficiency of face-to-face patient-physician interactions (13 interviewees) and promote confidence in the quality of care received (2 interviewees). Multiple participants, however, expressed concern that digitization could facilitate unauthorized access to PHI by corporations or health professionals not involved in their care (6 interviewees).

Relationship Between Patients and ED Staff

Interviewees suggested that relationships between patients and ED staff can be undermined when physical discomfort (1 interviewee), anxiety (2 interviewees), or feelings of being neglected during long wait times (2 interviewees) contribute to high tensions during in-person interactions. There were also concerns about poor accountability from ED staff in cases of medical error or professional misconduct (3 interviewees).

Participants generally expected relationships with ED staff to improve with PHI digitization (10 interviewees). By updating patients on their medical status in real time, ED portals may alleviate anxiety ahead of face-to-face interactions with care providers (2 interviewees) and offer a glimpse of behind-the-scenes care processes, providing reassurance that patients are not forgotten during their visit (2 interviewees). As one interviewee stated, "If I know the reason why I'm waiting in the emergency room is because they're just waiting for results and diagnostics... I know what I'm waiting for and don't feel like I've been deprioritized."

Two interviewees described how patient-ED relationships may worsen with PHI digitization. One stated that electronic access to historic medical records may facilitate the disclosure of stigmatizing information (eg, psychiatric conditions), biasing providers against patients. The other interviewee, a former ED

nurse, indicated that digital technologies may detract from the human aspect of care:

> *When you improve efficiency, you kind of lose the art of…from my perspective, nursing. Where you take the time to put a warm blanket on, to hold somebody's hand, to help them with their dentures, whatever is required.*

Participants also suggested that patient portals could be a tool for improving accountability from ED staff. Portals may allow patients to identify errors or discriminatory remarks in their chart (3 interviewees). One interviewee suggested that the opportunity for patients and providers to participate in mutual surveillance may deconstruct the power imbalance inherent in clinical relationships.

Self-advocacy in the ED

Several interviewees described how patient self-advocacy in the ED can be compromised by insufficient opportunity to process information from health professionals, with one participant stating:

> *A lot of what happens in healthcare is a one-way conversation. It's almost as an afterthought at the end of a whole bunch of information spewing towards you – do you have any questions? And you don't have enough time to really think about it and digest what you just heard to formulate a question quickly, especially if you're in the emergency department in pain.*

Concern about interrupting the ED workflow was also identified as a barrier to self-advocacy. One interviewee stated that she did not receive analgesia until the end of her visit as she did not know the appropriate way to voice her concern and "just didn't feel like bothering anyone."

Patient portals in the ED may allow patients to learn about their medical status ahead of in-person encounters, facilitating more informed decision-making (6 interviewees). Portals may also provide a nonintrusive process for bringing up care concerns, increasing the likelihood that they will be voiced (2 interviewees). Barriers to their use in the ED include medical incapacity (8 interviewees) and limited access to smart devices (1 interviewee), which may be minimized through patient-accessible smart devices in the ED or user controls authorizing portal access by designated family members during emergencies.

Self-management After the ED

Participants indicated that ED patients have limited access to visit details for postdischarge self-management. Medical incapacity and emotional stress can prevent patients from recalling visit details presented verbally by care providers (4 interviewees) and incidental findings are not consistently shared with patients (3 interviewees).

ED portals were suggested to enhance patients' understanding of their medical condition at discharge (14 interviewees), increase compliance with discharge instructions (5 interviewees), and facilitate online self-education (5 interviewees). One respondent remarked that visitor restrictions due to the COVID-19 pandemic made it more important for patients cognitively impaired by pain or illness to have a digital record of their visit postdischarge. Digital access to ED test results may also allow for follow-up of incidental findings. Two interviewees stated that they were diagnosed with medical conditions that could have been identified earlier had they been informed of abnormal results obtained in the ED.

Digital ED PHI access was expected to enhance information-sharing with family caregivers, allowing them to better support patients in decision-making and day-to-day care implementation (eg, transport to appointments; 3 interviewees). Digitization was also expected to improve information-sharing with allied health professionals, giving patients more autonomy in where they seek postdischarge care (6 interviewees).

Potential barriers to effective portal use postdischarge may include limited access to smart devices or the internet, particularly for rural-dwelling or low-income patients (7 interviewees), as well as difficulties using the portal interface or interpreting medical information (14 interviewees).

Effects of the COVID-19 Pandemic on Patient Attitudes Toward Digital PHI Technologies

In total, 6 interviewees stated that the COVID-19 pandemic has highlighted the importance of digital health technologies in modern health care delivery. In addition, 4 further expressed that the COVID-19 pandemic has provided government and health care organizations with the impetus to enact these technologies, with 1 participant describing how First Nations reservations in the BC Interior have recently established high-speed internet infrastructure to facilitate telehealth consultations.

Discussion

Principal Findings

Our findings suggest that the majority of participants are supportive of ED PHI digitization in the form of EHR and patient portal implementation. The anticipated benefits of PHI digitization on the patient emergency care experience can be grouped into four domains: (1) overcoming challenges of the ED environment by relieving anxiety and fostering relationships with staff, (2) facilitating access to information by ED staff and patients, (3) promoting self-advocacy by enhancing patient decision-making capacity and health care provider accountability, and (4) easing care transitions by facilitating medical self-management, self-education, and care planning with community providers. Users were interested in portal features consistent with these aims.

Although this is the first study to our knowledge that examines the perspectives of ED users on PHI digitization, these findings are consistent with primary care studies suggesting that portals can alleviate anxiety [22], increase patient activation [15,22], and facilitate collaborative relationships with clinicians [23,24]. Our results differ from those of previous studies by identifying barriers to portal use that are specific to the ED context, such as high medical acuity or difficulties with in-hospital internet and smart device access. In addition, whereas previous work in

the primary care context found that patient engagement in portals is contingent upon a pre-existing foundation of trust between patients and their providers [25], our results suggest that patient portals may work inversely in the emergency setting to foster trust in new providers.

Although the participants in our study were generally enthusiastic about PHI digitization and patient portals in the ED, positive perception may not translate to actual portal uptake. A recent study from the University of Iowa reported that only 8.9% of ED users used a portal to view their test results, possibly due to a lack of multilingual settings, internet and smart device access, or patient education on portal use [26]. It is therefore incumbent upon institutions to consult patients as stakeholders in the development of digital PHI tools and care providers to meaningfully engage patients in their use.

The minority of participants who opposed ED PHI digitization expressed concerns over information privacy and security. The potential for PHI compromise through third-party breaches or unauthorized release to employers or insurance companies is a common theme among studies exploring barriers to portal use [27]. Mitigation strategies include data minimization, encryption policies, proxy accounts providing family caregivers with access to preauthorized content, and audit trails allowing patients to view users who have accessed their EHR [28]. To safeguard patient confidence in digital PHI systems, the Canadian Medical Protective Association also recommends patient counselling on safe data practices and provider transparency regarding who has PHI access [29].

A major limitation of this study is that self-selection bias may have led to an overrepresentation of positive attitudes toward PHI digitization. Although our open recruiting strategy makes it challenging to determine the extent to which our survey cohort is representative of the general population of BC ED users, among our interview participants, 7 of 16 reported work experience in health care. There is evidence that health care workers self-report high levels of digital literacy and share homogenous, generally positive viewpoints toward PHI digitization [30]. Similarly, the perspectives of vulnerable and marginalized populations (eg, low socioeconomic status) were underrepresented in this study. Several interviewees stated that these populations may have unique perspectives on digitization, a suggestion supported by previous findings that lower engagement in eHealth activities is associated with lower socioeconomic status, ethnic minority status, and rural residency [31]. Future work should seek to capture the perspectives of a broader range of ED users to inform the creation of equitable digital PHI tools.

As of September 2021, COVID-19 continues to impact the global community. In British Columbia, a resurgence of cases emerged in July 2021 but began to stabilize as of late August 2021, with daily reported cases exceeding 600 in early September 2021 [32]. As we completed data collection in August 2020, we were unable to capture the ongoing effects of the COVID-19 pandemic on evolving patient attitudes in British Columbia. However, participant observations that the COVID-19 pandemic has spurred the health care system to implement overdue digital reforms allow us to hypothesize that support for PHI digitization is likely to remain robust as the global pandemic evolves.

Conclusion

Our findings suggest that BC ED users welcome PHI digitization and expect it to enhance their ED experience by increasing patient comfort, facilitating communication with ED health professionals, and improving post-ED care. The COVID-19 pandemic provides a window of opportunity for introducing digital PHI technologies to improve ED care as part of the larger digital revolution currently affecting health care internationally.

Acknowledgments

We are grateful to Helen Novak Lauscher, Giuseppe Carenini, Anito Ho, Richard Lester, Riyad Abu-Laban, Hubert Wong, and Colleen McGavin for their feedback on the design of this study. This work was supported by the Michael Smith Foundation for Health Research (grant F19-02689). Publishing fees were covered by Canada Health Infoway.

Authors' Contributions

SL and KH contributed to the conception and design of the study. SL and RT participated in data collection and analysis. The manuscript was prepared and approved by all three authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.
[DOCX File , 35 KB - medinform_v10i1e28981_app1.docx]

Multimedia Appendix 2
Interview questions.
[DOCX File , 16 KB - medinform_v10i1e28981_app2.docx]

References

1. Institute of Medicine. Six Aims for Improvement. In: Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001:41-53.
2. Brown K, Mace SE, Dietrich AM, Knazik S, Schamban NE. Patient and family-centred care for pediatric patients in the emergency department. CJEM 2008 Jan 21;10(1):38-43. [doi: 10.1017/s1481803500009994] [Medline: 18226317]
3. Hess EP, Knoedler MA, Shah ND, Kline JA, Breslin M, Branda ME, et al. The chest pain choice decision aid: a randomized trial. Circ Cardiovasc Qual Outcomes 2012 May;5(3):251-259 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.111.964791] [Medline: 22496116]
4. Smith S, Mitchell C, Bowler S. Standard versus patient-centred asthma education in the emergency department: a randomised study. Eur Respir J 2008 May 09;31(5):990-997 [FREE Full text] [doi: 10.1183/09031936.00053107] [Medline: 18216062]
5. Gordon J, Sheppard LA, Anaf S. The patient experience in the emergency department: A systematic synthesis of qualitative research. Int Emerg Nurs 2010 Apr;18(2):80-88. [doi: 10.1016/j.ienj.2009.05.004] [Medline: 20382369]
6. Pham JC, Trueger NS, Hilton J, Khare RK, Smith JP, Bernstein SL. Interventions to improve patient-centered care during times of emergency department crowding. Acad Emerg Med 2011 Dec;18(12):1289-1294 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01224.x] [Medline: 22168193]
7. Ly S, Ho K. Patient perspectives on the digitization of personal health information in the emergency department: Gaps and opportunities. Clinical eHealth 2020;3:5-6. [doi: 10.1016/j.ceh.2020.01.001]
8. The emerging benefits of electronic medical record use in community-based care. Canada Health Infoway. 2013. URL: https://www.infoway-inforoute.ca/en/component/edocman/1224-the-emerging-benefits-of-electronic-medical-record-use-in-community-based-care-full-report/view-document?Itemid=0 [accessed 2019-10-02]
9. 2017 CMA Physician Workforce Survey. Canadian Medical Association. 2017. URL: https://surveys.cma.ca/ [accessed 2021-01-08]
10. 2019 CMA Physician Workforce Survey. Canadian Medical Association. 2019. URL: https://surveys.cma.ca/ [accessed 2021-01-08]
11. Canadian Medical Association, Ipsos Group. The future of connected health care: reporting Canadians' perspective on the health care system. Canadian Medical Association. 2019 Aug. URL: https://www.cma.ca/sites/default/files/pdf/Media-Releases/The-Future-of-Connected-Healthcare-e.pdf [accessed 2021-01-08]
12. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
13. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
14. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
15. Ferreira A, Correia A, Silva A, Corte A, Pinto A, Saavedra A, et al. Why facilitate patient access to medical records. Stud Health Technol Inform 2007;127:77-90. [Medline: 17901601]
16. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. Am J Health Promot 2019 Jun;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]
17. Physician's guide to electronic medical records (EMRs). Doctors of BC. 2019. URL: https://www.doctorsofbc.ca/sites/default/files/dto_physicians_guide_to_emr_frequently_asked_questions_faqs.pdf [accessed 2021-01-13]
18. Excelleris Technologies Inc. MyCareCompass. 2021. URL: https://www.bc.mycarecompass.lifelabs.com/?province=BC#/ [accessed 2021-02-26]
19. REACH BC. Vancouver Coastal Health Research Institute. URL: https://www.reachbc.ca [accessed 2021-12-06]
20. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
21. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
22. Shah SD, Liebovitz D. It Takes Two to Tango: Engaging Patients and Providers With Portals. PM R 2017 May 18;9(5S):S85-S97. [doi: 10.1016/j.pmrj.2017.02.005] [Medline: 28527507]
23. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2017 Apr;26(4):262-270 [FREE Full text] [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
24. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
25. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

26. Foster B, Krasowski MD. The Use of an Electronic Health Record Patient Portal to Access Diagnostic Test Results by Emergency Patients at an Academic Medical Center: Retrospective Study. J Med Internet Res 2019 Jun 28;21(6):e13791 [FREE Full text] [doi: 10.2196/13791] [Medline: 31254335]
27. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using Electronic Health Record Portals to Improve Patient Engagement: Research Priorities and Best Practices. Ann Intern Med 2020 Jun 02;172(11 Suppl):S123-S129 [FREE Full text] [doi: 10.7326/M19-0876] [Medline: 32479176]
28. Latulipe C, Quandt SA, Melius KA, Bertoni A, Miller DP, Smith D, et al. Insights Into Older Adult Patient Concerns Around the Caregiver Proxy Portal Use: Qualitative Interview Study. J Med Internet Res 2018 Nov 02;20(11):e10524 [FREE Full text] [doi: 10.2196/10524] [Medline: 30389654]
29. Patient portals: considerations for safety and medico-legal risk. Canadian Medical Protective Association. 2021. URL: https://www.cmpa-acpm.ca/en/advice-publications/browse-articles/2021/patient-portals-considerations-for-safety-and-medico-legal-risk [accessed 2021-09-20]
30. Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health Informatics J 2020 Mar;26(1):592-612 [FREE Full text] [doi: 10.1177/1460458219839613] [Medline: 30983476]
31. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
32. British Columbia (BC) COVID-19 Situation Report Week 35: August 29 - September 4, 2021. BC Centre for Disease Control. 2021. URL: http://www.bccdc.ca/Health-Info-Site/Documents/COVID_sitrep/Week_35_2021_BC_COVID-19_Situation_Report.pdf [accessed 2021-09-08]

Abbreviations

BC: British Columbia
ED: emergency department
eHEALS: eHealth Literacy Scale
EHR: electronic health record
PHI: personal health information

Edited by G Eysenbach, S Hagens; submitted 20.03.21; peer-reviewed by CY Yeh, F Ghezelbash, J Pecina; comments to author 13.09.21; revised version received 27.09.21; accepted 23.11.21; published 06.01.22.

Please cite as:
Ly S, Tsang R, Ho K
Patient Perspectives on the Digitization of Personal Health Information in the Emergency Department: Mixed Methods Study During the COVID-19 Pandemic
JMIR Med Inform 2022;10(1):e28981
URL: https://medinform.jmir.org/2022/1/e28981
doi:10.2196/28981
PMID:34818211

Original Paper

Assessment of Natural Language Processing Methods for Ascertaining the Expanded Disability Status Scale Score From the Electronic Health Records of Patients With Multiple Sclerosis: Algorithm Development and Validation Study

Zhen Yang[1], MSc; Chloé Pou-Prom[1], MSc; Ashley Jones[2], BSc; Michaelia Banning[1], BA (Hons); David Dai[1], MSc; Muhammad Mamdani[1,3,4,5,6], PharmD, MA, MPH; Jiwon Oh[2,4,7], MD, PhD; Tony Antoniou[1,3,8,9], PhD

[1]Data Science and Advanced Analytics, Unity Health Toronto, Toronto, ON, Canada
[2]Division of Neurology, Department of Medicine, St. Michael's Hospital, Toronto, ON, Canada
[3]Li Ka Shing Knowledge Institute, Unity Health Toronto, Toronto, ON, Canada
[4]Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[5]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada
[6]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada
[7]Department of Neurology, Johns Hopkins University, Baltimore, MD, United States
[8]Department of Family and Community Medicine, Unity Health Toronto, Toronto, ON, Canada
[9]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Tony Antoniou, PhD
Department of Family and Community Medicine
Unity Health Toronto
30 Bond Street
Toronto, ON, M5B 1W8
Canada
Phone: 1 416 867 7460 ext 77056
Email: tony.antoniou@unityhealth.to

Abstract

Background: The Expanded Disability Status Scale (EDSS) score is a widely used measure to monitor disability progression in people with multiple sclerosis (MS). However, extracting and deriving the EDSS score from unstructured electronic health records can be time-consuming.

Objective: We aimed to compare rule-based and deep learning natural language processing algorithms for detecting and predicting the total EDSS score and EDSS functional system subscores from the electronic health records of patients with MS.

Methods: We studied 17,452 electronic health records of 4906 MS patients followed at one of Canada's largest MS clinics between June 2015 and July 2019. We randomly divided the records into training (80%) and test (20%) data sets, and compared the performance characteristics of 3 natural language processing models. First, we applied a rule-based approach, extracting the EDSS score from sentences containing the keyword "EDSS." Next, we trained a convolutional neural network (CNN) model to predict the 19 half-step increments of the EDSS score. Finally, we used a combined rule-based–CNN model. For each approach, we determined the accuracy, precision, recall, and F-score compared with the reference standard, which was manually labeled EDSS scores in the clinic database.

Results: Overall, the combined keyword-CNN model demonstrated the best performance, with accuracy, precision, recall, and an F-score of 0.90, 0.83, 0.83, and 0.83 respectively. Respective figures for the rule-based and CNN models individually were 0.57, 0.91, 0.65, and 0.70, and 0.86, 0.70, 0.70, and 0.70. Because of missing data, the model performance for EDSS subscores was lower than that for the total EDSS score. Performance improved when considering notes with known values of the EDSS subscores.

Conclusions: A combined keyword-CNN natural language processing model can extract and accurately predict EDSS scores from patient records. This approach can be automated for efficient information extraction in clinical and research settings.

(JMIR Med Inform 2022;10(1):e25157) doi:10.2196/25157

KEYWORDS

machine learning; multiple sclerosis; natural language processing

Introduction

Multiple sclerosis (MS) is the most common cause of neurological disability in young adults in the developed world [1]. Although the majority of individuals present initially with relapsing-remitting disease, neurological disability can accumulate over time, resulting in significant functional impairment in a substantial portion of people with MS [1,2]. However, there is considerable individual heterogeneity in MS disease progression, such that validated measures of disability are required to monitor functional decline and response to disease-modifying therapies.

The Kurtzke Expanded Disability Status Scale (EDSS) is the most widely used validated measure to quantify and monitor changes in MS-related disability over time [3,4]. The EDSS is a clinician-administered ordinal rating system that quantifies disability in 8 functional systems, increasing from 0 (no disability) to 10 (death due to MS) in increments of 0.5 units. EDSS subscores can also be determined for each of the individual functional systems comprising the total score, using a scale that ranges from 0 to 5 or 6 [3,4]. Because the EDSS score is used for both clinical and research purposes, it is typically extracted or derived manually from electronic medical records and transcribed in clinical and research databases to monitor trends in disease evolution and response to treatment [5-7]. However, the EDSS score may not be determined at all visits, introducing missing data when patient records are used for research and clinical monitoring [8]. Moreover, extracting and deriving the EDSS score from patient records is time-consuming and inefficient because of the unstructured nature of clinical records [9].

Natural language processing is a field of artificial intelligence that is increasingly being applied to extract and transform unstructured notes in electronic medical records into coded data that can be used for clinical, quality improvement, and research purposes [10,11]. Natural language processing has been studied in a variety of clinical settings, including oncology, emergency medicine, and primary care, for applications as varied as case ascertainment, risk assessment, and disease staging [12-16]. Within the field of MS, comparatively few studies have investigated the use and performance of natural language models. Specific areas of application have included identifying patients with MS from clinical databases, extracting disease-specific variables, detecting genotype-phenotype associations for MS from an electronic medical record–linked DNA biorepository, identification and sentiment analysis of MS-related content on social media, biomedical literature mining, and using clinical variables to derive a disease severity score [9,17-24]. Existing studies thus far have largely evaluated rule-based natural language processing approaches, wherein clinicians provide keywords and a predetermined set of rules to locate specific text in a note that denotes a particular finding as either present or absent. Deep learning natural language processing approaches, in which machine learning algorithms are trained to capture specific outcomes from text, have been less well studied in the MS field. Our objective was to compare rule-based and deep learning natural language processing algorithms for detecting and predicting the total EDSS score and EDSS functional system subscores from clinic notes.

Methods

Setting and Data Sources

The Barlo MS Centre of Unity Health Toronto is one of the largest MS clinics in Canada, providing specialized care to over 7000 Ontario residents living with MS. The clinic database contains comprehensive information on all patients, including demographic data, relapse and treatment history, imaging results, and findings from neurological examinations, including EDSS and functional system scores. For this study, we extracted all clinical notes generated for patients seen at the clinic between June 2015 and July 2019, and randomly divided all notes in the study period into training (80%) and test (20%) data sets. We divided notes at the patient level to prevent data leakage (ie, same patient appearing in both training and test data sets).

Data Preprocessing

To prepare notes for rule-based and deep learning natural language processing, we first removed all redundant information, including patient and physician names within the header and footer of each note, date and time of visit, fax number, and document number. We also removed identifying information such as home addresses, phone numbers, patient identification number, and dates of birth and electronic signatures, as well as nonletter characters such as punctuation, symbols, and left-over whitespace. Next, we removed stop words using the Natural Language Toolkit default list [25]. Stop words are commonly used terms (eg, “and,” “it,” “the,” etc) that have little value with respect to the meaning of clinical text. We completed these steps so that only the most relevant parts of the document would be provided as input to the text classification model. Finally, we encoded each note into a sequence of integers, setting the maximum sequence length to 1000 words, which is within the limit of most notes included for study. We zero-padded sequences with smaller word counts, and removed the last few words when the sequence count exceeded the maximum length. Preprocessing steps were automated, applicable to the test-time/application-time, and did not require manual review.

Natural Language Processing

We compared the performance characteristics of 3 natural language processing models in outputting 1 EDSS score for each note. First, we used a rule-based approach, wherein the preprocessed text was divided into sentences, and extracted the EDSS score on the first occasion when “EDSS” and a numeric value between 0.0 and 10.0 appeared in the same sentence. To extract EDSS functional system subscores, MS clinic staff were

consulted to develop rules that paired keyword patterns representing clinical findings relevant to a specific functional system (eg "ataxia" for the cerebellar subscore and "indwelling catheter" for the bowel and bladder score) with adjectives denoting the varying levels of disability related to each functional system, such as "mild," "moderate," or "significant." These rules were based on Neurostatus definitions and scoring for neurological examinations [26]. Using this approach, EDSS subscores were extracted or derived for each functional system.

Because it is possible that multiple keywords can appear in the same note (eg, "EDSS was 5.0 in the previous visit. …EDSS is 6.0 in this visit."), the rule-based approach may result in errors when extracting the most recent EDSS score, highlighting the potential limitations of this approach and the need to evaluate alternative models. We therefore trained separate convolutional neural network (CNN) models to predict the 19 half-step increments of the total EDSS score and the functional system subscores. CNNs are artificial neural networks that are being increasingly used for applications as varied as image detection and natural language processing [27-29]. In the case of the latter, text must first be converted into a numerical form known as a word vector before it can be fed into a CNN model. To do this, we experimented with various approaches, including Bidirectional Encoder Representations from Transformers (BERT) [30], BioBERT [31], deep contextualized word representations (Embeddings from Language Models [ELMo]) [32], and pretrained Word2Vec (trained on PubMed, Wiki, and PubMed Central) [33]. A comparison of these approaches found that Word2Vec trained on our hospital data had superior performance and runtime relative to the other approaches. Moreover, Word2Vec embeddings trained on our data were able to capture semantic relationships between MS-related terms. For example, the terms RRMS ("relapsing-remitting multiple sclerosis"), AMS ("active multiple sclerosis"), and CIS ("clinically isolated syndrome") are identified as nearest neighbors of the term "MS," using our approach. We therefore trained a 200-dimensional Word2Vec embedding with all neurologist specialty notes from the clinic using Gensim [34]. Word2Vec is a 2-layer neural network net that transforms inputted text into numerical vectors, or embeddings, of a given size (eg, 200 dimensions) that can be processed by CNNs [35]. This is done by grouping the vectors mathematically based on word similarity, with similar words being closer to each other when mapped in multidimensional space, while unrelated words are separated by greater distance. For all of the CNN models, we used 200-dimensional Word2Vec embeddings trained on all clinical notes from the MS clinic. Word embeddings were trained using a window size of 10 and a minimum count of 2, yielding an embedding matrix with a dimension of 1000×200, reflecting the maximum sequence length of 1000 words, that acted as an embedding layer in the CNN models. We chose a 1000-word maximum sequence length based on premodeling determinations of the word count of the consult notes comprising our data set demonstrating that most notes fell within this limit. The CNN model is based on a well-known CNN structure used for sentence classification (Figure 1) [29]. First, a section of the note is represented as a numeric feature (ie, word embedding with a dimension of 1000×200). Next, convolutional layers with multiple filters of different kernel sizes (sizes 3, 4, and 5) are applied to obtain multiple features (with dropout rate 0.5 and maximum pooling on each of the convolutions). Features are then passed to a fully connected layer whose output is the probability distribution over the list of EDSS classes. Therefore, in addition to the embedding layer, CNN models also contained convolutional layers with maximal pooling and fully connected layers with Softmax output (Figure 1) [29]. We implemented the model using Keras 2.0 API [36], and trained the model using the RMSprop optimizer and early stopping to prevent overfitting from too many iterations. We experimented with different learning rates, epochs, batch sizes, and patience for early stopping, choosing the hyperparameters that delivered the best accuracy for our test data. We also tried shallow neural networks (unigram features and a cutoff of 5000 features ordered by term frequency) with term-frequency inverse document frequency features and recurrent neural networks (RNNs) for our study. In the case of the former, we found that these models did not adequately represent word relations and context-based information. Moreover, these approaches created extremely high dimensional sparse input vectors. Although findings with RNNs were comparable, we elected to proceed with the CNN and Word2Vec approach because these models were faster to train.

Finally, we used a keyword-CNN model to ascertain whether the combination of the 2 approaches would yield better performance metrics than either model alone. We reasoned that a combined model would balance the strengths and limitations of each model separately. Specifically, while CNN models perform well with large data volumes and are less time-intensive than rule-based approaches, these models typically lack transparency and explainability, leaving users with little understanding of how predictions and decisions are made. Moreover, CNN models may not perform well when data volumes are small, such as for patients at the highest extremes of EDSS scores. In contrast, while rule-based approaches are transparent and explainable (ie, extracted keyword patterns in notes can be shown to users), and have good performance for rare outcomes, they will predict mostly unknown results when keywords are not explicitly found in the reference text. To account for these strengths and weaknesses, we developed a combined model that involves 2 steps. First, the model uses a rule-based approach to detect whether the EDSS score is explicitly written in a given note. In such a case, the model outputs the extracted EDSS score. In the event that keywords are not explicitly written, the note is passed on to the CNN, which will provide a prediction for the EDSS score (Figure 2).

Figure 1. Convolutional neural network model structure. EDSS: Expanded Disability Status Scale.

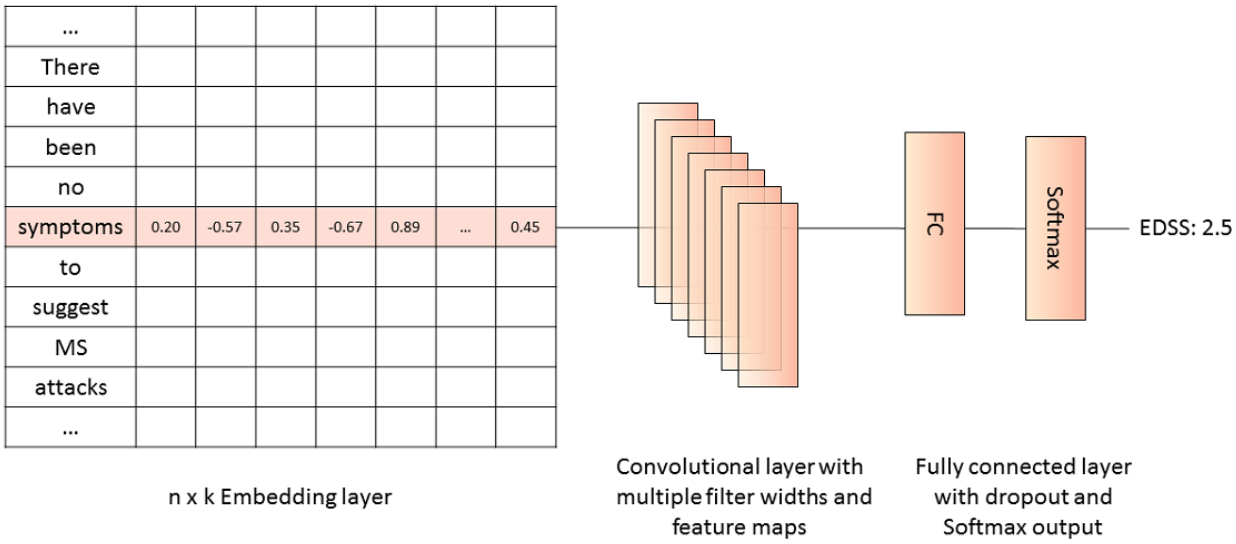
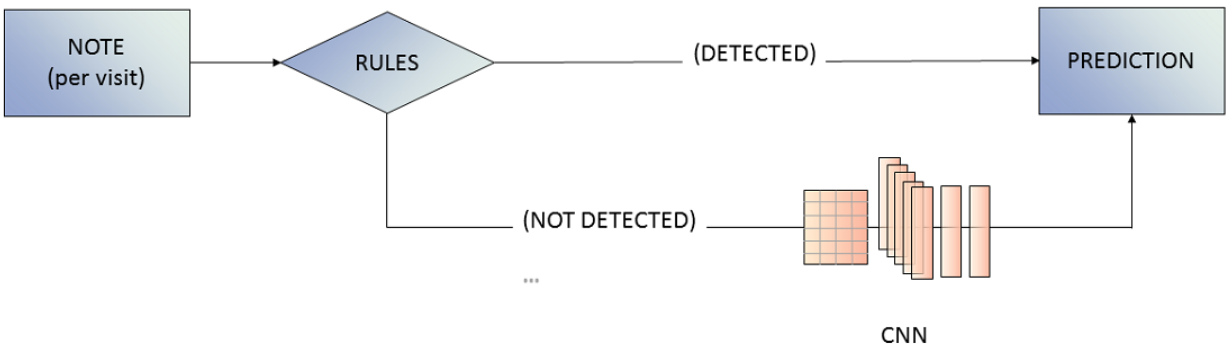


Figure 2. Combined rule-based–CNN model. CNN: convolutional neural network.



Statistical Analysis

After training, all models were evaluated on the 3493 notes comprising the test set. Our primary outcome was the performance of each model for abstracting and/or deriving the total EDSS score. We determined the accuracy, precision, recall, and F-score of each model compared with the reference standard, which was the manually labeled EDSS scores in the clinic database. Accuracy is the ratio of correct predictions made (ie, true positives plus true negatives) to the total number of predictions made (ie, sum of true positives, false positives, true negatives, and false negatives). For total EDSS scores, predictions were considered accurate if they were identical to those recorded in patient records. For functional subscores, predictions were considered accurate if they were within +/−1 of their referent values. Precision is calculated by dividing the number of true positive predictions by the sum of true and false positives, whereas recall is defined as the number of true positives over the total number of positives (ie, sum of true positives and false negatives). To determine precision and recall, we considered each score as a class, and obtained true positive, false positive, true negative, and false negative rates for each class. Finally, the F-score is a metric that combines precision and recall into a single number using the harmonic mean, thereby taking both false positives and false negatives into account. Compared with accuracy, the F1-score provides a more robust measure of incorrectly classified cases in imbalanced class settings such as ours. In all cases, we determined macro average performance measures, obtained by first calculating each class metric and then taking the average of these. We used Pitman permutation tests to determine whether model differences in accuracy and F1-scores were statistically significant [37]. In secondary analyses, we determined the performance of each model in abstracting functional system EDSS subscores. In a sensitivity analysis, we replicated our analyses using 10-fold cross-validation on the training set. For each fold, we used 90% of the notes for training and 10% for validation, and then applied the hyperparameters producing the best results in the cross-validation toward evaluating the test set.

Ethics Approval

This study was approved by the Research Ethics Board of Unity Health Toronto, Toronto, Canada (reference #16-371).

Results

Our data set comprised 17,452 clinic notes for 4906 patients seen at the MS clinic between June 2015 and July 2019. Overall, the mean age of the patients was 49.5 (SD 12.4) years, and 3534 (72%) were female. The majority of notes (n=10,881, 62.3%) had an EDSS score explicitly dictated. There was considerable class imbalance in the EDSS labels, with 13,880 (79.5%) and 1386 (7.9%) scores being in the range of 0.0 to 4.0 and above 6.0, respectively.

In our main analysis, the rule-based model delivered greater precision than the CNN model (0.91 vs 0.71) for predicting the total EDSS score. Conversely, the CNN model had greater accuracy (0.86 vs 0.57) and slightly better recall (0.70 vs 0.65) relative to the rule-based model (Multimedia Appendix 1). In a qualitative error analysis of the validation set (n=3493 notes), the numbers and proportions of instances where the EDSS score was captured by both models, captured only by the rule-based method, captured only by the CNN, and missed by both models were 1864 (53.4%), 122 (3.5%), 1155 (33.1%), and 352 (10.1%), respectively. Model performance varied at the extremes of the EDSS score, with the rule-based approach performing worse at the lower ranges where patient disability is minimal, while the CNN model underpredicted EDSS scores in patients with very high levels of disability (Multimedia Appendix 2). Specifically, the F-scores for the rule-based and CNN models at EDSS scores of 0 to 4 were 0.69 and 0.89, respectively, while those for EDSS scores greater than 4 were 0.78 and 0.54, respectively. We observed similar patterns when comparing notes that did (n=2172, 62.2%) and did not (n=1321, 37.8%) report an EDSS score (Multimedia Appendix 3). For notes with an explicit EDSS score, the accuracies of the rule-based and CNN models were 0.87 and 0.93, respectively, with the rule-based model achieving greater performance at higher EDSS scores and slightly lower performance at lower EDSS scores, in part because of lower recall when the EDSS score is 0.0. For notes lacking an explicit EDSS score, the accuracy of the CNN model was 0.74, while the rule-based model was unable to return an EDSS score, with all predictions being labeled as "unknown."

When compared with each model individually, the combined rule-based–CNN model performed best for predicting the total EDSS score, with accuracy, precision, recall, and an F-score of 0.90, 0.82, 0.83, and 0.83, respectively (Multimedia Appendix 1). We obtained similar results for the combined model using 10-fold cross-validation, with accuracy and an F-score of 0.87 and 0.81, respectively. The differences in accuracy and F1-score between the combined rule-based–CNN model and both the rule-based and CNN models were statistically significant (P<.001). The proportions of records with an unknown EDSS score prediction with the rule-based model, CNN model, and combined model were 44.43% (1552/3493), 3.06% (107/3493), and 2.83% (99/3493), respectively.

Similar to the total EDSS score, the combined model performed best for predicting EDSS functional system subscores (Multimedia Appendix 1). However, relative to the total EDSS score, functional system subscores had higher rates of unknown values in patient records, ranging from 8.2% for the ambulation subscore to 33.3% for the cerebral subscore. Consequently, performance measures were generally lower for combined models predicting EDSS functional system subscores relative to the total score (Multimedia Appendix 1). We therefore determined a post-hoc converted accuracy by excluding unknown values from the analysis and calculating performance metrics from notes with valid scores. The converted accuracy exceeded 0.90 for all EDSS functional system subscores, ranging from 0.94 for the sensory function subscore to 0.98 for brainstem and bowel/bladder function subscores.

Discussion

In our study, we found that a combined rule-based–CNN natural language processing approach can accurately extract the EDSS score from the clinic notes of people with MS. Moreover, the combined model was able to derive the EDSS score in notes that did not explicitly contain this information using available MS-specific variables. These results highlight the feasibility of developing automated algorithms for the extraction of clinically relevant information that would be otherwise challenging to abstract manually from unstructured data sources.

Our work confirms and builds upon earlier work using natural language processing methods in the field of MS in several ways [9,17-24]. First, while previous studies have used rule-based approaches to develop classification algorithms for identifying patients with MS and extracting clinically relevant information from electronic health records, we compared 3 separate natural language processing models for extracting the EDSS score, demonstrating that the combination of a CNN and rule-based algorithm leverages the strengths of each method while overcoming the limitations inherent in each approach. Specifically, the rule-based model exhibited greater precision, excelling when the keyword "EDSS" and an associated score appeared explicitly in the note, but had lower recall, particularly for patients at the lowest extreme of EDSS scores where physicians may be more likely to provide a qualitative summary of a patient's disability status with no accompanying EDSS score (eg, "neurological exam remains normal"). In such cases, the rule-based approach will return an EDSS score of "unknown," signifying no extraction of any score. Additionally, the rule-based approach struggled with cases where there were multiple EDSS scores in the note (eg, "she previously had an EDSS score of 5.0 and her current score is of 6.0"), or when the EDSS score was written in a format not accounted for in our rules (eg, "EDSS was three"). These limitations were reduced by the CNN model, which derived an EDSS score using high-level text features in the note and performed well in predicting EDSS scores in the lower range. Conversely, class imbalance in the higher range of EDSS scores undermined the performance of the CNN model, resulting in underprediction of the EDSS score among the very few patients with extremely high scores (Multimedia Appendix 2). This weakness was mitigated when combined with the rule-based model, which

performed well for high EDSS scores by capturing relevant keyword patterns. By combining the 2 models, we leveraged the strengths of each to optimize performance for both low and high EDSS scores.

Second, although previous studies have demonstrated that natural language processing models can extract the EDSS score and the related MS severity score from patient records containing these data [9,21,23], we demonstrated that a combined rule-based–CNN model could derive the EDSS score from notes where this measure was not explicitly provided, a phenomenon observed in approximately one-third of the notes available for study. The ability to automate EDSS score derivation using available clinical data may address issues of missing data within electronic health records and facilitate the use of these databases for quality improvement and research purposes.

Finally, we examined whether natural language processing models could extract functional EDSS subscores from electronic health records. Our model was able to extract the subscores, albeit with less precision than the total EDSS score. This is a line of inquiry that has not been addressed in prior studies.

Our study has some limitations. Although there were a sufficient number of notes available for ascertaining model performance related to the total EDSS score, data were sparser for our secondary analyses of the functional system subscores. These findings should therefore be considered hypothesis generating, and they warrant further evaluation with larger data sets. In addition, our models were developed and validated using the records of a single MS clinic embedded within a large academic teaching hospital. Consequently, the portability of our models is unknown. Finally, our models identify cross-sectional associations and cannot be considered as algorithms that predict disability progression in patients with MS. However, our models may automate the extraction of this information for use as inputs in future studies of machine learning approaches for predicting outcomes in patients with MS.

In conclusion, we found that a combined rule-based–CNN model was superior to either model alone for extracting and/or deriving EDSS scores from the records of patients with MS. This approach can be harnessed to establish and maintain clinical and research databases of people with MS, which may otherwise be too time-consuming and labor-intensive to maintain.

Acknowledgments

This study was funded by grants from St. Michael’s Hospital Foundation and Li Ka Shing Foundation, and an unrestricted research grant from Roche (Canada).

Authors' Contributions

Conception and/or design of the study: ZY, CP-P, AJ, MB, DD, MM, JO, and TA; data acquisition/analysis: ZY, CPP, and AJ; interpretation of results: ZY, CP-P, AJ, MB, DD, MM, JO, and TA; drafting of the manuscript: TA; revision of the manuscript: ZY, CP-P, AJ, MB, DD, MM, JO, and TA; final approval of the manuscript: ZY, CP-P, AJ, MB, DD, MM, JO, and TA.

Conflicts of Interest

JO reports grants from MS Society of Canada, The Barford and Love MS Fund of St. Michael’s Hospital Foundation, National MS Society, Brain Canada, Biogen-Idec, Roche, and EMD-Serono; and personal fees for consulting or speaking from Biogen-Idec, EMD-Serono, Roche, Sanofi-Genzyme, Novartis, and Celgene.

Multimedia Appendix 1
Model performance for predicting the total Expanded Disability Status Scale score and functional system subscores.
[DOCX File , 19 KB - medinform_v10i1e25157_app1.docx]

Multimedia Appendix 2
Perclass model performance for the rule-based, convolutional neural network, and combined models.
[DOCX File , 15 KB - medinform_v10i1e25157_app2.docx]

Multimedia Appendix 3
Performance of the rule-based, convolutional neural network, and combined models stratified by the presence or absence of the Expanded Disability Status Scale score in notes.
[DOCX File , 17 KB - medinform_v10i1e25157_app3.docx]

References

1. Murray TJ. Diagnosis and treatment of multiple sclerosis. BMJ 2006 Mar 04;332(7540):525-527 [FREE Full text] [doi: 10.1136/bmj.332.7540.525] [Medline: 16513709]
2. Compston A, Coles A. Multiple sclerosis. The Lancet 2008 Oct;372(9648):1502-1517. [doi: 10.1016/s0140-6736(08)61620-7]
3. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov 01;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]

4. Meyer-Moock S, Feng Y, Maeurer M, Dippel F, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. BMC Neurol 2014 Mar 25;14(1):58 [FREE Full text] [doi: 10.1186/1471-2377-14-58] [Medline: 24666846]
5. Uitdehaag BMJ. Disability Outcome Measures in Phase III Clinical Trials in Multiple Sclerosis. CNS Drugs 2018 Jun 20;32(6):543-558 [FREE Full text] [doi: 10.1007/s40263-018-0530-8] [Medline: 29926371]
6. Inojosa H, Schriefer D, Ziemssen T. Clinical outcome measures in multiple sclerosis: A review. Autoimmun Rev 2020 May;19(5):102512. [doi: 10.1016/j.autrev.2020.102512] [Medline: 32173519]
7. Rae-Grant A, Bennett A, Sanders AE, Phipps M, Cheng E, Bever C. Quality improvement in neurology: Multiple sclerosis quality measures. Neurology 2015 Sep 02;85(21):1904-1908. [doi: 10.1212/wnl.0000000000001965]
8. Davis MF, Haines JL. The intelligent use and clinical benefits of electronic medical records in multiple sclerosis. Expert Review of Clinical Immunology 2014 Dec 11;11(2):205-211. [doi: 10.1586/1744666x.2015.991314]
9. Davis MF, Sriram S, Bush WS, Denny JC, Haines JL. Automated extraction of clinical traits of multiple sclerosis in electronic medical records. J Am Med Inform Assoc 2013 Dec 01;20(e2):e334-e340 [FREE Full text] [doi: 10.1136/amiajnl-2013-001999] [Medline: 24148554]
10. Kimia AA, Savova G, Landschaft A, Harper MB. An Introduction to Natural Language Processing. Pediatric Emergency Care 2015;31(7):536-541. [doi: 10.1097/pec.0000000000000484]
11. Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K, et al. Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances. J Biomed Inform 2018 Dec;88:11-19 [FREE Full text] [doi: 10.1016/j.jbi.2018.10.005] [Medline: 30368002]
12. Kehl KL, Elmarakeby H, Nishino M, Van Allen EM, Lepisto EM, Hassett MJ, et al. Assessment of Deep Natural Language Processing in Ascertaining Oncologic Outcomes From Radiology Reports. JAMA Oncol 2019 Oct 01;5(10):1421-1429 [FREE Full text] [doi: 10.1001/jamaoncol.2019.1800] [Medline: 31343664]
13. Yim W, Yetisgen M, Harris WP, Kwan SW. Natural Language Processing in Oncology: A Review. JAMA Oncol 2016 Jun 01;2(6):797-804. [doi: 10.1001/jamaoncol.2016.0213] [Medline: 27124593]
14. St-Maurice J, Kuo M, Gooch P. A Proof of Concept for Assessing Emergency Room Use with Primary Care Data and Natural Language Processing. Methods Inf Med 2012 Dec 07;52(01):33-42. [doi: 10.3414/me12-01-0012]
15. Sterling NW, Patzer RE, Di M, Schrager JD. Prediction of emergency department patient disposition based on natural language processing of triage notes. Int J Med Inform 2019 Sep;129:184-188. [doi: 10.1016/j.ijmedinf.2019.06.008] [Medline: 31445253]
16. Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, Johnson A, et al. Predicting Intensive Care Unit admission among patients presenting to the emergency department using machine learning and natural language processing. PLoS One 2020 Mar 3;15(3):e0229331 [FREE Full text] [doi: 10.1371/journal.pone.0229331] [Medline: 32126097]
17. Krysko KM, Ivers NM, Young J, O'Connor P, Tu K. Identifying individuals with multiple sclerosis in an electronic medical record. Mult Scler 2015 Feb 12;21(2):217-224. [doi: 10.1177/1352458514538334] [Medline: 24948683]
18. Middleton R, Akbari A, Lockhart-Jones H, Jones J, Owen C, Hughes S, et al. Clinical Validation of the UKMS Register Minimal Dataset utilising Natural Language Processing. IJPDS 2017 Apr 18;1(1):268. [doi: 10.23889/ijpds.v1i1.288]
19. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 2010 Apr 09;86(4):560-572 [FREE Full text] [doi: 10.1016/j.ajhg.2010.03.003] [Medline: 20362271]
20. Palakal M, Bright J, Sebastian T, Hartanto S. A comparative study of cells in inflammation, EAE and MS using biomedical literature data mining. J Biomed Sci 2007 Jan 3;14(1):67-85. [doi: 10.1007/s11373-006-9120-8] [Medline: 17082901]
21. Xia Z, Secor E, Chibnik LB, Bove RM, Cheng S, Chitnis T, et al. Modeling disease severity in multiple sclerosis using electronic health records. PLoS One 2013 Nov 11;8(11):e78927 [FREE Full text] [doi: 10.1371/journal.pone.0078927] [Medline: 24244385]
22. Chase HS, Mitrani LR, Lu GG, Fulgieri DJ. Early recognition of multiple sclerosis using natural language processing of the electronic health record. BMC Med Inform Decis Mak 2017 Feb 28;17(1):24 [FREE Full text] [doi: 10.1186/s12911-017-0418-4] [Medline: 28241760]
23. Damotte V, Lizée A, Tremblay M, Agrawal A, Khankhanian P, Santaniello A, et al. Harnessing electronic medical records to advance research on multiple sclerosis. Mult Scler 2019 Mar 09;25(3):408-418. [doi: 10.1177/1352458517747407] [Medline: 29310490]
24. Nelson RE, Butler J, LaFleur J, Knippenberg K, Kamauu AWC, DuVall SL. Determining Multiple Sclerosis Phenotype from Electronic Medical Records. J Manag Care Spec Pharm 2016 Dec;22(12):1377-1382. [doi: 10.18553/jmcp.2016.22.12.1377] [Medline: 27882837]
25. Natural language processing with Python. Natural Language Toolkit. URL: http://nltk.org [accessed 2020-09-24]
26. Neurostatus. URL: https://www.neurostatus.net/ [accessed 2021-12-08]
27. Bernal J, Kushibar K, Asfaw DS, Valverde S, Oliver A, Martí R, et al. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif Intell Med 2019 Apr;95:64-81. [doi: 10.1016/j.artmed.2018.08.008] [Medline: 30195984]

28. Hughes M, Li I, Kotoulas S, Suzumura T. Medical Text Classification Using Convolutional Neural Networks. Stud Health Technol Inform 2017;235:246-250. [Medline: 28423791]
29. Kim Y. Convolutional neural networks for sentence classification. arXiv. 2014. URL: https://arxiv.org/pdf/1408.5882.pdf [accessed 2021-04-07]
30. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2019 Jun Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics; June 2019; Minneapolis, MN URL: https://www.aclweb.org/anthology/N19-1423/
31. Lee J, Yoon W, Kim S, Kim D, Kim S, So C, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
32. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep contextualized word representations. arXiv. 2018. URL: https://arxiv.org/pdf/1802.05365.pdf [accessed 2021-04-07]
33. Pyysalo S, Ginter F, Moen H, Salakoski T, Ananiadou S. Distributional semantics resources for biomedical text processing. Biomedical natural language processing tools and resources. URL: https://bio.nlplab.org/pdf/pyysalo13literature.pdf [accessed 2021-04-07]
34. Gensim Models — Word2vec Embeddings. Radim Řehůřek. URL: https://radimrehurek.com/gensim/models/word2vec.html [accessed 2021-04-07]
35. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionalitys. arXiv. 2013. URL: https://arxiv.org/pdf/1310.4546.pdf [accessed 2021-04-07]
36. Chollet F. Keras: Deep learning for humans. GitHub. URL: https://github.com/keras-team/keras [accessed 2021-04-07]
37. Dror R, Baumer G, Shlomov S, Reichart R. The Hitchhiker's Guide to Testing Statistical Significance in Natural Language Processing. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018 Presented at: 56th Annual Meeting of the Association for Computational Linguistics; July 2018; Melbourne, Australia URL: https://www.aclweb.org/anthology/P18-1128/ [doi: 10.18653/v1/p18-1128]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
CNN: convolutional neural network
EDSS: Expanded Disability Status Scale
MS: multiple sclerosis
RNN: recurrent neural network

Edited by C Lovis; submitted 20.10.20; peer-reviewed by M Torii, H Suominen; comments to author 18.12.20; revised version received 08.04.21; accepted 19.11.21; published 12.01.22.

Please cite as:
Yang Z, Pou-Prom C, Jones A, Banning M, Dai D, Mamdani M, Oh J, Antoniou T
Assessment of Natural Language Processing Methods for Ascertaining the Expanded Disability Status Scale Score From the Electronic Health Records of Patients With Multiple Sclerosis: Algorithm Development and Validation Study
JMIR Med Inform 2022;10(1):e25157
URL: https://medinform.jmir.org/2022/1/e25157
doi:10.2196/25157
PMID:35019849

Original Paper

The Development History and Research Tendency of Medical Informatics: Topic Evolution Analysis

Wenting Han[1], PhD; Xi Han[2], PhD; Sijia Zhou[3], PhD; Qinghua Zhu[4], PhD

[1]School of Management & Engineering, Nanjing University, Nanjing, China

[2]School of Business Administration, Guangdong University of Finance & Economics, Guangzhou, China

[3]Department of Information Systems, City University of Hong Kong, Hong Kong, Hong Kong

[4]School of Information Management, Nanjing University, Nanjing, China

Corresponding Author:
Xi Han, PhD
School of Business Administration
Guangdong University of Finance & Economics
21 Luntou Road
Guangzhou, 510320
China
Phone: 86 13512791305
Email: hanx015@163.com

Abstract

Background: Medical informatics has attracted the attention of researchers worldwide. It is necessary to understand the development of its research hot spots as well as directions for future research.

Objective: The aim of this study is to explore the evolution of medical informatics research topics by analyzing research articles published between 1964 and 2020.

Methods: A total of 56,466 publications were collected from 27 representative medical informatics journals indexed by the Web of Science Core Collection. We identified the research stages based on the literature growth curve, extracted research topics using the latent Dirichlet allocation model, and analyzed topic evolution patterns by calculating the cosine similarity between topics from the adjacent stages.

Results: The following three research stages were identified: early birth, early development, and rapid development. Medical informatics has entered the fast development stage, with literature growing exponentially. Research topics in medical informatics can be classified into the following two categories: data-centered studies and people-centered studies. Medical data analysis has been a research hot spot across all 3 stages, and the integration of emerging technologies into data analysis might be a future hot spot. Researchers have focused more on user needs in the last 2 stages. Another potential hot spot might be how to meet user needs and improve the usability of health tools.

Conclusions: Our study provides a comprehensive understanding of research hot spots in medical informatics, as well as evolution patterns among them, which was helpful for researchers to grasp research trends and design their studies.

(JMIR Med Inform 2022;10(1):e31918) doi:10.2196/31918

KEYWORDS

medical informatics; research hotspot; LDA model; topic evolution analysis; mobile phone

Introduction

Background

Medical informatics is a discipline that has received much attention in recent years. It has flourished with the development of information technology [1]. In 1959, Ledley and Lusted [2] suggested using computers to support medical decisions, which combined information technology with the medical domain. In the 1970s, the International Federation for Information Processing proposed the term *medical informatics*. It was defined as “the application of computer technology to all fields of medicine—medical care, medical teaching, and medical research.”

Systematic reviews of a research area are impactful because they can help researchers grasp future research trends and better design their studies. There have been many reviews of medical

informatics conducted over the past 5 decades. Methods including bibliometric methods, visualization technologies, and social network analysis were always used in these reviews. For example, previous research used cocitation networks and co-occurring keywords to uncover knowledge structures in medical informatics [3], as well as keyword analysis [4] (such as keyword-frequency statistics and keyword clustering) to discover research topics. Visualization tools [5], including VOSviewer and CiteSpace, were used to reveal the scientific networks. In addition, some researchers brought MeSH (Medical Subject Headings) terms into medical informatics studies to extract high-quality research topics [6] or journals [7].

After reviewing medical informatics, we found that most systematic reviews in this field discovered research trends using bibliometric methods based on paper keywords, which summarized research contents into several words. Keywords, by contrast, had fewer semantic information compared with abstracts.

Objectives

In this study, we chose the latent Dirichlet allocation (LDA) model to extract research topics from research article abstracts. Furthermore, we attempted to explore topic evolution patterns to predict future research trends. In conclusion, our study will be guided by the following three issues: (1) What are the research stages in the development of medical informatics, and what are the features of each stage? (2) What are the research hot spots in medical informatics and at different stages? Do these research hot spots change over time? (3) How have these research topics evolved over time? What will be the future research trends?

Methods

Data Collection

This study collected publications indexed by the Web of Science Core Collection database. To fully retrieve articles in medical informatics, we chose papers published by 27 representative medical informatics journals (Textbox 1) according to the medical informatics journal list supplied by the Journal Citation Reports. By limiting the document types into research articles and setting the published time before 2020, we downloaded the total records of 56,466 articles on April 16, 2021.

Textbox 1. Twenty-seven representative medical informatics journals (ranked by initials).

Titles of journals

1. *Applied Clinical Informatics*
2. *Artificial Intelligence in Medicine*
3. *Biomedical Engineering—Biomedizinische Technik*
4. *BMC Medical Informatics and Decision Making*
5. *Cin—Computers Informatics Nursing*
6. *Computer Methods and Programs in Biomedicine*
7. *Health Informatics Journal*
8. *Health Information Management Journal*
9. *IEEE Journal of Biomedical and Health Informatics*
10. *Informatics for Health & Social Care*
11. *International Journal of Medical Informatics*
12. *International Journal of Technology Assessment in Health Care*
13. *Internet interventions—The Application of Information Technology in Mental and Behavioral Health*
14. *JMIR Medical Informatics*
15. *JMIR mHealth and uHealth*
16. *JMIR Serious Games*
17. *Journal of Biomedical Informatics*
18. *Journal of Evaluation in Clinical Practice*
19. *Journal of Medical Internet Research*
20. *Journal of Medical Systems*
21. *Journal of the American Medical Informatics Association*
22. *Medical & Biological Engineering & Computing*
23. *Medical Decision Making*
24. *Methods of Information in Medicine*
25. *Statistical Methods in Medical Research*
26. *Statistics in Medicine*
27. *Therapeutic Innovation & Regulatory Science*

Research Design

Research Stage Identification

To determine how research topics evolve over time, we need to divide the history of medical informatics during the last 5 decades into several time units. Previous studies that analyzed publications released in the last 5-10 years usually took a year as a time unit [8]. When the time span exceeds decades, evidence for distinguishing time units, such as the life cycle theory [9], is necessary. In this study, we choose the literature growth curve of Price [10] to identify time units because this theory provides the quantitative features of literature growth in each stage. In the early stage, the number of research papers is minimal and increases unsteadily. At this point, no mathematical model perfectly fits the growth curve. Then, the number of research publications rises dramatically in the development stage, following the exponential increase model. In the mature stage, the number of papers grows slowly and steadily, with a growth trend that is consistent with the linear increase model. Finally, in the last stage of discipline, the number of papers declines as theories and research in 1 discipline become saturated. Furthermore, the growth curve would either gradually parallel the horizontal axis or fluctuate irregularly.

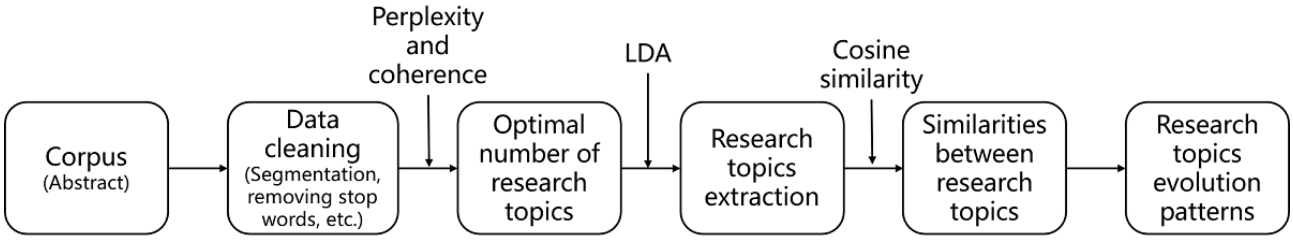
According to the literature growth curve of Price [10], a discipline's development history can be divided into stages based on the rate of literature growth. To divide the past 5 decades of medical informatics into distinct stages, we used the piecewise regression algorithm to fit the curve of the annual cumulative number of research papers. The time point that can separate the development stages occurs when the curve slopes are significantly distinguished. After identifying these time points, we attempted to match the literature growth curve in every stage with various mathematical models (linear increase model, exponential increase model, etc) to find the features of each stage.

Topic Evolution Analysis

Topic evolution analysis was adopted in this study to extract research topics and explore their evolution patterns. There are many topic extraction methods, including those based on word frequency, co-occurrence, and topic models. Compared with the first 2 methods, extracting topics through topic models, which can mine topics from a semantic perspective and show a better topic distribution, is suitable for our research. From various topic models, we chose the LDA model [11] for topic extraction. The LDA model uses the Dirichlet distribution to perform probability modeling at three levels: document, topic, and word. It calculates the semantic similarities between topics, documents, topics, and keywords. Many previous studies have shown that this model is effective in research topic mining and research trend prediction [12,13]. Before extracting topics using the LDA model, we had to determine the optimal number of topics extracted. Perplexity [11] and coherence [14] are always chosen as indicators. The optimal number of topics occurs when the value of perplexity is low, and the value of coherence is high.

Then, we needed to calculate the similarity between topics from adjacent stages to identify their relationships. Previous studies have used semantic similarity between keywords under 2 topics to represent topic similarity [15,16]. If the similarity of 2 keyword vectors exceeds a threshold, the evolutionary relationship between 2 topics is identified; otherwise, it is not. Typical measures of word vector similarity include Jensen-Shannon divergence, Kullback-Leibler divergences, and cosine similarity [16,17]. In this study, we used Python coding programs to calculate the cosine similarity between the 2 topics. The cosine similarity value ranges from 0 to 1, with higher values indicating greater similarity. It is reasonable to take 0.5 as a threshold. Figure 1 provides an overview of the topic evolution analysis process.

Figure 1. The process of topic evolution analysis. LDA: latent Dirichlet allocation.



Results

Identify Research Stages

As stated previously, we counted the annual cumulative number of research papers and plotted the literature growth curves in Figure 2.

Then, to find the points that significantly distinguish the rate of literature growth, we used the piecewise regression algorithm in Python to fit the curve of the annual cumulative number of papers in Figure 2. The fitting results are shown in Figure 3.

Figure 3 indicates that the curve was inflected in 1992 and 2010. Therefore, we divided the past 5 decades into three stages: 1964-1991, 1992-2009, and 2010-2020. We then adopted SPSS (IBM Corporation) to fit the growth curve for each stage. Curve fitting yielded the following results.

Figure 2. Annual distribution of the cumulative number of research papers.

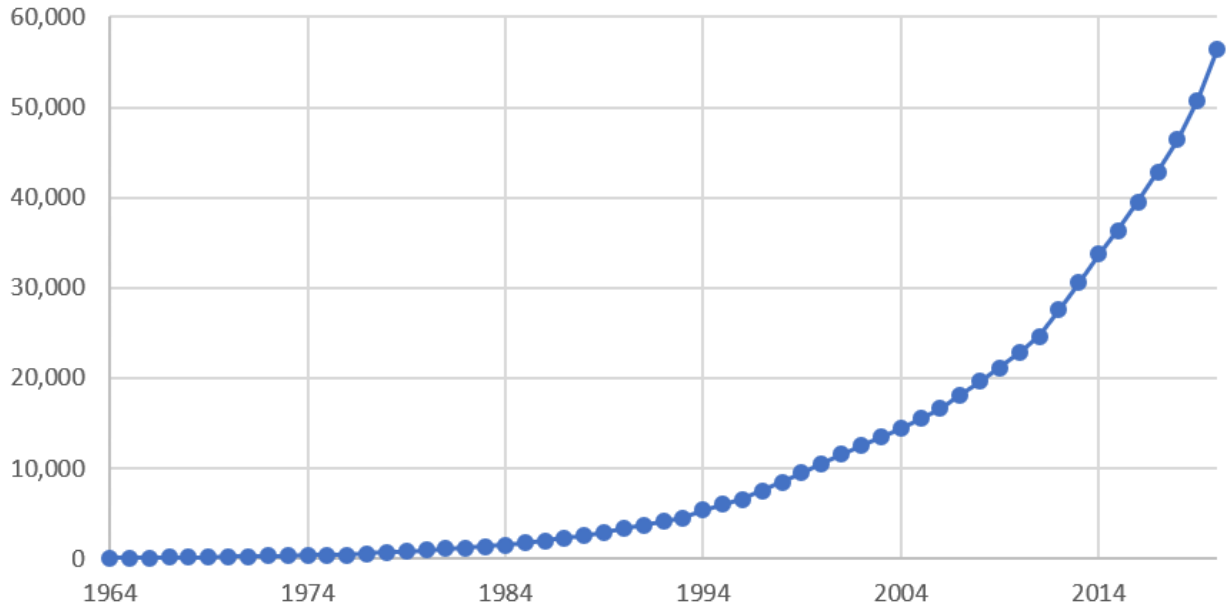
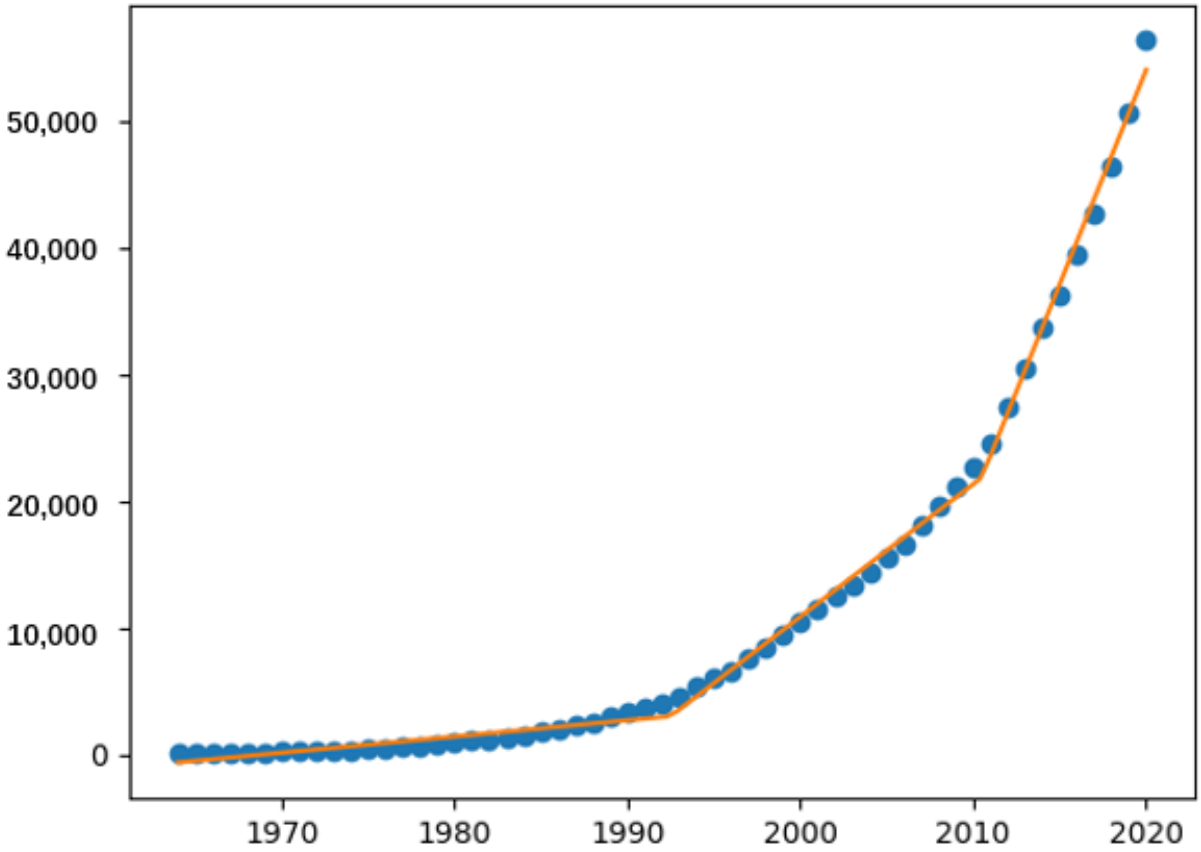


Figure 3. The result of piecewise regression fitting.



The literature growth curve between 1964 and 1991 was difficult to fit any mathematical models. The literature growth curve from 1992 to 2009 (Figure 4) was consistent with the linear increase model, and the adjusted R^2 was 0.988. The literature growth curve from 2010 to 2020 (Figure 5) followed the exponential increase model, and the adjusted R^2 was 0.998. Then, we can summarize the 3 stages of medical informatics: the period from 1964 to 1991 belonged to the early birth stage of medical informatics. There were fewer papers at this point, and the rising speed was unstable. The period of 1992-2009 could be regarded as the early development stage, as the number of papers began to increase and the rate of growth fitted a linear increase model but had not yet reached an exponential increase. Finally, between 2010 and 2020, medical informatics came to a rapid development stage. Some emerging technologies, such as deep learning algorithms and open-source tools for artificial intelligence, have been released and boomed up with the big data era. How to use these technologies in medical informatics has been widely discussed. Therefore, the number of publications increased significantly, and the growth curve followed the exponential increase model.

Figure 4. Results of curve fitting (1992-2009).

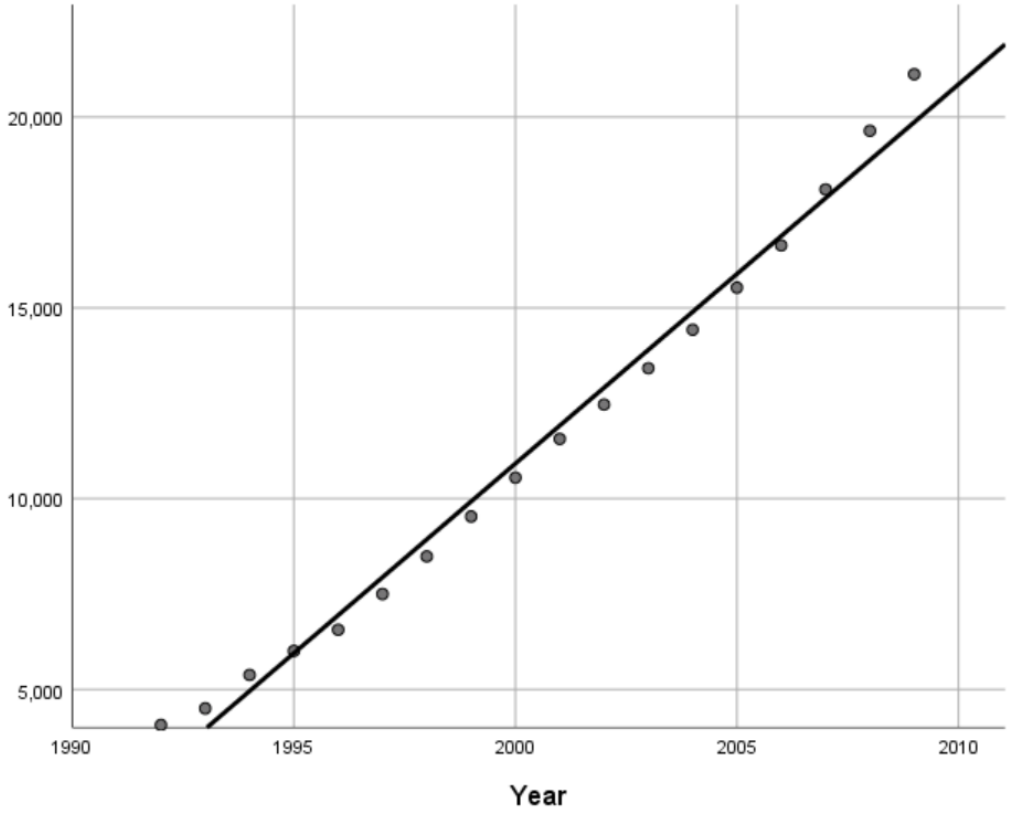
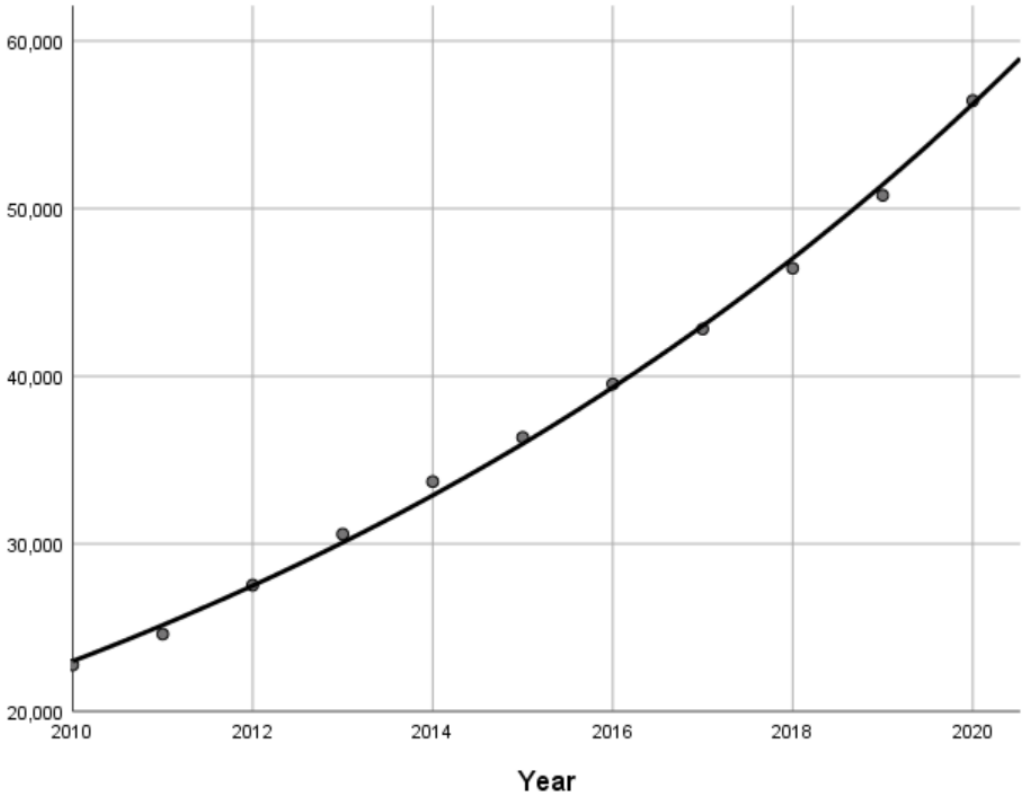


Figure 5. Results of curve fitting (2010-2020).



Topic Evolution Analysis

Overview

We used the LDA model to extract research topics from all corpora and corpora of each stage. As mentioned above, the abstracts of the research articles were chosen as corpora because the abstract, as a paragraph of text, had a clearer semantic logic and a more complete summary of the paper's content, making it more appropriate for LDA-based research topic extraction.

Optimal Topic Number Identification

Perplexity and coherence were calculated to identify the optimal number of topics extracted. Figures 6-9 show the perplexity and coherence curves drawn by Python coding programs.

Perplexity is an index that measures the information generalized by the topic model. A lower perplexity value indicates that the topic model provides more information. Coherence measures the degree of semantic similarity between keywords within a topic. Because topics learned by topic models are not always fully interpretable, coherence is proposed to distinguish between interpretable and artificial topics [14]. A higher coherence score indicates that the topic model offers some meaningful topics. We need to balance perplexity and coherence to choose the optimum number of topics with lower perplexity and higher coherence. We also proposed that higher coherence was more significant because we tended to get more relevant topics.

Figure 6 shows that the optimum number of topics in all corpora was 10, with maximum coherence and minimum perplexity. Figure 7 shows that the coherence reached its maximum when the number of topics was 6, whereas the perplexity was lowest for 7 topics. However, we determined to extract 6 topics from the corpora of stage 1. As seen in Figures 8 and 9, the coherence curve reached the end of the rapid growth when the number of topics was 9. Meanwhile, perplexity was relatively low at 9 topics. We then decided to extract 9 topics from the corpora of stages 2 and 3.

Figure 6. The perplexity and coherence curve of all corpora.

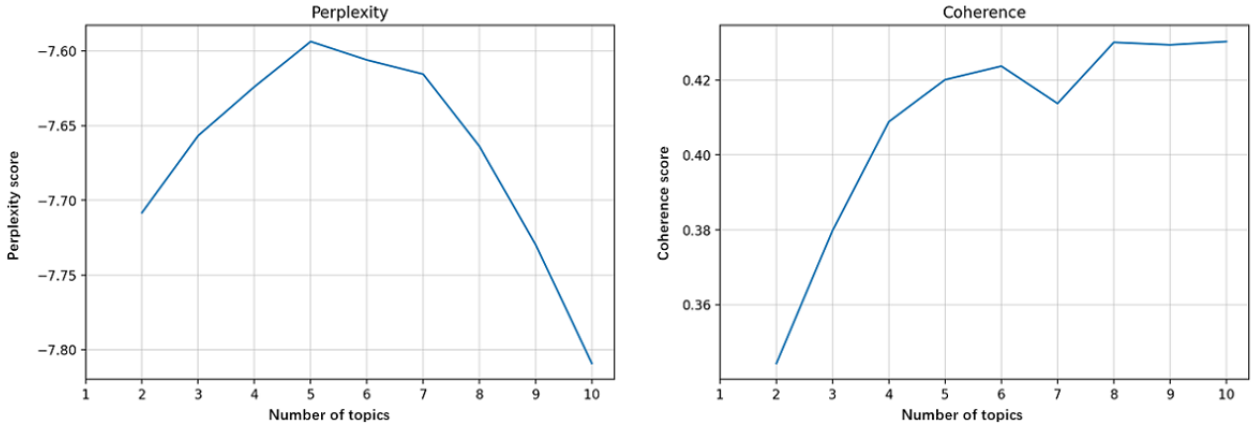


Figure 7. The perplexity and coherence curve of corpora in research stage 1 (1964-1991).

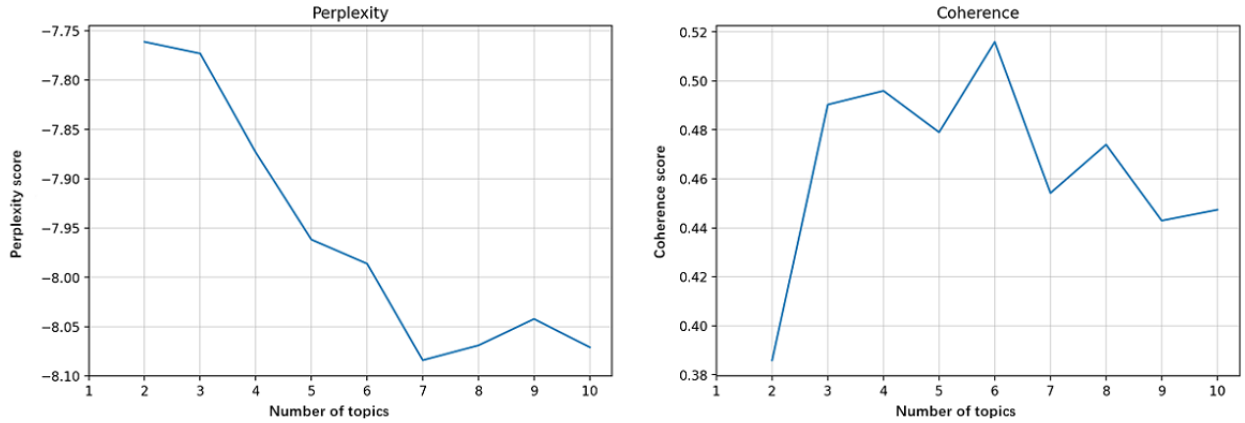


Figure 8. The perplexity and coherence curve of corpora in research stage 2 (1992-2009).

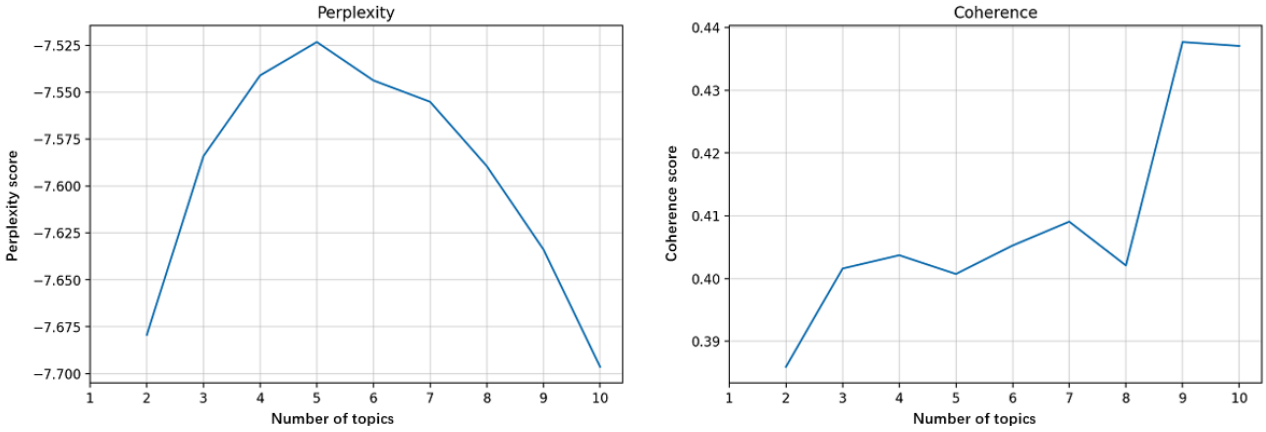
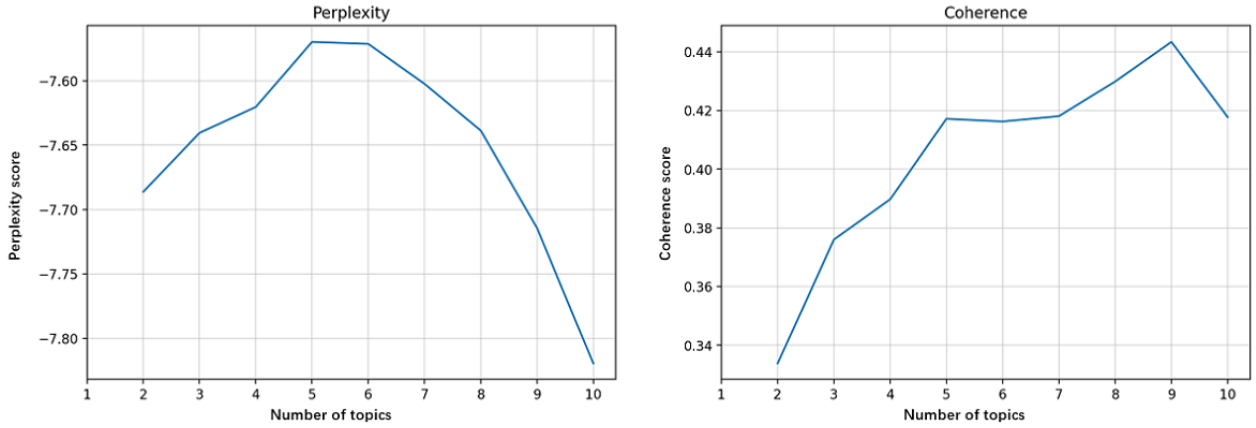


Figure 9. The perplexity and coherence curve of corpora in research stage 3 (2010-2020).



Research Topic Extraction

We adopted the LDA model to extract research topics from the abstracts of 56,466 research articles. The Python library Gensim was used to conduct the LDA model. Gensim is a Python library for topic modeling, document indexing, and similarity retrieval with large corpora. *Alpha* and *beta* are hyperparameters that affect topics' sparsity. According to the Gensim docs, they both default to 1.0/number of topics prior. The number of topics extracted was set to 10, and the top 20 keywords were displayed under each topic. The topic extraction results for all corpora are shown in Table 1.

Table 1 provides an overview of the 10 research hot spots in medical informatics. Topic 1 focused primarily on the medical system, and the keywords under this topic indicate that development and usage, medical system technology, and users' needs are all explored. Topic 2 mainly concerned health-related measurement, with researchers focusing on developing health domain scales, for example, health literacy. Questionnaire design and item optimization are important research questions on this topic. Topic 3 was related to patient care. Physicians, clinicians, treatment, and risk become the top keywords with high weights in this topic. Studies under topic 4 were largely concerned with web-based health information, including the search, use, and evaluation of web-based health information. In addition, user profiling and participation in web-based health communities are hot spots under this topic. Topic 5 can be summarized as medical image processing. Under this topic, researchers were interested in the development and optimization of image-processing algorithms. The keywords under topic 6 were mostly connected to health data analysis. The use of mathematical models and information technologies, such as simulations, in health data analysis has attracted many researchers. Topic 7 was primarily concerned with medication management. Researchers have emphasized drug prescription, dose, safety, and surveillance. Topic 8 emphasized the studies on electronic medical records, especially the management, analysis, and application of medical records. The major research content under topic 9 concerned health interventions. The experimental method is commonly used in health intervention studies. Participants were recruited and divided into groups with different types of interventions. The efficacy of interventions was verified by comparing the performances of various groups. Finally, topic 10 was mainly concerned with the analysis of physiological data collected from patients, such as electroencephalography and heart rate. Relevant keywords include signal, frequency, and flow rate.

We also extracted research hot spots at each stage. Table 2 shows the research topics and top keywords in the 3 stages.

The keywords in stage 1 indicated that the research topics were more medically connected, with a concentration on medical

data analysis. For example, topic 4 mainly focused on the analysis of patients' physiological data (blood, flow, signal, arterial, etc). The analysis and application of data in medical systems was the focus of topic 5 (system, datum, program, etc). Meanwhile, researchers were interested in learning how to analyze the aforementioned medical data. Model, variable, estimation, linear, and other keywords in topic 6 suggested that mathematical models and computational techniques were effective methods in medical data analysis.

Stages 1 and 2 covered some comparable topics, with topics 2 and 4 in stage 2 maintaining the focus on medical system development and medical data analysis methods. Meanwhile, topics in stage 2 revealed some new patterns. For example, the types of medical data were enlarged in the focus of medical data analysis, with medical image processing emerging as a new research hot spot (topic 3 in stage 2). Furthermore, topic 5 suggested that researchers were concerned about the search, application, and users' need for web-based health information. Furthermore, topics in stage 2 revealed that the attention on patients began to increase, such as topic 1, which focused on patient care and treatment, and topic 8, which addressed patients' need to improve medical institutions' services.

Topics in stage 3 inherited the focus on medical data analysis from stage 1 and stage 2, including analysis of medical system data (topic 1), methods of medical data analysis (topic 2), analysis of patients' electronic medical records (topic 3), medical image processing (topic 8), and analysis of disease-related data (topic 9). The keywords in these topics indicated that the goal of medical data analysis is gradually shifting to human-centered, such as improving medical systems based on patients' needs (topic 1), providing better care for patients (topic 3), identifying health risks, and predicting disease for patients (topic 9).

There were a few new topics in stage 3. First, it is worth noting that the development of health tools has become a research hot spot. Topic 4 showed how health tools, such as sensing devices, were used to collect users' physiological data and help them with self-health management. Furthermore, mobile health tools were used for health interventions (topic 5). Meanwhile, as seen in topic 7, which addressed the measurement of health tool usability, user experience has been one of the research hot spots in medical informatics. Finally, the researchers emphasized the importance of standard medical information. The keywords in topic 6 revealed that the construction of concepts, terms, and ontologies became a popular topic in stage 3.

Table 1. Research topics and top keywords in all corpora.

Topics	Keywords
Topic 1	0.026*system, 0.014*information, 0.014*health, 0.012*datum, 0.011*medical, 0.010*care, 0.009*technology, 0.008*process, 0.008*base, 0.008*clinical, 0.007*support, 0.007*user, 0.007*provide, 0.007*use, 0.007*develop, 0.007*development, 0.006*application, 0.006*need, 0.005*implementation, 0.005*tool
Topic 2	0.017*score, 0.016*use, 0.015*measure, 0.013*student, 0.011*scale, 0.010*assess, 0.009*quality, 0.009*test, 0.009*assessment, 0.008*health, 0.008*questionnaire, 0.008*age, 0.008*item, 0.008*group, 0.008*high, 0.007*factor, 0.007*difference, 0.007*mean, 0.006*level, 0.006*year
Topic 3	0.069*patient, 0.020*care, 0.019*cost, 0.017*decision, 0.016*physician, 0.015*treatment, 0.014*clinical, 0.010*practice, 0.008*guideline, 0.008*evidence, 0.007*effectiveness, 0.006*use, 0.006*outcome, 0.006*decision_make, 0.006*benefit, 0.005*clinician, 0.005*primary, 0.005*quality, 0.005*year, 0.005*risk
Topic 4	0.047*health, 0.031*information, 0.013*use, 0.013*internet, 0.011*online, 0.011*search, 0.010*survey, 0.008*relate, 0.007*web, 0.007*user, 0.007*access, 0.006*public, 0.006*identify, 0.006*age, 0.006*population, 0.005*question, 0.005*community, 0.005*source, 0.005*report, 0.005*woman
Topic 5	0.018*image, 0.017*use, 0.014*base, 0.013*propose, 0.012*feature, 0.010*classification, 0.008*performance, 0.008*datum, 0.007*model, 0.007*accuracy, 0.007*algorithm, 0.007*system, 0.006*technique, 0.006*set, 0.006*analysis, 0.006*network, 0.005*detection, 0.005*different, 0.005*show, 0.005*present
Topic 6	0.030*model, 0.017*datum, 0.014*use, 0.013*estimate, 0.011*effect, 0.011*analysis, 0.011*test, 0.010*trial, 0.009*time, 0.008*propose, 0.008*study, 0.007*treatment, 0.007*base, 0.006*simulation, 0.006*distribution, 0.005*parameter, 0.005*compare, 0.005*case, 0.005*outcome, 0.005*variable
Topic 7	0.043*drug, 0.021*medication, 0.013*order, 0.013*dose, 0.012*alert, 0.011*error, 0.011*safety, 0.009*rate, 0.009*report, 0.008*prescription, 0.008*system, 0.008*infection, 0.007*increase, 0.007*surveillance, 0.007*time, 0.007*period, 0.006*use, 0.006*event, 0.006*prescribe, 0.006*identify
Topic 8	0.029*patient, 0.024*datum, 0.018*clinical, 0.018*use, 0.012*record, 0.012*system, 0.011*model, 0.010*hospital, 0.009*medical, 0.007*identify, 0.007*disease, 0.006*base, 0.006*time, 0.006*concept, 0.006*develop, 0.006*database, 0.006*report, 0.006*set, 0.006*code, 0.005*diagnosis
Topic 9	0.026*intervention, 0.020*group, 0.019*participant, 0.011*base, 0.011*app, 0.010*program, 0.010*self, 0.008*use, 0.008*month, 0.007*control, 0.007*health, 0.007*change, 0.007*week, 0.006*behavior, 0.006*follow, 0.006*time, 0.006*user, 0.006*increase, 0.005*day, 0.005*mobile
Topic 10	0.014*use, 0.011*signal, 0.008*model, 0.008*time, 0.007*system, 0.006*measurement, 0.006*parameter, 0.006*frequency, 0.005*show, 0.005*change, 0.005*flow, 0.005*rate, 0.005*subject, 0.005*high, 0.004*analysis, 0.004*heart, 0.004*control, 0.004*increase, 0.004*different, 0.004*measure

Table 2. Research topics and keywords in the 3 stages.

Topics	Keywords
Stage 1 (1964-1991)	
Topic 1	0.022*patient, 0.008*subject, 0.008*risk, 0.008*use, 0.007*record, 0.007*analysis, 0.006*image, 0.006*dose, 0.005*power, 0.005*procedure, 0.005*disease, 0.005*number, 0.005*measure, 0.005*system, 0.005*calculate, 0.005*datum, 0.005*average, 0.005*step, 0.005*present, 0.005*base
Topic 2	0.018*provide, 0.015*medical, 0.015*increase, 0.015*physician, 0.013*include, 0.010*several, 0.009*use, 0.008*year, 0.008*report, 0.008*risk, 0.008*practice, 0.007*patient, 0.007*model, 0.007*analysis, 0.007*probability, 0.006*condition, 0.006*factor, 0.006*investigate, 0.006*value, 0.006*heart
Topic 3	0.015*test, 0.014*clinical, 0.012*trial, 0.012*estimate, 0.011*analysis, 0.010*model, 0.009*diagnostic, 0.008*treatment, 0.008*compare, 0.008*medical, 0.008*problem, 0.008*rate, 0.007*decision, 0.007*present, 0.006*population, 0.006*development, 0.006*need, 0.006*effect, 0.006*statistical, 0.006*multiple
Topic 4	0.017*blood, 0.017*measurement, 0.016*flow, 0.016*model, 0.015*analysis, 0.014*pressure, 0.013*use, 0.011*electrode, 0.010*signal, 0.008*measure, 0.008*human, 0.007*effect, 0.006*impedance, 0.006*spectral, 0.006*arterial, 0.006*volume, 0.005*parameter, 0.005*frequency, 0.005*distribution, 0.005*skin
Topic 5	0.042*system, 0.027*datum, 0.013*clinical, 0.012*information, 0.011*program, 0.011*use, 0.011*medical, 0.011*knowledge, 0.010*computer, 0.010*base, 0.009*analysis, 0.009*image, 0.009*develop, 0.007*management, 0.007*describe, 0.007*process, 0.006*processing, 0.006*trial, 0.006*procedure, 0.005*study
Topic 6	0.019*use, 0.019*model, 0.010*time, 0.009*study, 0.008*control, 0.008*propose, 0.008*variable, 0.008*individual, 0.008*analysis, 0.007*first, 0.006*describe, 0.006*datum, 0.006*make, 0.006*number, 0.006*non, 0.005*estimation, 0.005*response, 0.005*linear, 0.005*examine, 0.005*base
Stage 2 (1992-2009)	
Topic 1	0.032*patient, 0.018*use, 0.012*diagnosis, 0.011*diagnostic, 0.011*datum, 0.010*classification, 0.009*test, 0.009*case, 0.009*accuracy, 0.008*performance, 0.008*model, 0.008*hospital, 0.007*sensitivity, 0.007*clinical, 0.007*system, 0.007*set, 0.006*compare, 0.006*disease, 0.006*rate, 0.006*prediction
Topic 2	0.027*system, 0.018*datum, 0.015*medical, 0.012*information, 0.012*base, 0.011*use, 0.011*clinical, 0.010*model, 0.009*knowledge, 0.008*application, 0.007*develop, 0.007*describe, 0.007*support, 0.006*provide, 0.006*process, 0.006*concept, 0.005* software, 0.005*present, 0.005*database, 0.005*tool
Topic 3	0.018*image, 0.017*use, 0.013*signal, 0.008*time, 0.008*analysis, 0.007*base, 0.007*system, 0.007*technique, 0.006*frequency, 0.005*obtain, 0.005*show, 0.005*present, 0.005*feature, 0.005*high, 0.005*measurement, 0.005*subject, 0.004*parameter, 0.004*noise, 0.004*different, 0.004*measure
Topic 4	0.025*model, 0.016*datum, 0.014*use, 0.014*test, 0.011*analysis, 0.011*estimate, 0.011*effect, 0.010*trial, 0.008*propose, 0.007*treatment, 0.007*time, 0.007*base, 0.007*study, 0.006*distribution, 0.006*parameter, 0.005*compare, 0.005*simulation, 0.005*error, 0.005*clinical, 0.005*procedure
Topic 5	0.018*information, 0.011*use, 0.009*health, 0.009*evidence, 0.009*report, 0.009*search, 0.008*user, 0.008*clinical, 0.008*internet, 0.007*evaluation, 0.007*identify, 0.007*base, 0.006*assessment, 0.006*question, 0.006*study, 0.006*technology, 0.005*web, 0.005*quality, 0.005*medical, 0.005*provide
Topic 6	0.018*model, 0.011*cell, 0.009*use, 0.008*gene, 0.008*dose, 0.007*tissue, 0.007*increase, 0.006*flow, 0.006*pressure, 0.006*blood, 0.006*change, 0.005*drug, 0.005*control, 0.005*measure, 0.004*show, 0.004*current, 0.004*high, 0.004*response, 0.004*experimental, 0.004*value
Topic 7	0.028*patient, 0.018*cost, 0.014*treatment, 0.011*health, 0.010*use, 0.009*group, 0.008*risk, 0.007*measure, 0.007*intervention, 0.007*decision, 0.007*year, 0.006*analysis, 0.006*quality, 0.006*compare, 0.006*score, 0.006*high, 0.006*utility, 0.006*outcome, 0.005*life, 0.005*state
Topic 8	0.026*health, 0.024*care, 0.020*patient, 0.015*system, 0.013*information, 0.009*medical, 0.009*practice, 0.008*physician, 0.008*hospital, 0.008*technology, 0.007*service, 0.007*clinical, 0.006*base, 0.006*computer, 0.006*use, 0.005*need, 0.005*management, 0.005*record, 0.005*support, 0.004*implementation
Topic 9	0.021*model, 0.021*disease, 0.019*risk, 0.017*datum, 0.015*estimate, 0.013*time, 0.011*population, 0.011*age, 0.011*rate, 0.010*use, 0.009*exposure, 0.009*case, 0.008*incidence, 0.008*cancer, 0.007*infection, 0.007*child, 0.007*mortality, 0.006*year, 0.005*individual, 0.005*prevalence
Stage 3 (2010-2020)	
Topic 1	0.023*health, 0.018*datum, 0.015*information, 0.012*system, 0.008*technology, 0.008*clinical, 0.008*medical, 0.007*process, 0.007*provide, 0.007*use, 0.007*support, 0.007*care, 0.006*base, 0.006*need, 0.006*development, 0.005*develop, 0.005*service, 0.005*patient, 0.005*user, 0.005*healthcare
Topic 2	0.031*model, 0.016*datum, 0.014*use, 0.012*effect, 0.011*estimate, 0.011*treatment, 0.010*trial, 0.010*analysis, 0.009*time, 0.008*propose, 0.008*study, 0.008*test, 0.007*outcome, 0.006*base, 0.006*simulation, 0.005*compare, 0.005*risk, 0.004*clinical, 0.004*show, 0.004*variable

Topics	Keywords
Topic 3	0.056*patient, 0.024*care, 0.016*system, 0.016*hospital, 0.012*physician, 0.010*use, 0.009*clinical, 0.009*electronic, 0.009*medication, 0.009*record, 0.008*time, 0.008*information, 0.008*health, 0.007*medical, 0.006*ehr, 0.006*provider, 0.006*improve, 0.006*practice, 0.006*quality, 0.005*decision
Topic 4	0.013*use, 0.009*model, 0.007*time, 0.007*patient, 0.007*system, 0.006*measurement, 0.006*parameter, 0.006*subject, 0.006*show, 0.005*control, 0.005*rate, 0.005*change, 0.005*device, 0.005*measure, 0.005*sensor, 0.005*high, 0.004*activity, 0.004*signal, 0.004*heart, 0.004*increase
Topic 5	0.020*health, 0.017*intervention, 0.015*participant, 0.013*group, 0.010*use, 0.010*app, 0.009*base, 0.007*self, 0.007*online, 0.007*internet, 0.006*user, 0.006*conclusion, 0.005*report, 0.005*web, 0.005*behavior, 0.005*high, 0.005*program, 0.005*information, 0.005*increase, 0.005*treatment
Topic 6	0.012*system, 0.012*model, 0.011*use, 0.011*clinical, 0.011*concept, 0.008*base, 0.007*term, 0.007*medical, 0.007*text, 0.007*semantic, 0.007*ontology, 0.007*biomedical, 0.007*knowledge, 0.006*information, 0.006*cell, 0.006*structure, 0.006*domain, 0.006*query, 0.005*different, 0.005*document
Topic 7	0.016*use, 0.013*score, 0.013*usability, 0.012*student, 0.011*test, 0.010*base, 0.010*user, 0.009*item, 0.009*evaluation, 0.009*training, 0.008*tool, 0.008*assessment, 0.008*group, 0.008*assess, 0.008*evaluate, 0.008*questionnaire, 0.007*develop, 0.007*scale, 0.007*nursing, 0.007*task
Topic 8	0.019*propose, 0.019*use, 0.018*image, 0.015*feature, 0.014*base, 0.011*classification, 0.010*performance, 0.009*accuracy, 0.007*datum, 0.007*detection, 0.007*algorithm, 0.007*model, 0.006*technique, 0.006*system, 0.006*signal, 0.006*show, 0.005*analysis, 0.005*classifier, 0.005*high, 0.005*dataset
Topic 9	0.021*datum, 0.020*disease, 0.017*use, 0.015*drug, 0.014*identify, 0.011*cancer, 0.011*risk, 0.010*clinical, 0.010*patient, 0.008*base, 0.008*gene, 0.007*diagnosis, 0.007*develop, 0.007*set, 0.006*predict, 0.006*case, 0.006*high, 0.006*prediction, 0.005*record, 0.005*accuracy

Topic Evolution Pattern Construction

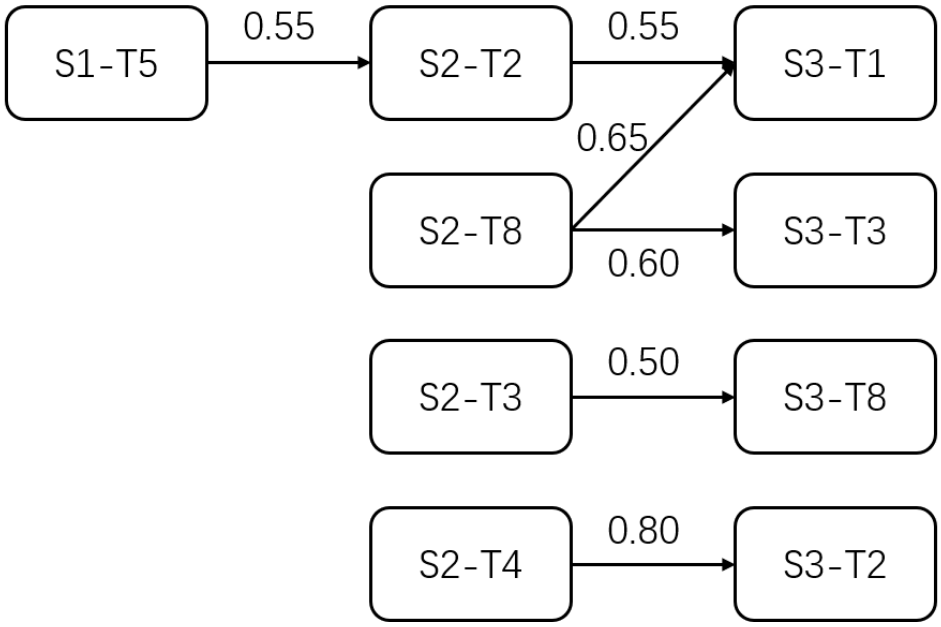
As previously stated, there were several research topics that were comparable between 2 adjacent stages. To determine the evolution pattern, we used the Python coding program to calculate the cosine similarity of keywords between 2 research topics from 2 adjacent stages. A total of 2 topics were connected if the cosine similarity between them was more than 0.5. Figure 6 illustrates the connections between topics from stages 1 to 3. Here, S1-T5 refers to topic 5 in stage 1.

Figure 10 shows that the connections between stage 1 and stage 2 were weaker than those between stage 2 and stage 3. The reason for this could be that, in the early stage of medical informatics, there was less research literature and the focus of these studies was primarily on the medical field, whereas as medical informatics developed, research became more interdisciplinary as knowledge and research methods from other fields, such as computer science, library science, and psychology, were integrated into medical informatics. Therefore, research topics in stages 2 and 3 were more diverse and less similar to those in stage 1.

There was an evolution line from stage 1 to stage 3, starting at topic 5 in stage 1, moving through topic 2 in stage 2, and ending at topic 1 in stage 3. The focus of these topics was mainly on medical systems, with the difference that topic 5 in stage 1 and topic 2 in stage 2 concentrated more on technologies for medical system development and optimization, such as software and database construction, whereas topic 1 in stage 3 addressed the user needs to improve the service of the medical system.

There were several evolution lines between topics in stages 2 and 3. First, topic 8 in stage 2 was split into topic 1 and topic 3 in stage 3. The keywords of topic 8 in stage 2 emphasized the importance of patient needs. As a result, topic 1 in stage 3 evaluated patient needs in the progress of medical system development, and ’topic 3 in stage 3 considered patient needs in the improvement of health care service. Second, topic 8 in stage 3 was inherited from topic 3 in stage 2, indicating that medical image processing has been one of the research hot spots in medical informatics since the 1990s. Finally, topic 4 in stage 2 evolved into topic 2 in stage 3, with the focus of this evolution line being primarily on methods of medical data analysis. Researchers have been working hard to develop efficient methods for analyzing medical data, such as using mathematical models and constructing computing algorithms.

Figure 10. Research topics' evolution patterns.



Discussion

Principal Findings

This study explored the research stages, research hot spots, and their evolution patterns in medical informatics. We found that medical informatics has gone through three stages: (1) the early birth stage (1964-1991), with a small number of papers and an unstable growth speed; (2) the early development stage (1992-2009), with an increasing number of papers and a steadily rising speed; and (3) the fast development stage (2010-2020), with a large number of papers and an exponential growth speed.

In the first stage (1964-1991), researchers focused on medical data analysis, including the analysis of patients' physiological data, such as pulmonary data [18], cerebrum data [19], and renal data [20], as well as the analysis of medical images, such as electroencephalogram [21] and electromyography [22]. Medical data analysis studies in this period served a primary role in in the field of medicine, such as providing therapy for patients or assisting physicians with disease diagnosis. In addition, methodologies and technologies used in medical data analysis became a research hot spot in this period. Researchers used some mathematical models (regression [23], Bayesian [24,25], Markov [26], etc) and computer technologies (database [27], information system [28], simulation [29], etc) to improve the efficiency and precision of medical data analysis.

In the second stage (1992-2009), research topics inherited features from the previous stage while also developing new ones. First, research topics in the second phase maintained the focus on medical data analysis and its related methodologies and technologies [30-32]. Medical image processing became a dependent hot spot, indicating that studies on medical image processing grew rapidly during this period [33-35]. Furthermore, as medical informatics became increasingly interdisciplinary, studies were no longer limited to analyzing data from medical institutions or medical systems. Web-based health information also attracted the attention of researchers, including studies on internet users' information behavior (search [36], application [37], and evaluation [38] of web-based health information). Finally, the topics in stage 2 reflected the shift in emphasis from data to people, with more studies aimed at meeting patients' health care needs [39-41] and improving users' satisfaction [42,43].

In the third stage (2010-2020), medical data analysis remained one of the research hot spots. Derived from topics in stage 2, the purpose of medical informatics research always took user needs into account, including the needs of patients [44] and doctors [45]. Meanwhile, studies in this period also paid more attention to applying new emerging technologies in health data analysis, such as deep learning [46], blockchain [47], and artificial intelligence [48]. Furthermore, with the growing use of smartphones and wearables, a variety of health tools have enabled users to generate their own private health logs and manage their health conditions, such as weight control [49], chronic disease treatment [50], and mental health management [51]. Particularly during the COVID-19 pandemic, the use of digital health tools to provide health care and mental support for people became a significant issue [52]. However, as mobile health tools such as health apps have become widely used, researchers should pay attention to emerging problems such as the digital divide [53] and the patients' privacy disclosure [54], especially older adults' acceptance of information and communications technology [55].

On the basis of the results of research topic extraction in all corpora, we concluded that the focus of research in medical informatics could be divided into two aspects: data-centered studies and people-centered studies. In data-centered studies, medical records, medical images, and disease data were analyzed, which used mathematical methods and computing technologies to increase the efficiency and precision of data analysis. People-centered studies emphasized user needs and satisfaction, intending to improve health care service and health tool usability.

Furthermore, topic evolution patterns revealed that medical data analysis has always been a research hot spot since the beginning of medical informatics, particularly the methods and technologies used in data analysis. This is consistent with the results of previous studies [9,56]. The reason for this might be attributed to the development of emerging technologies, which prompted the exploration of data analysis methods. We could infer that future medical informatics research will continue to focus on the application of emerging technologies, such as deep learning, artificial intelligence, and blockchain, in medical data analysis. The topic evolution patterns also showed that people-centered topics arose in the second stage and were integrated with data-centered topics in the third stage. This tendency may be emphasized in future medical informatics studies. As mentioned previously, people-centered studies have considered user needs and satisfaction. It is possible that the usability of health tools such as health apps and wearables, as well as their effect on health behavior intervention, could be important issues for future research.

Limitations

There are several limitations to this study. First, the Web of Science database did not index the abstracts of all papers, especially those in the early stage. As a result, we might have missed some topics in the research topic extraction. Second, we chose 27 representative journals in medical informatics without regard to the journals' starting years. Journals that started in the earlier period would cover different topics from later ones, which might influence topic extraction results. Finally, while identifying the research stages, we only considered the annual cumulative number of research papers according to the literature growth curve of Price [10]. The journal amount, paper work, and web-based submission were also important indexes to consider when determining research stages.

Comparison With Prior Work

We reviewed the development history of medical informatics from 1964 to 2020. Previous literature reviews have mostly focused on papers published within the last 10 to 20 years [3]. By contrast, our study attempted to provide a comprehensive review of medical informatics based on the results of a thorough survey.

In previous studies, research stages were usually divided intuitively based on the annual number of papers curve, with no quantitative model fitting [9]. In our study, we used the piecewise regression model to fit the curve of the annual cumulative number of papers to identify the research stages. We also used several mathematical models to fit curves in different stages to determine the literature growth features of each stage. We find that medical informatics is at a fast development stage, with an exponential increase in the literature. In fact, medical informatics has attracted research interest from various fields. Our findings are consistent with the current situation.

Previous studies that extracted research topics in medical informatics simply discussed and summarized the content of these topics [56]. In this study, we further divided the research topics into data- and people-centered topics. Furthermore, we found an integration tendency between these 2 types of topics according to their evolution patterns. However, previous studies have only emphasized the importance of medical data analysis [9].

Conclusions

Our study offers a comprehensive understanding of research hot spots and their evolution patterns in medical informatics, and it could be helpful for predicting future research trends in this field. We found that medical informatics was in the fast development stage, with rapid growth in the literature. Medical data analysis has always been an important research topic since the birth of medical informatics to the current developmental stage. Many researchers are interested in data analysis methodologies and technologies, such as mathematical models and computer science technologies. In addition, the concentration of medical data has shifted from data to people. Recent studies have focused on improving medical systems and health tools, such as how to deliver better patient care and how to support users' self-health management. We predicted that the application of emerging computer technologies in medical data analysis and the usability of mobile health tools would become a research hot spots in future medical informatics studies.

Acknowledgments

This study was supported by the Guangdong Basic and Applied Basic Research Foundation (2021A1515011956) and the National Natural Science Foundation of China (72174083). The authors would like to thank all the anonymous reviewers and editors for their valuable comments.

Authors' Contributions

WH analyzed the data and drafted the study. XH contributed to the research design. SZ reviewed the paper. QZ conducted the research focus into medical informatics. All authors contributed to the study's preparation and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Haux R. Medical informatics: past, present, future. Int J Med Inform 2010 Sep;79(9):599-610. [doi: 10.1016/j.ijmedinf.2010.06.003] [Medline: 20615752]
2. Ledley RS, Lusted LB. Reasoning Foundations of Medical Diagnosis: symbolic logic, probability, and value theory aid our understanding of how physicians reason. Science 1959 Jul 03;130(3366):9-21. [doi: 10.1126/science.130.3366.9] [Medline: 13668531]
3. Saheb T, Saheb M. Analyzing and visualizing knowledge structures of health informatics from 1974 to 2018: a bibliometric and social network analysis. Healthc Inform Res 2019 Apr;25(2):61-72 [FREE Full text] [doi: 10.4258/hir.2019.25.2.61] [Medline: 31131140]
4. Kim Y, Delen D. Medical informatics research trend analysis: a text mining approach. Health Informatics J 2018 Dec;24(4):432-452 [FREE Full text] [doi: 10.1177/1460458216678443] [Medline: 30376768]
5. Park HA, Lee JY, On J, Lee JH, Jung H, Park SK. 2016 year-in-review of clinical and consumer informatics: analysis and visualization of keywords and topics. Healthc Inform Res 2017 Apr;23(2):77-86 [FREE Full text] [doi: 10.4258/hir.2017.23.2.77] [Medline: 28523205]
6. Wang L, Topaz M, Plasek JM, Zhou L. Content and trends in medical informatics publications over the past two decades. Stud Health Technol Inform 2017;245:968-972. [doi: 10.3233/978-1-61499-830-3-968] [Medline: 29295244]
7. Lyu PH, Yao Q, Mao J, Zhang SJ. Emerging medical informatics research trends detection based on MeSH terms. Inform Health Soc Care 2015;40(3):210-228. [doi: 10.3109/17538157.2014.892490] [Medline: 25188551]
8. Deng S, Xia S, Hu J, Li H, Liu Y. Exploring the topic structure and evolution of associations in information behavior research through co-word analysis. J Librariansh Inf Sci 2020 Jul 23;53(2):280-297. [doi: 10.1177/0961000620938120]
9. Feng J, Mu X, Li F, Shen Y, Wang W, Xu Y. Analyzing research fronts in medical informatics field based on topic model and life cycle theory. Res Square 2020. [doi: 10.21203/rs.2.22157/v1]
10. Price D. Science since Babylon. Am J Phys 1961 Dec;29(12):863-864. [doi: 10.1119/1.1937650]
11. Blei DM, Ng AY, Jordan ML. Latent dirichlet allocation. J Mach Learn Res 2003;1:993-1022 [FREE Full text] [doi: 10.1162/jmlr.2003.3.4-5.993]
12. Hassan SU, Haddawy P. Analyzing knowledge flows of scientific literature through semantic links: a case study in the field of energy. Scientometrics 2015 Jan;103(1):33-46. [doi: 10.1007/s11192-015-1528-3]
13. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr;55(4):77-84. [doi: 10.1145/2133806.2133826]
14. Kapadia S. Evaluate topic models: latent dirichlet allocation (LDA). Towards Data Science. 2019. URL: https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0 [accessed 2021-09-26]
15. Hurtado JL, Agarwal A, Zhu X. Topic discovery and future trend forecasting for texts. J Big Data 2016 Apr 14;3(1):7. [doi: 10.1186/s40537-016-0039-2]
16. Chen B, Tsutsui S, Ding Y, Ma F. Understanding the topic evolution in a scientific domain: an exploratory study for the field of information retrieval. J Informetr 2017 Nov;11(4):1175-1189. [doi: 10.1016/j.joi.2017.10.003]
17. Guan P, Wang Y, Fu Z. Analyzing topic semantic evolution with LDA: case study of lithium ion batteries. Data analysis and knowledge discovery. Data Anal Knowl Discov 2019;3(7):61-72 [FREE Full text] [doi: 10.11925/infotech.2096-3467.2018.1404]
18. Huygen P, Feenstra B, Hoorn E, Jansen J, Zwart A. PDPS: a pulmonary data processing system for assessment of gas exchange properties by multiple gas wash-out. Comput Methods Programs Biomed 1991 Dec;36(4):223-235. [doi: 10.1016/0169-2607(91)90091-7] [Medline: 1663442]
19. Stassen HH. The octave approach to EEG analysis. Methods Inf Med 1991 Oct;30(4):304-310. [Medline: 1762585]
20. Speiman FA, Öberg PA. Continuous measurement of renal cortical blood flow and renal arterial blood flow during stimulation of the renal nerve. Med Biolog Eng Comput 1991 Mar;29(2):121-128. [doi: 10.1007/bf02447096]
21. Dunstan FD, Marshall RW. The detection of artefacts in EEG series. Stat Med 1991 Nov;10(11):1719-1731. [doi: 10.1002/sim.4780101110] [Medline: 1792466]
22. Stålberg E, Stålberg S, Melander M, Arimura K. A personal computer based system used in electromyography for interpretation and reporting. Comput Methods Programs Biomed 1991;34(2-3):219-227. [doi: 10.1016/0169-2607(91)90046-v] [Medline: 2060293]
23. Liu X, Liang K. Adjustment for non-differential misclassification error in the generalized linear model. Stat Med 1991 Aug;10(8):1197-1211. [doi: 10.1002/sim.4780100804] [Medline: 1925152]
24. Herskovits EH, Cooper GF. Algorithms for Bayesian belief-network precomputation. Methods Inf Med 1991;30(02):81-89. [doi: 10.1055/s-0038-1634820]
25. Mollie A, Richardson S. Empirical Bayes estimates of cancer mortality rates using spatial models. Stat Med 1991 Jan;10(1):95-112. [doi: 10.1002/sim.4780100114] [Medline: 2006359]
26. Munro AJ, Warde PR. The use of a Markov process to simulate and assess follow-up policies for patients with malignant disease: surveillance for stage I nonseminomatous tumors of the testis. Med Decis Making 1991;11(2):131-139. [doi: 10.1177/0272989X9101100211] [Medline: 1907708]
27. Rouan G, Marks E, Tuchfarber B, Redington TJ, Lorenzi N. Development, effects, and educational outcomes of reporting from a clinical database. Med Decis Making 1991;11(4_suppl):80-89. [doi: 10.1177/0272989X9101104s16] [Medline: 1663196]

28. Finkelstein MW, Johnson LA, Lilly GE. A computer management system for patient simulations. Comput Methods Programs Biomed 1991 Apr;34(4):257-261. [doi: 10.1016/0169-2607(91)90109-7]
29. Malik M, Camm AJ. Computer simulation of electronic interactions during excitation and repolarisation of myocardial tissue. Med Biolog Eng Comput 1991 Jul;29(4):425-432. [doi: 10.1007/bf02441665]
30. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997 May 15;16(9):965-980. [doi: 10.1002/(sici)1097-0258(19970515)16:9<965::aid-sim509>3.0.co;2-o]
31. Friedman JH, Meulman JJ. Multiple additive regression trees with application in epidemiology. Stat Med 2003 May 15;22(9):1365-1381. [doi: 10.1002/sim.1501] [Medline: 12704603]
32. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993;13(4):322-338. [doi: 10.1177/0272989X9301300409] [Medline: 8246705]
33. Lehmann T, Aach T, Witte H. Sensor, signal and image informatics. Yearb Med Inform 2006;15(01):57-67. [doi: 10.1055/s-0038-1638479]
34. Zhuang X, Meng Q. Local fuzzy fractal dimension and its application in medical image processing. Artif Intell Med 2004 Sep;32(1):29-36. [doi: 10.1016/j.artmed.2004.01.016] [Medline: 15350622]
35. Gupta S, Chauhan RC, Sexana SC. Wavelet-based statistical approach for speckle reduction in medical ultrasound images. Med Biol Eng Comput 2004 Mar;42(2):189-192. [doi: 10.1007/bf02344630]
36. Hoogendam A, Stalenhoef AF, Robbé PF, Overbeke AJ. Answers to questions posed during daily patient care are more likely to be answered by UpToDate than PubMed. J Med Internet Res 2008 Oct 03;10(4):e29 [FREE Full text] [doi: 10.2196/jmir.1012] [Medline: 18926978]
37. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008 May 27;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
38. Breckons M, Jones R, Morris J, Richardson J. What do evaluation instruments tell us about the quality of complementary medicine information on the internet? J Med Internet Res 2008 Jan 22;10(1):e3 [FREE Full text] [doi: 10.2196/jmir.961] [Medline: 18244894]
39. Bock BC, Graham AL, Whiteley JA, Stoddard JL. A review of web-assisted tobacco interventions (WATIs). J Med Internet Res 2008 Nov 06;10(5):e39 [FREE Full text] [doi: 10.2196/jmir.989] [Medline: 19000979]
40. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008 May 15;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]
41. Rizo CA, Lupea D, Baybourdy H, Anderson M, Closson T, Jadad AR. What Internet services would patients like from hospitals during an epidemic? Lessons from the SARS outbreak in Toronto. J Med Internet Res 2005 Aug 03;7(4):e46 [FREE Full text] [doi: 10.2196/jmir.7.4.e46] [Medline: 16236698]
42. Holmes-Rovner M, Kroll J, Schmitt N, Rovner DR, Breer ML, Rothert ML, et al. Patient satisfaction with health care decisions: the satisfaction with decision scale. Med Decis Making 1996;16(1):58-64. [doi: 10.1177/0272989X9601600114] [Medline: 8717600]
43. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
44. James G, Nyman E, Fitz-Randolph M, Niklasson A, Hedman K, Hedberg J, et al. Characteristics, symptom severity, and experiences of patients reporting chronic kidney disease in the PatientsLikeMe online health community: retrospective and qualitative study. J Med Internet Res 2020 Jul 15;22(7):e18548 [FREE Full text] [doi: 10.2196/18548] [Medline: 32673242]
45. Maassen O, Fritsch S, Gantner J, Deffge S, Kunze J, Marx G, et al. Future mobile device usage, requirements, and expectations of physicians in German University hospitals: web-based survey. J Med Internet Res 2020 Dec 21;22(12):e23955 [FREE Full text] [doi: 10.2196/23955] [Medline: 33346735]
46. Jeong YU, Yoo S, Kim Y, Shim WH. De-identification of facial features in magnetic resonance images: software development using deep learning technology. J Med Internet Res 2020 Dec 10;22(12):e22739 [FREE Full text] [doi: 10.2196/22739] [Medline: 33208302]
47. Cadoret D, Kailas T, Velmovitsky P, Morita P, Igboeli O. Proposed implementation of Blockchain in British Columbia's health care data management. J Med Internet Res 2020 Oct 23;22(10):e20897 [FREE Full text] [doi: 10.2196/20897] [Medline: 33095183]
48. Ko H, Chung H, Kang WS, Park C, Kim DW, Kim SE, et al. An artificial intelligence model to predict the mortality of COVID-19 patients at hospital admission time using routine blood samples: development and validation of an ensemble model. J Med Internet Res 2020 Dec 23;22(12):e25442 [FREE Full text] [doi: 10.2196/25442] [Medline: 33301414]
49. Turner-McGrievy GM, Beets MW, Moore JB, Kaczynski AT, Barr-Anderson DJ, Tate DF. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. J Am Med Inform Assoc 2013 May;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]

50. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov 13;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
51. Torous J, Friedman R, Keshavan M. Metadata correction: smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR mHealth uHealth 2014 Jul 07;2(3):e34 [FREE Full text] [doi: 10.2196/mhealth.3668]
52. Petracca F, Ciani O, Cucciniello M, Tarricone R. Harnessing digital health technologies during and after the COVID-19 pandemic: context matters. J Med Internet Res 2020 Dec 30;22(12):e21815 [FREE Full text] [doi: 10.2196/21815] [Medline: 33351777]
53. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. J Am Med Inform Assoc 2017 Nov;24(6):1142-1148 [FREE Full text] [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
54. Chiauzzi E, Wicks P. Digital trespass: ethical and terms-of-use violations by researchers accessing data from an online patient community. J Med Internet Res 2019 Feb 21;21(2):e11985 [FREE Full text] [doi: 10.2196/11985] [Medline: 30789346]
55. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):209-231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
56. Feng J, Mu X, Wang W, Xu Y. A topic analysis method based on a three-dimensional strategic diagram. J Inform Sci 2020 Jun 24;47(6):770-782. [doi: 10.1177/0165551520930907]

Abbreviations

LDA: latent Dirichlet allocation
MeSH: Medical Subject Headings

Edited by M Focsa; submitted 09.07.21; peer-reviewed by J Plasek, G Kolostoumpis, H Lin; comments to author 02.09.21; revised version received 15.11.21; accepted 19.12.21; published 27.01.22.

Please cite as:
Han W, Han X, Zhou S, Zhu Q
The Development History and Research Tendency of Medical Informatics: Topic Evolution Analysis
JMIR Med Inform 2022;10(1):e31918
URL: https://medinform.jmir.org/2022/1/e31918
doi:10.2196/31918
PMID:35084351

Original Paper

A Platform and Multisided Market for Translational, Software-Defined Medical Procedures in the Operating Room (OP 4.1): Proof-of-Concept Study

Magdalena Görtz[1], MD, Dr med; Michael Byczkowski[2], Dipl-Math; Mathias Rath[1], MD; Viktoria Schütz[1], MD, Dr med; Philipp Reimold[1], MD, Dr med; Claudia Gasch[1], MD, Dr med; Tobias Simpfendörfer[1], MD, Dr med; Keno März[3], PhD; Alexander Seitel[3], PhD; Marco Nolden[3], PhD; Tobias Ross[3]; Diana Mindroc-Filimon[3]; Dominik Michael[3]; Jasmin Metzger[3]; Sinan Onogur[3]; Stefanie Speidel[4], Prof Dr; Lars Mündermann[5]; Johannes Fallert[5], PhD; Michael Müller[6], PhD; Magnus von Knebel Doeberitz[7], Prof Dr; Dogu Teber[8], Prof Dr; Peter Seitz[9]; Lena Maier-Hein[3], Prof Dr; Stefan Duensing[10], Prof Dr; Markus Hohenfellner[1], Prof Dr

[1]Department of Urology, Heidelberg University Hospital, Heidelberg, Germany

[2]SAP SE, Walldorf, Germany

[3]German Cancer Research Center, Heidelberg, Germany

[4]National Center for Tumor Diseases, Dresden, Germany

[5]KARL STORZ SE & Co. KG, Tuttlingen, Germany

[6]mbits imaging GmbH, Heidelberg, Germany

[7]Department of Applied Tumor Biology, Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany

[8]Department of Urology, Städtisches Klinikum Karlsruhe, Karlsruhe, Germany

[9]Siemens Healthineers AG, Forchheim, Germany

[10]Section of Molecular Urooncology, Department of Urology, University of Heidelberg School of Medicine, Heidelberg, Germany

Corresponding Author:
Magdalena Görtz, MD, Dr med
Department of Urology
Heidelberg University Hospital
Im Neuenheimer Feld 420
Heidelberg, 69120
Germany
Phone: 49 6221568820
Email: magdalena.goertz@med.uni-heidelberg.de

Abstract

Background: Although digital and data-based technologies are widespread in various industries in the context of Industry 4.0, the use of smart connected devices in health care is still in its infancy. Innovative solutions for the medical environment are affected by difficult access to medical device data and high barriers to market entry because of proprietary systems.

Objective: In the proof-of-concept project OP 4.1, we show the business viability of connecting and augmenting medical devices and data through software add-ons by giving companies a technical and commercial platform for the development, implementation, distribution, and billing of innovative software solutions.

Methods: The creation of a central platform prototype requires the collaboration of several independent market contenders, including medical users, software developers, medical device manufacturers, and platform providers. A dedicated consortium of clinical and scientific partners as well as industry partners was set up.

Results: We demonstrate the successful development of the prototype of a user-centric, open, and extensible platform for the intelligent support of processes starting with the operating room. By connecting heterogeneous data sources and medical devices from different manufacturers and making them accessible for software developers and medical users, the cloud-based platform OP 4.1 enables the augmentation of medical devices and procedures through software-based solutions. The platform also allows for the demand-oriented billing of apps and medical devices, thus permitting software-based solutions to fast-track their economic development and become commercially successful.

Conclusions: The technology and business platform OP 4.1 creates a multisided market for the successful development, implementation, distribution, and billing of new software solutions in the operating room and in the health care sector in general. Consequently, software-based medical innovation can be translated into clinical routine quickly, efficiently, and cost-effectively, optimizing the treatment of patients through smartly assisted procedures.

(JMIR Med Inform 2022;10(1):e27743) doi:10.2196/27743

KEYWORDS

cloud-based platform; data; eHealth; Internet of Medical Things; IoT; medical apps; multisided market; perioperative medicine; software-defined healthcare; translational research

Introduction

Background

Innovation and growth in the health care sector could be significantly improved by supporting the rapid translation of software-based medical research and its results into clinical routine, increasing patient outcomes at scale. Technological advances such as the improved visualization of target structures during surgery by means of augmented reality [1,2] promise to further enhance the outcomes of surgery for the greater benefit of the patient. However, the development and translation of navigated, software-based innovations into commercial solutions are affected by two main challenges: because of proprietary systems, access to medical devices and their data is difficult [3], and there are high barriers to the transfer of research results into clinical practice, particularly a successful market entry [4]. Compared with Industry 4.0, the health care sector so far has not prepared for similar developments. As an example, medical devices commonly represent highly specialized but unconnected stand-alone solutions, optimized for their task but with limited flexibility and extensibility [5]. Start-ups offer high promises of disruptive innovation in the health care sector because they are highly flexible and make use of new technologies [6]. However, a fast and efficient go-to-market is especially difficult for small companies, start-ups, and spin-offs of research institutes. High barriers to market entry result in delays or failure to bring innovative solutions into clinical routine where their benefits could help larger numbers of patients [7].

The proof-of-concept project OP 4.1 addresses this issue by providing software-based solutions with connection to devices and data, consequently supporting their translation into clinical routine. With OP 4.1, we demonstrate for the first time an open and extensible platform prototype that is not only *open to join*, creating an open ecosystem, but also allows comprehensive connectivity and augmentation of the physical capabilities of medical devices through software-based add-ons, enabling the fast implementation of new solutions in the operating room. The open and extensible design of the platform offers developers well-established and standardized interfaces for stakeholders to connect their apps easier and more efficiently compared with closed systems with proprietary interfaces. The use of open, nonproprietary interfaces in the OP 4.1 platform eases interoperability and data exchange among stakeholders and is important for widespread adoption. In addition, the platform is not limited to a fixed set of interfaces; it can be extended to provide future standards or individual needs.

Software-Defined Healthcare

We introduce the term *software-defined healthcare* in accordance with the definition of *software-defined vehicles*, highlighting that software is becoming the driving factor of innovation and a key value generator, whereas hardware is becoming more and more standardized and eventually commoditized, mostly acting as a base to build differentiating capabilities through software on top [8]. The term builds on the notion of *software-defined systems*, where the software components are segregated from the underlying hardware by means of different abstraction layers [9]. As can be observed in various industries (eg, in the automotive industry), it is easier to update software without having to change hardware, and this also has great potential for continuous innovation and further differentiation. An early example from the automotive sector is the intercompany collaboration on car platforms as well as the use of digital (entertainment) systems as a means of differentiation in current marketing by vendors. In this paper, we focus on software-defined medical procedures in the operating room and how they can be enabled technically and commercially by introducing an underlying platform with a multisided market. We proofed the efficacy of the proposed solution by successfully developing 4 apps for use in the operating room (*augmented reality app*, *live perfusion measurement app*, *precision puncture app*, and *mobile information service app*). Through the introduction of the business and technology platform OP 4.1, we effectively created a multisided market for medical device services, allowing for fast commercialization of software-based research solutions in the operating room and in the health care sector in general.

Methods

Stakeholders Required to Design a Multisided Market

The creation of a central platform prototype requires the active collaboration of several independent market contenders. These consist of *consumers*, *providers*, and platform *suppliers* (Figure 1).

The primary *consumers* on the OP 4.1 platform are medical staff, the hospital administration, and the information technology (IT) department. The *consumers* consume services made available by the *providers* on the platform. The platform's various user spaces must be designed around the requirements of the *consumer* and the medical staff, considering the process of buying, testing, and rating the solutions. In this proof-of-concept study, the *consumer* was the Department of

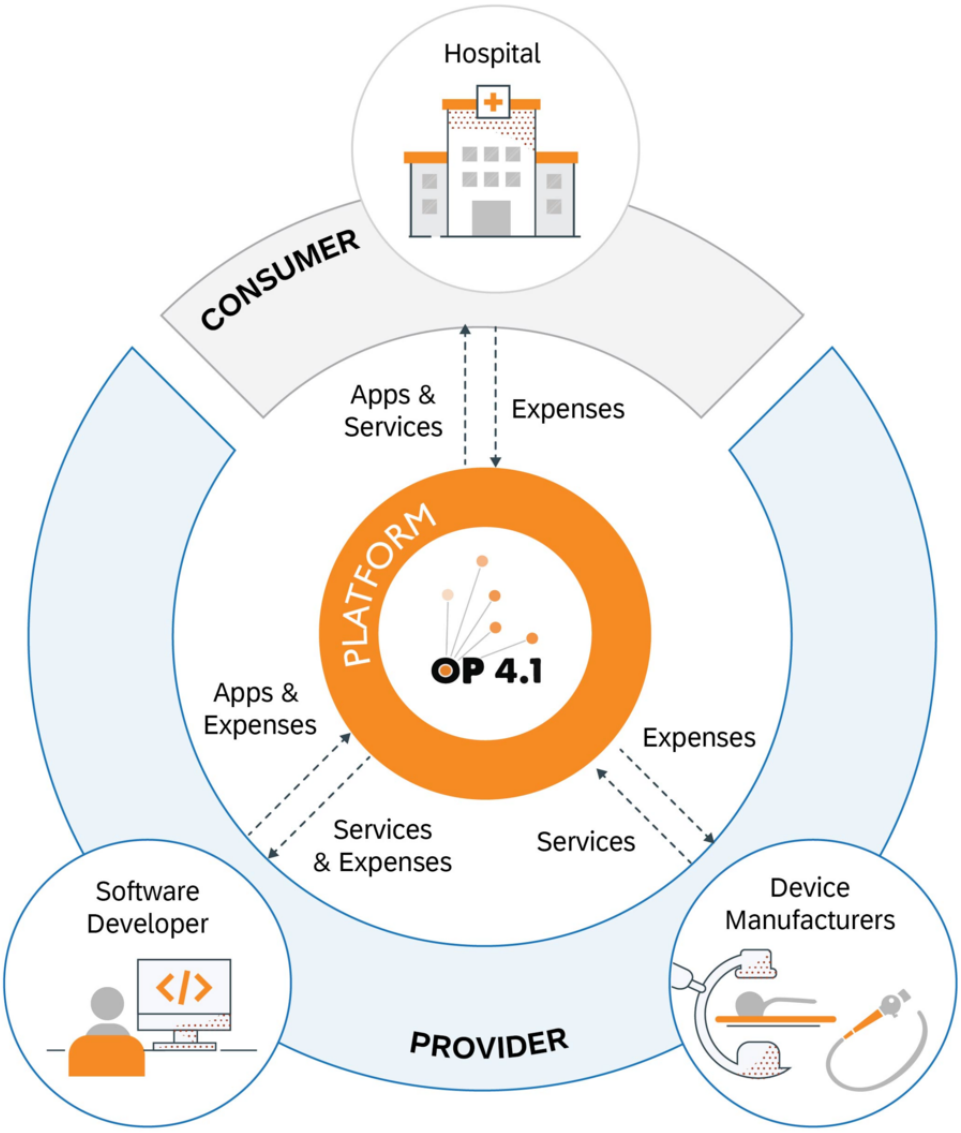
Urology at Heidelberg University Hospital, which initiated the project OP 4.1.

Providers in the context of the OP 4.1 platform are primarily software developers and their companies (eg, start-up or research organizations with their potential spin-offs) and device manufacturers. A software developer in the OP 4.1 consortium was the Deutsches Krebsforschungszentrum (DKFZ; German Cancer Research Center). The role of the DKFZ in the project OP 4.1 was to provide a starting set of services to extend the platform's capabilities, to develop basic infrastructural components for developers to extend the OP 4.1 platform, and to use the OP 4.1 platform to develop apps. The goal was to exemplify how to implement existing research projects of the DKFZ on the OP 4.1 platform in the form of apps [10-12]. The set of research projects of the DKFZ comprised three apps: 1 for preoperative planning and intraoperative assistance of laparoscopic kidney tumor resections (*augmented reality app*) [10], 1 for live perfusion monitoring based on multispectral imaging data (*live perfusion measurement app*) [11], and 1 for a marker-less navigation concept for high-precision needle punctures (*precision puncture app*) [12]. The goal of the fourth app, *mobile information service*, was to disseminate information about the current state and progress of surgeries to mobile devices connected to the OP 4.1 platform. This contribution to OP 4.1 was provided by the start-up company mbits imaging GmbH, the other software developer in the consortium.

As outlined in the previous paragraph, medical device manufacturers are also *providers* on the OP 4.1 platform. They could connect their devices to the platform to supply data (eg, usage data) and provide their device-generated data to software developers according to a price stipulated by them. In the OP 4.1 project, this role was undertaken by the medical device manufacturer KARL STORZ SE & Co. KG, whose contribution to OP 4.1 was to supply a gateway for the standardized acquisition of surgical data streams in the operating room and to facilitate the interface between clinical devices and the cloud-based integration platform. Another medical device manufacturer who joined the OP 4.1 project was Siemens Healthineers AG to enable the *augmented reality* app with intraoperative 3D imaging.

The platform connects *providers* and *consumers* and generates network effects. For the success of the platform, a neutral platform owner, one that is neither *consumer* nor *provider*, is advisable to attract as many market contenders as possible without them having to fear direct competition. This setup is particularly relevant for the medical device manufacturers' side to be able to provide as much choice to the consumers as possible. With SAP SE, a partner for software applications, user-centric design, and, with the SAP Cloud Platform, service-oriented commercial platform solutions, strengthened the OP 4.1 consortium. The goal of SAP SE as the platform *supplier* was also to create a holistic business platform model for this specific scenario based on earlier conceptual work by Cigaina [13].

Figure 1. OP 4.1 platform business model. The OP 4.1 platform approach creates value for all parties involved by facilitating exchanges among several independent market contenders. Through the platform, consumers (eg, hospitals) and providers (eg, software developers and device manufacturers) interact with each other, with providers interacting among themselves as well. Hospitals pay software developers and device manufacturers for apps and services, software developers consuming device services pay device manufacturers for consuming these services, and software developers charge device manufacturers for software solutions such as predictive maintenance.



Approach Required to Implement a Platform for a Multisided Market

In the OP 4.1 consortium, an agile design and development process was used to ensure that the requirements for the technical platform and apps include the needs of all envisioned user types, for example, physicians and developers. The project culminated in the final prototype that was presented at the conclusion of the project in January 2020 at a demonstrator event for the German Federal Ministry for Economic Affairs and Energy, the project governance body Deutsche Gesellschaft für Luft- und Raumfahrt (German Aerospace Center), the general public, and the media.

Technology Architecture of the OP 4.1 Platform to Establish a Multisided Market

The OP 4.1 platform consists of two instances: 1 that is cloud-based and built on SAP Cloud Platform [14] and 1 that is on the premises at the hospital, consisting of the OP 4.1 gateway, connected medical devices, and relevant parts of the hospital's IT landscape (Figure 2).

On the cloud side, a Cloud Foundry (Cloud Foundry Foundation) subaccount [15] holds the OP 4.1 core components such as identity authentication, application programming interface (API) management, cloud database, and internet of things (IoT) services, as well as conversational artificial intelligence (AI), complemented by functionalities related to operating rooms, for example, timestamp tracking, surgery summary, speech recognition, pay per use, and invoice. These OP 4.1 core

components leverage standard SAP Cloud Platform services such as SAP Subscription Billing [16], SAP Consent Repository [17], SAP Credential Store [18], and SAP LiveLink 365 [19]. Access to the functionalities and services is provided by the OP 4.1 user interfaces, primarily the service cockpits, the provided OP 4.1 apps, and the surgery dashboard. These can be accessed by the hospital user and software developers through their common interaction devices such as screens in the operating room for the surgeons and office workstations for the developers.

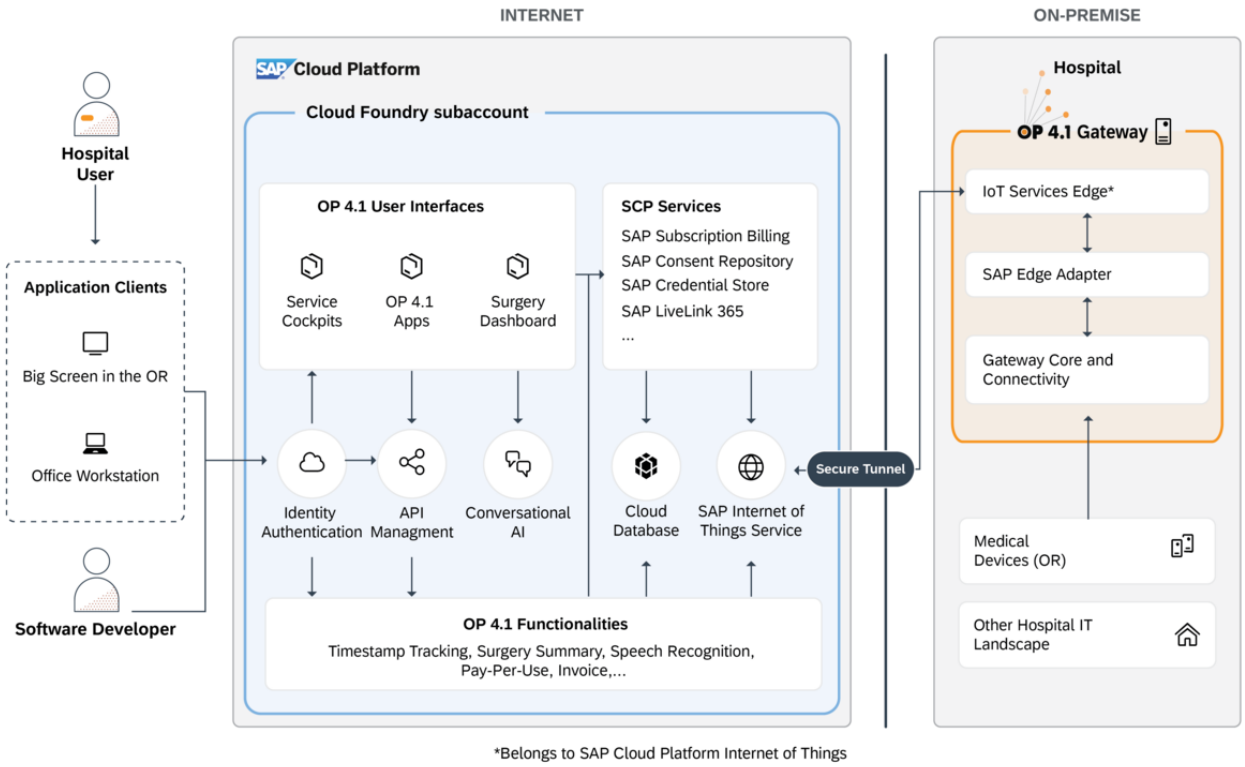
On the on-premises side, we developed the OP 4.1 gateway for the standardized acquisition of surgical data streams in the operating room and for the facilitation of the interface between clinical devices and the cloud-based OP 4.1 integration platform through a secure tunnel. The OP 4.1 gateway provides *connectivity* (eg, data import and preprocessing) by transferring relevant data from devices and existing data sources using interfaces, *interoperability* (eg, adjustment of data types and formats) by definition of data formats, standardization, and dedicated selection of data for analysis, as well as *distribution*, by provisioning data to further systems.

In the OP 4.1 project, a number of interfaces were implemented to connect various medical devices and IT systems. These include, for example, the STORZ Communication Bus [20] as well as emerging standards for the interoperability of medical systems, such as IEEE 11073 Service-Oriented Device Connectivity [21,22], HL7/Fast Healthcare Interoperability Resources [23,24], and IHE Patient Care Device [25]; video capture; and a digital imaging and communications in medicine [26] node. An Apache Kafka (Apache Software Foundation) [27] cluster was integrated to convert the incoming data streams where necessary and to provide real-time data feeds to the cloud instance. The interfaces also allow the connecting of existing parts of the hospital's IT landscape, such as the picture archiving and communication system or hospital information system (HIS), to the OP 4.1 platform, augmenting their capabilities and enabling a holistic overview of all related information. Thereby, the existing data from these systems are not copied to the OP 4.1 platform; rather, related metadata are exposed to enable system access when required.

The generic OP 4.1 platform can be regarded as having interoperability at the structural level (level 2) of the Health Information Management and Systems Society categories [28]. To reach higher levels of interoperability, there are 2 approaches available: on the one hand, platform extensions that supply support for additional interface standards; on the other hand, content extensions that provide semantic layers. The latter, in particular, represents a commercial opportunity for third parties implementing Systemized Nomenclature of Medicine [29] or related standards. As the platform is open and extensible, all conceivable standards that are not yet inherently delivered with a productive instance of the platform can be provided by third parties. Software developers who would like to leverage a semantic layer can build proprietary add-ons to integrate data into a semantic model. The open and extensible OP 4.1 platform also allows software developers to create and make available their own semantic layer to other software developers so that it can be used not only for individual apps, but it is also established as a central element, extending the platform. Given such extensions, it is envisionable to be able to reach higher levels of interoperability, such as semantic level 3 and beyond.

In an active setting, the use of a medical device would trigger an OP 4.1 gateway usage event, which is later processed by SAP IoT Services Edge [30] and provisioned to the cloud-based event handler. The event handler in turn creates a usage record with SAP Subscription Billing. Once triggered, the invoicing functionality processes all available usage records and generates a consolidated invoice according to the respective billing rules.

Figure 2. OP 4.1 architecture. The OP 4.1 platform consists of two instances: one that is cloud-based and built on SCP and one that is on the premises at the hospital, consisting of the OP 4.1 gateway, connected medical devices, and relevant parts of the hospital's IT landscape. On the cloud side, a Cloud Foundry subaccount holds OP 4.1 core components such as identity authentication, API management, cloud database, and IoT services, as well as conversational AI, complemented by functionalities related to ORs, for example, timestamp tracking, surgery summary, speech recognition, pay per use, and invoice. These OP 4.1 core components leverage standard SCP services such as SAP Subscription Billing, SAP Consent Repository, SAP Credential Store, and SAP LiveLink 365 supporting various billing scenarios, documenting patient consent, and managing user authorization. Access to the functionalities and services is provided by the OP 4.1 user interfaces, primarily the service cockpits, the provided OP 4.1 apps, and the surgery dashboard. These can be accessed by hospital user and software developers through screens in the OR and office workstations. Direct access to the platform for software developers is also provided. On the on-premises side, we developed the OP 4.1 gateway for the standardized acquisition of surgical data streams in the OR and for the facilitation of the interface between medical devices and the cloud-based OP 4.1 integration platform through a secure tunnel. The SAP Edge Adapter and the IoT Services Edge together with the Gateway Core and Connectivity implement the required data models to publish the device data for SCP. The IoT Services Edge enables decentralized data processing at the edge of the network. This affords the possibility to process data and services locally. Through the IoT Services Edge, processed or aggregated data can be sent from the OP 4.1 gateway to the cloud. The SAP Edge Adapter we developed for this project is a Kafka-to–MQ Telemetry Transport adapter and connects the Gateway Core to the IoT Services Edge. The Gateway Core contains, among others, a STORZ Communication Bus client, an Online Certificate Status Protocol client, and the Apache Kafka component. AI: artificial intelligence; API: application programming interface; IoT: internet of things; IT: information technology; OR: operating room; SCP: SAP Cloud Platform.



Data Protection and Security Concept Required to Apply a Platform in Clinical Routine

To ensure broad acceptance of the OP 4.1 platform, we created a data protection and security concept. When used in a real-life scenario, OP 4.1 requires conformity with relevant data protection laws when handling patient data [31]. Since May 2018, the General Data Protection Regulation 2016/679, a regulation in European Union law on data protection and privacy, has been the binding directive in the European Union and the European Economic Area [32]. On the basis of this data protection regulation, we generated an OP 4.1 data protection checklist for software developers. Data processing systems must be protected against unauthorized use, and only authorized persons should have access to the data for which they have been granted specific access rights. This was technically implemented using the Open Authorization 2.0 authentication protocol [33] and several other security technologies and capabilities provided by the underlying SAP Cloud Platform such as the SAP Authorization and Trust Management Service [34], the SAP Cloud Identity Services [35], and the SAP Trust Center [36], which are set up in a fenced network. The OP 4.1 platform prototype runs in a shared environment where the data are isolated from each other and the traffic is controlled by firewalls. Administrative access is performed through terminal services that require strong authentication. All communication channels are protected with the transport layer security protocol. Proper user authorizations are also required and respected: for example, when an HIS is connected to the platform, the platform is able to allow its users access to the HIS through the dashboard. However, the user needs to be authorized against the platform, and as in well-established single–sign-on scenarios, the platform passes on the user request to the HIS for further processing. There, the user request is checked to determine if the user is

authorized to access the requested parts of the target system (eg, the HIS) so that access can be granted. Thus, users do not have automatic access to all data or systems through the platform; only users who have permission to use the respective parts of the connected systems (eg, the HIS) can access it through the platform.

We established that the integration platform enables apps to log access to data, including but not limited to username and access date. All processes need to include an option for correction, anonymization, and deletion. In real-world use, affected patients must agree that their health-related data are collected, processed, and used on the cloud-based platform and made available to its respective parts.

Results

The OP 4.1 Platform as a Basis for the Implementation of Translational Apps in Clinical Routine

The ultimate goal of OP 4.1 is to create a multisided market by providing a technical and business platform to help research-based solutions to fast-track their economic development and become commercially successful (Figure 3). To support the effective translation of software-based solutions into clinical routine, we developed the OP 4.1 platform prototype. As described previously, the OP 4.1 platform has the capability to provide standardized and open interfaces to devices and data sources, integrate heterogeneous data, and provide central services (eg, data modeling and processing, user administration, and access management) as well as development and commercial support, all combined with a user-centric design (Figure 2).

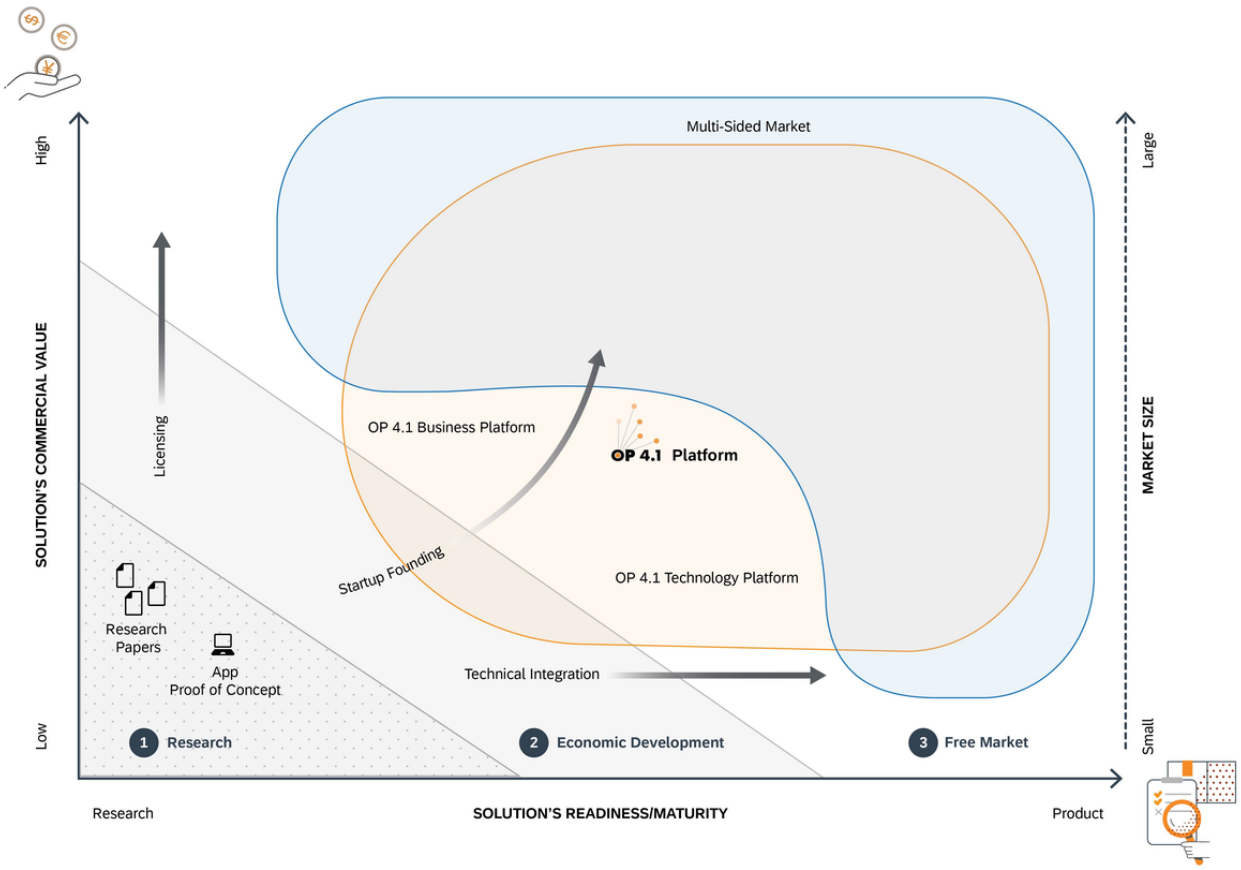
The platform provides clinical process data (eg, image and video data, vital signs, and device data) and an expandable selection of platform services (APIs) to its users. Moreover, platform apps can be built using the software development kit of the OP 4.1 platform with predefined design and interaction concepts, the documentation of previous apps, and the OP 4.1 API Hub. The API Hub serves as the central instance for searching for APIs to use within apps on the OP 4.1 platform prototype. The API Hub itself is a web portal application, documenting use and prices and allowing APIs to be evaluated, downloaded, and tested. The APIs, which are provided as services to all apps, are functions that are technically validated by the platform provider, medical device manufacturer, or software developer through the certification of the API-providing device or data source, and they extend the core platform capabilities. The APIs are provided on the platform to developers for creating software or for interacting with systems. Thus, the APIs often serve common functionalities to reduce the development effort for software developers and to enable easy access to device and other data, supporting intuitive development on the platform. The concept allows third-party companies to develop their software solutions directly on the development instance of the OP 4.1 platform, testing them against digital devices offered by medical device manufacturers, running quality checks on the platform's quality instance, ideally having their solutions' technical interfaces automatically precertified, and eventually deploying them into the hospital's own platform space. It needs to be noted that although the OP 4.1 platform can provide a certification of the technical interoperability of apps and medical devices at an interface level to ensure seamless integration with the platform, medical certifications for each app need to be fulfilled additionally by the respective software developer. Medical products, including apps, can be certified modularly and independently of each other, and their sensible combination is then also permissible. In the case of OP 4.1, not only the respective apps and the connected medical devices, but also the OP 4.1 gateway as a specific medical device must be certified.

However, the OP 4.1 platform makes all usage data available for authorized users and provides a unified environment for developing, testing, deployment, and runtime of apps. Thus, by providing the relevant data and direct information to support medical certification in a standardized and repeatable manner, the translation of software solutions into clinical routine can be fast-tracked through the OP 4.1 platform compared with the time-consuming process of certification of solutions for different stand-alone environments. In addition, the platform's standard environment with modular functionality means that certification can take place more efficiently because the basic underlying functionality remains the same across apps and its descriptions and reviews can be easily reused for new use cases on top of the OP 4.1 platform.

Although many apps will be cloud-only, thus mostly soft real-time, there are specific use cases when solutions from OP 4.1 are not only deployed in the customer's cloud space, but will also be partially delivered on special appliances at runtime because they might require highly hardware-dependent functionality, including but not limited to hard real-time (eg, high-definition 3D videos and graphics processing unit arrays for augmented reality or deep learning algorithms). In this case, the platform still controls all related metadata and data storage locations to perform the platform's services such as authorization, updates, connection, charging, billing, and invoicing.

To our knowledge, the OP 4.1 platform is the first holistic cloud-based platform that supports developers in designing, coding, testing, deploying, maintaining, and commercializing their innovative solutions for the operating room.

Figure 3. OP 4.1 approach. Most research organizations are funded through research grants; their work results in research papers and, in the area of software-defined healthcare, apps as proofs of concept. The readiness and maturity of such solutions is often low, and there is very limited commercial value to capture because of their nature as proofs of concept. To overcome these challenges, develop economically, and generate revenue outside of research grants, there are 3 different dimensions to consider: first, protect and license the innovation, which increases its commercial value; second, increase the solution's readiness through further technical integration, effectively providing apps that come closer to a more mature product state; and third, found a start-up to commercialize the innovation or invention. By choosing the third option, the start-up needs to not only show increasing commercial value to investors but also increase the solution's readiness beyond the initial proof of concept. Eventually, there needs to be a market, especially for such purely software-based solutions, which, because they mostly require medical devices to operate with, is a multisided market with different types of market contenders interacting on a platform. This is where the OP 4.1 platform comes into play. On the side of the solution's readiness, the technology platform side of OP 4.1 provides an environment to efficiently execute the solution, supporting the entire development and deployment cycle. On the side of commercial value, the focus is on OP 4.1 as a business platform: by being able to not only bring a solution to market but also charge for it, commercialization becomes possible, and the access to the market created by the OP 4.1 platform enlarges the addressable market of the start-up's solution, effectively increasing its commercial value. As a result, OP 4.1 creates a multisided market by providing a technical and business platform to help research-based solutions to fast-track their economic development and become commercially successful.



The OP 4.1 Gateway to Connect the Medical World With the IT World

To apply the OP 4.1 platform in the operating room, it needs to be connected to the various medical devices and appliances. As a well-defined link among these independent medical devices with IT solutions, including a cloud-based integration platform such as OP 4.1, we introduced the OP 4.1 gateway (Figure 2). Different interfaces were implemented to connect various medical devices, data sources, and IT systems (as described in the *Methods* section). The gateway locally aggregates the intraoperatively acquired multimodal data, preprocesses them, and makes them available to the central platform layer. The gateway also ensures proper import, standardization, and distribution of data, while promoting data connectivity and interoperability. Data persistence is achieved through connected systems and data lakes such as the picture archiving and communication system.

The OP 4.1 Dashboard and Apps to Prove the Translational Capabilities in the Clinic

To demonstrate the efficacy of the OP 4.1 platform, we created a platform dashboard and 4 platform apps. The dashboard is a proof of concept for using platform information and for making the integrated platform visible to the end user. With the context-sensitive dashboard based on a user-centric interaction concept, we demonstrate that the physician in the operating room can interact intuitively with the platform. The dashboard also enables its users to gain access to external sources such as the HIS and start the required apps during surgery. We introduced 4 proof-of-concept apps to demonstrate how new software solutions can be easily integrated into the platform. The entire OP 4.1 prototype was tested with the medical use

case *kidney tumor*. The aim of the dashboard is to provide the physician with the appropriate information at the right time (Figure 4). The required patient information, including information from a connected HIS, can be requested using natural language through built-in speech recognition or by manual input through a touch screen. Apps can be started centrally by a physician or a nurse through the dashboard. By sending push notifications to mobile devices through the *mobile information service* app, relevant stakeholders such as the surgery team are informed about the start or delay of planned operations, independent of their location. This is based on the app's ability to retrieve current intraoperative time stamps from the cloud platform through APIs. By combining images of computed tomography scans from connected systems through the central cloud platform with live ultrasound images, the *precision puncture* app (Figure 5) enables a safe and fast percutaneous puncture of target structures, for example, in the kidney. The *augmented reality* app can display risk structures intraoperatively based on the segmentation of computed tomography images and the overlay of a 3D view of the kidney tumor in the laparoscopic video stream. The *live perfusion measurement* app enables the continuous quantification of renal tissue oxygenation with multispectral image analysis and machine learning. This allows the physician in partial kidney resection to determine which part of the kidney is still perfused after selective arterial clamping, helping to reduce the risk of kidney injury compared with hilar clamping [37].

The functionality of the OP 4.1 apps and data streams was demonstrated in real time at the conclusion of the project OP 4.1.

Figure 4. OP 4.1 user-centric dashboard with access to medical data and apps. The context-sensitive dashboard enables users to control the entire OP 4.1 platform before, during, and after surgery. It also provides seamless access to information from external sources such as the hospital information system and allows the starting of preselected apps.

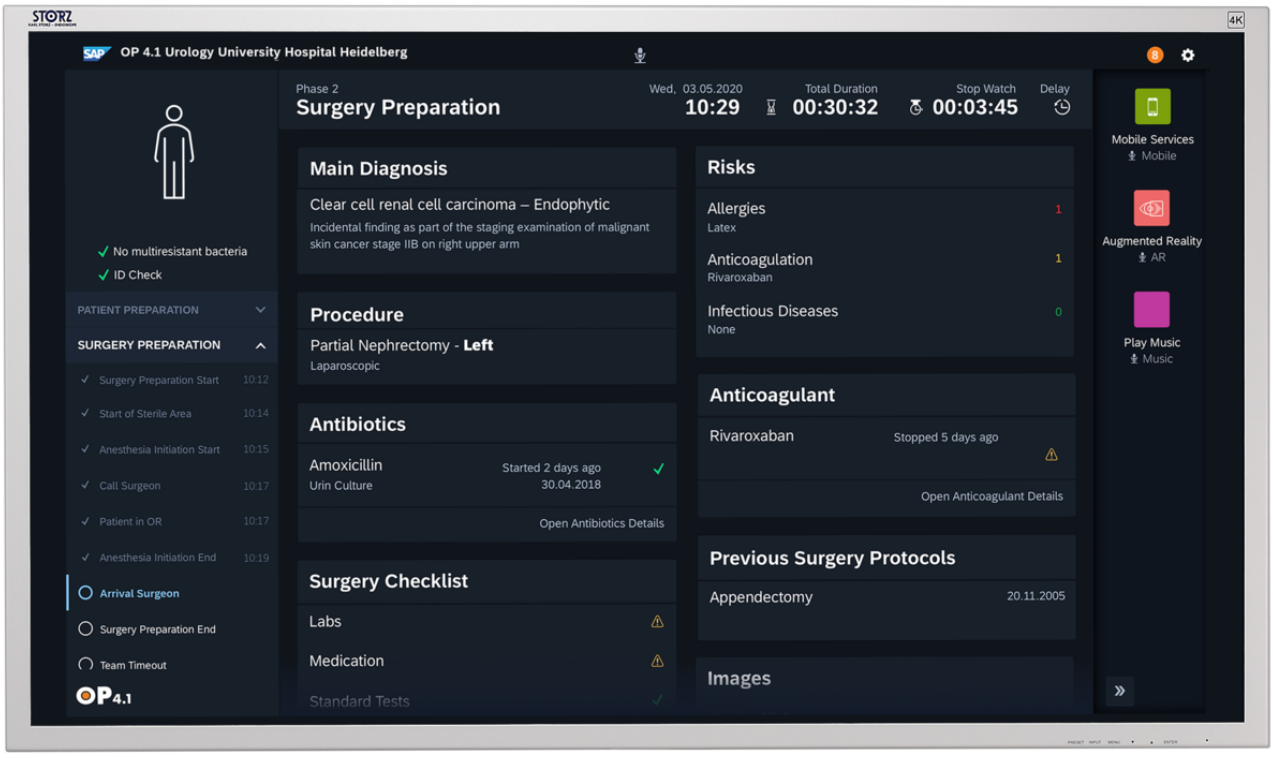


Figure 5. OP 4.1 precision puncture app. By combining live ultrasound images with segmented computed tomography images from connected systems through the central OP 4.1 platform, this app enables a safe and fast percutaneous puncture of target structures, for example, in the kidney. Left: ultrasound image with the projection of the needle and the segmented structures from the computed tomography images. Right: 3D scene, including the ultrasound plane and the segmentations, made possible by multimodal image fusion.

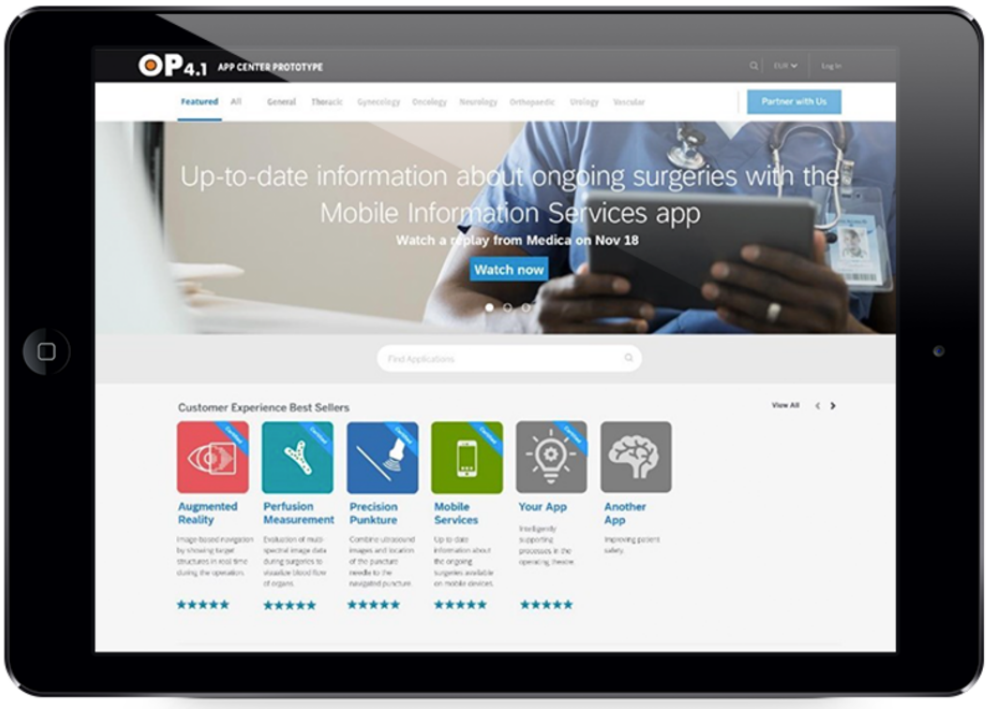
The OP 4.1 Platform Business Model Concept to Create Tangible Value

Besides the development of a technology platform, we created a concept for a platform business model. To be commercially viable, the OP 4.1 platform approach needs to create value for all parties involved by facilitating exchanges among several independent market contenders: on the platform, consumers and providers interact with each other, with providers interacting among themselves as well, exchanging value (Figure 1). Software developers provide apps and digital content, and device manufacturers provide access to medical devices and related information in real time. Essential for a platform business model is a constant revenue stream for all market contenders as well as transparent pricing. Here, hospitals pay software developers and device manufacturers according to their respective agreed-upon payment model, software developers consuming device services pay device manufacturers for consuming these services, and software developers charge device manufacturers for software solutions such as predictive maintenance.

To access the platform, various channels and interfaces such as mobile apps, websites, digital stores, web services, or physical devices can be envisioned. Being the central entity, the platform provides interfaces to consumers, providers, and platform developers: consumers select and test apps on the OP 4.1 app center (Figure 6), providers expose and publish their microservices in the OP 4.1 API Hub as well as upload their apps, and platform developers can extend the platform through new standard services or additional medical gateways. Eventually, the platform will provide rule-based quality assurance of interfaces and certification for technical compliance, machine learning–based quality assurance, and social-based quality assurance through the ratings on the OP 4.1 app center as well as consumer feedback.

It is worth noting that the OP 4.1 platform prototype is a cooperation project among small and large German companies and research institutions, with an open and extensible approach. During the project, other companies (eg, Drägerwerk AG & Co. KGaA, Intuitive Surgical, Inc., and RaySearch Laboratories AB) expressed their interest in the OP 4.1 prototype and became associated partners. Such new partners could expand and enrich the platform with additional functionality.

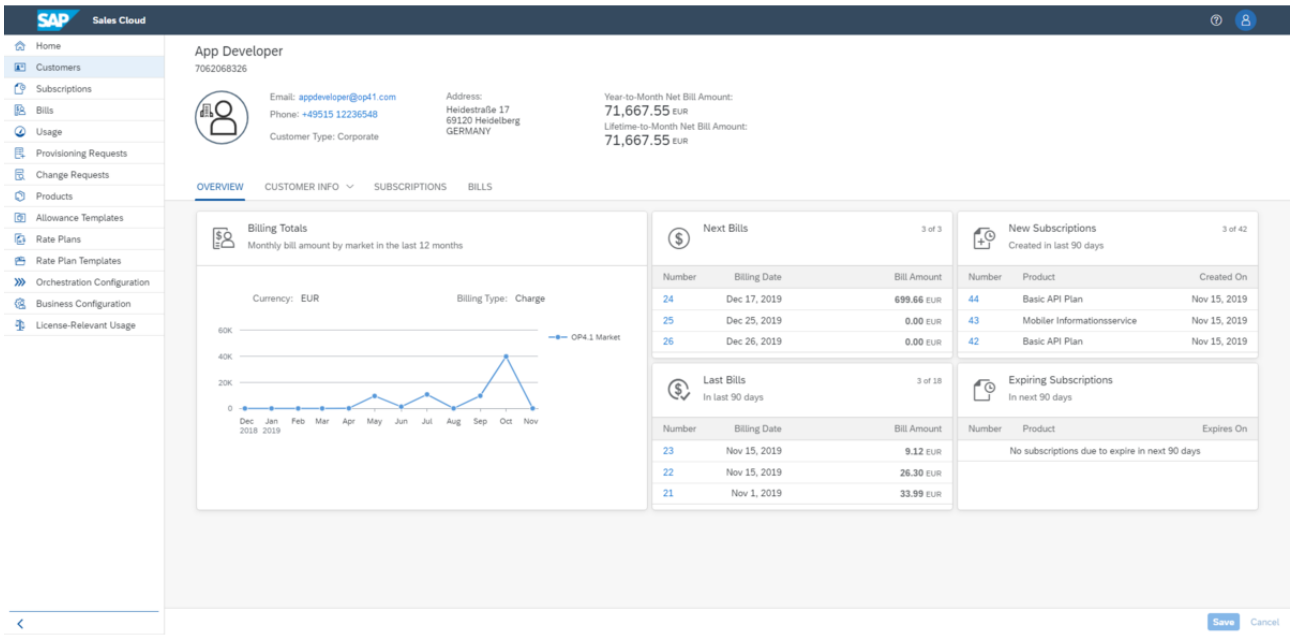
Figure 6. OP 4.1 app center. At the app center, users such as physicians and nurses could preselect apps for a specific surgery based on their filter criteria, recommendation, and previous social feedback from other users.



The Pay-Per-Use Model to Effectively Capture Value

To help research institutes to monetize their inventions and intellectual property, we designed a flexible pay-per-use model as part of the OP 4.1 platform business model. Currently, medical devices are usually financed through a one-time payment of the product price as capital expenses plus recurring service and support fees as operational expenses. This makes the availability of highly innovative new solutions in the health care sector difficult because of limitations in the availability of capital. A business platform such as the OP 4.1 prototype can help to fill a market gap by supporting innovative payment terms for medical devices and software-based innovation. Information about device activity and specific modes will be transmitted to the integration platform. The platform can automatically provide the use records of devices and apps, thus enabling direct time-based or use case–based billing (Figure 7).

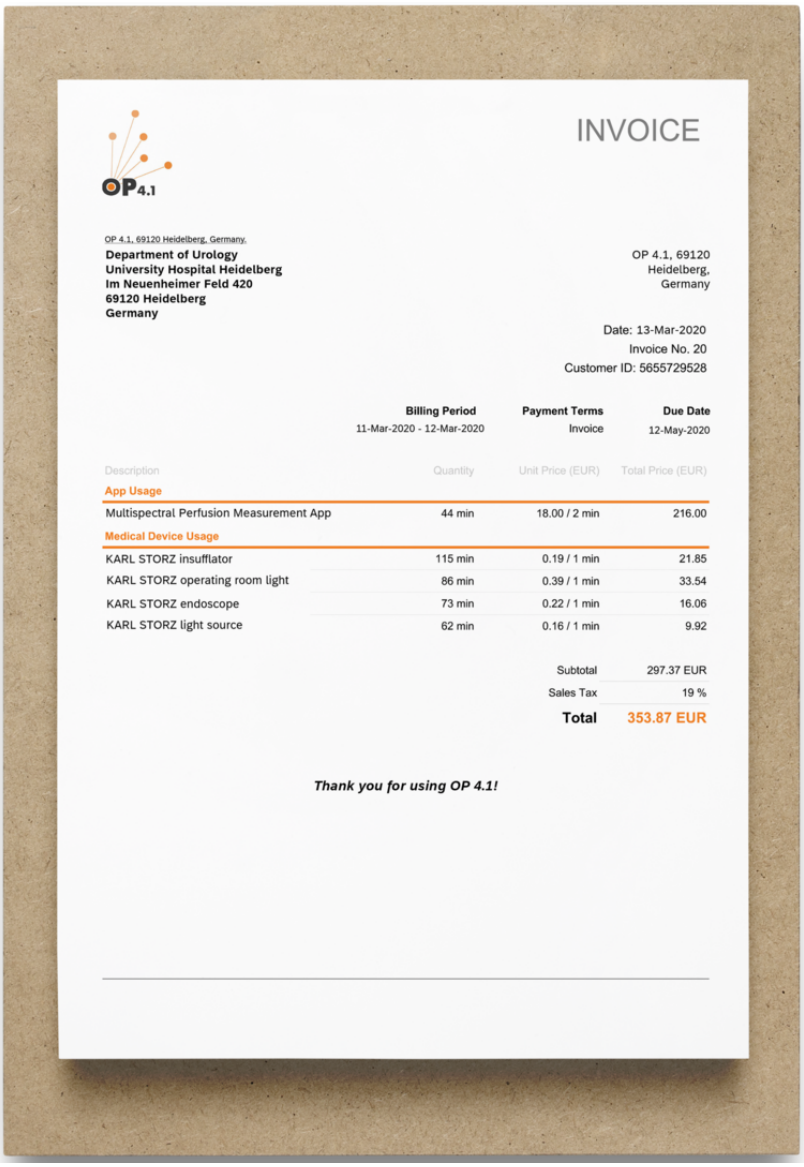
Figure 7. OP 4.1 commercial dashboard. On a single dashboard, the providers on the OP 4.1 platform can review and manage the commercial side of their activities. Among many other features, they can drill down into their customer base, see an overview of paid subscriptions, review bills and usage, and manage their products and commercial plans.



We proved how to generate a consolidated invoice for the OP 4.1 showcase as shown in Figure 8. Such a service-oriented, pay-per-use model would have the advantage of not only spreading high, one-time investment costs over time, but it would also allow converting of capital expenses for highly innovative and beneficial equipment into operational expenses by allocating the exact costs during the intervention on a patient-related basis.

Currently, there is a strong trend from product sales through selling services to selling outcomes, which materializes in seven atomic business models across industries: classical, physical price models such as (1) give away device for free *(gift)* and (2) pay per device *(buy physical part)* as well as use-bound pricing models such as (3) pay per time that device is owned *(rent)*, (4) pay per time that device is used *(pay per time used)*, (5) pay per capability that device has delivered *(pay per [micro]service provisioned)*, (6) pay per number of cases that device contributed to *(pay per incident)*, and (7) pay per total outcome of device's use *(pay by outcome)*. For the OP 4.1 platform, it can be envisioned to offer pricing models based on transaction fees (eg, pay per activity or use), subscription fees (eg, pay per time), and lead-generation fees (eg, pay by transaction initiated or facilitated). Building on this concept, it can be foreseen that all potential device use scenarios eventually map to a combination of these pricing models, enabling even more exotic pricing models on the platform to support the varying use cases encountered in real-world hospital scenarios.

Figure 8. OP 4.1 platform invoice and pay-per-use model. The invoice generated by the OP 4.1 platform consolidates all resources used for a specific procedure, for example, medical devices, apps, and expendables, and charges for them in a single invoice according to their respective underlying commercial models.



Discussion

Principal Findings

The rapid translation of software-based medical research and its results into clinical routine to improve patient outcomes is confronted with structural hurdles, even in the era of Industry 4.0. Innovative solutions are affected by difficult access to medical device data and high barriers to market entry because of proprietary systems. In this proof-of-concept project, we demonstrate, to the best of our knowledge for the first time, how these issues can be addressed by a technology and business platform for the operating room. Through the integration of different perioperative process data and the connection of clinical systems with patient data, it offers a basis for the development and implementation of new and innovative solutions in the form of apps. In the project OP 4.1, we successfully developed a cloud-based integration platform that also provides a business model and related platform. This business platform not only enables fast access to published device data and relevant patient data, but also allows charging for apps or devices, that is, through pay per use. Thus, we effectively created a new market for purely software-based solutions, helping to transfer technical innovation into clinical routine and to become commercially successful. We anticipate that with the distribution of smart, software-based medical solutions through the platform the treatment of patients can become safer and more precise.

The OP 4.1 platform could create benefit for *public health*, *physicians* and *hospitals*, *device manufacturers*, *software developers*, and ultimately *patients*. With the percentage of people aged ≥65 years doubling until 2050 and with an increasing population in general [38,39], *public health* is confronted with demands to be both productive and progressive. Medicine is facing large numbers of newly developed medical devices that need to interact with each other. Whereas other industries have already embraced IoT, in health care the Internet of Medical Things [40] is still in its infancy in terms of supporting the treatment of patients. Further digitalization of the health care sector can help to create a medical environment that is more predictive, preventive, personalized, and participatory [41].

At *hospitals*, *physicians* are the primary active users of platforms such as OP 4.1, leveraging apps offered through app centers. The expansion of software-defined healthcare helps to strengthen navigated, smartly assisted procedures, allowing *physicians* to take better and quicker decisions. *Physicians* benefit from quality assurance in real time (eg, live perfusion measurements) and potential procedural improvements (eg, improving the outcome of a surgery with augmented reality apps). We also anticipate that platforms could help physicians to increase productivity (eg, transfer from surgery data into surgery protocol), leveraging an existing HIS connected to the platform. In addition, through the transparent logging of process data, such platforms facilitate quality control for *hospitals* and the implementation of new models for commercialization of apps and devices through pay per use, allowing the conversion of investment costs into intervention-related operational costs. The business model is attractive to hospitals not only for being able to shift cost types such as capital expenses to operational expenses, but also for being able to allocate expenses that are normally considered overheads to individual patient treatments, subject to local regulations. The business model also allows hospitals to charge device manufacturers and software developers for evaluating new products and for providing user feedback. In addition, apps allow hospitals to obtain further benefits through more efficient processes and cost savings. The more efficient use of technology, information, and personnel in interventions can contribute to long-term cost reduction as well as optimization of work processes at *hospitals*.

Established *device manufacturers* can use platforms such as the OP 4.1 prototype to create and offer an ecosystem of connected solutions and even vendor-brand them. Traditionally, products sold by *device manufacturers* to hospitals represent optimized but self-contained stand-alone solutions. The usage data provided by the platform enables *device manufacturers* to continuously improve and adapt their solutions based on the needs of users that are derived from the information available in the operating room. The provision of medical device data is important for data-driven applications, including resource management, process optimization, quality assessment, performance analysis, context-sensitive assistance, and predictive maintenance. This leads to the use of devices in new scenarios and faster and more effective innovation cycles.

For *software developers*, the successful integration of the 4 apps into the OP 4.1 prototype acts as a solid reference and proof of concept on how to implement apps on the platform and exemplifies the simplification of translation of innovation into clinical routine. The 4 apps developed as part of the OP 4.1 prototype act as templates for designing and coding for other software developers. As functionality is replicated across the platform, once developers know how to create a module, further development can be performed more quickly. In addition, the platform provides central services to developers for developing and exposing their solutions to the customers through an app center. *Software developers* creating new software solutions get access to other providers' device simulations through the platform and can augment device capabilities by applying new software solutions. Eventually, *software developers* will potentially require less time and effort to deliver new solutions (eg, through predefined design concepts). Ultimately, translational research departments benefit from a quick and cost-effective transfer of their innovative results to multiple customers. The platform prototype gives researchers access to a large customer market (eg, hospitals that already participate in such platforms). This avoids a scenario where only *patients* of selected hospitals benefit from the latest innovation while the respective clinical trial is open and active. Cultural and socioeconomic health care systems vary widely across Europe and worldwide, with many systems approaching health care as a business, sometimes at the risk of implementing 2-tier or multitier medicine. New models for commercialization of apps and devices through pay per use do not work in this dimension, but they offer commercial benefits such as allowing the conversion of investment costs into intervention-related operational costs. In new business models, all *patients* could

benefit from innovation, even if invoicing is just occurring for *patients* with, for example, private insurance. Software-defined healthcare can even be cheaper for hospitals because it is often not necessary to buy new equipment; instead, the functionality of the existing equipment can be extended by means of software upgrades or add-ons.

Thus, the OP 4.1 platform can lower the barrier to market entry and efficiently make innovations available to all platform participants as well as patients.

Comparison With Previous Work

New opportunities for data collection have been created by the ubiquitous availability of mobile devices and wearables. Advances in health platforms (eg, Apple HealthKit and Google Fit) allow the bundling of fitness and medical data from different sources and make these available for sharing with health care professionals [42]. The aim of the SMART project was to develop an open platform to enable medical apps to run unmodified across different health care IT systems, promoting interoperable and vendor-independent apps [43]. The Medical Device Plug-and-Play Interoperability Program has been promoting medical device interoperability to enable the creation of complete electronic health records and cost-effective development of medical apps when using networked medical devices in clinical routine [44]. Another example is OR.NET, which defined cross-manufacturer concepts for the dynamic and secure networking of medical devices and IT systems [23]. Further research was performed in the project InnOPlan, which developed a SmartData platform for the real-time provision and analysis of medical device data to enable data-driven services in the operating room [45]. The HiGHmed Consortium aims at establishing data integration centers and an open platform architecture in cooperation with health care providers in the fields of oncology, cardiology, and infection control so that integration and reuse of data are facilitated [46]. The HiGHmed Consortium is one of the members of *the GermanMedical Informatics Initiative*, whose goal is to advance digitization in health care research and the exchanging of patient data, specifically among university hospitals. Data integration centers will enable research data to be collected and integrated across several institutions and locations [47].

Our approach goes far beyond previous efforts that focus primarily on data integration in medicine. With the OP 4.1 platform, we demonstrated for the first time a cloud-based integration platform that also provides a business model and related platform, thus effectively creating a new market for purely software-based solutions and helping to transfer technical innovation into the operation room and clinical routine and to become commercially successful.

Future Directions

Going further, it is straightforward to envision several dimensions in which to expand our findings in subsequent projects: big *data* feeding AI algorithms, the *expansion* of stakeholders on the platform, and the *transfer* of the platform concept into other medical disciplines.

The available *data* in the operating room can be permeated through the integration of devices from various companies, with the OP 4.1 platform enabling the development of solutions that span >1 device type. By introducing a consistent system to capture information across the operating room, it will be possible to create solutions analyzing current situations and patients' states as well as simulating outcomes, consequently predicting future states and proposing next steps for the interventions. High-quality data sets dynamically generated on a per-patient basis through medical devices and central clinical *data* collection are ideal for machine learning algorithm training [48,49]. Big data analysis will play a significant role in transforming medicine, and technology that enables the central organization, processing, and security of these data is critical [50-52]. The integrative and expandable OP 4.1 platform concept can provide the registration and analysis of diverse results from diagnostics and therapy in real time and over time. As a result, a data infrastructure could be created, supporting the next major step into data-based individualized medicine with its personalized and customized therapies [53].

The *expansion* of a platform such as OP 4.1 could attract many companies, thus starting self-supporting network effects that lead to higher numbers of devices being made available on the platform and more developers developing on the platform, resulting in more solutions available on the platform, and finally more consumers using these solutions and hospitals subscribing to the platform. In the long term, an ecosystem of producers and consumers, including various hospitals, device manufacturers, and software developers, can be curated on the platform.

In addition, the concept of the OP 4.1 prototype is not meant to be restricted to an operating room; a *transfer* to other medical disciplines is a viable option. Various disciplines could benefit, for example, through less effort having to be made for point-to-point integration, quality control, as well as the simple integration of systems by different vendors. The use of the OP 4.1 platform in multiple departments of a hospital enables central patient data management and analysis across the existing clinical information systems, exploiting potential synergies.

A business platform such as the OP 4.1 platform prototype can support several rising industry trends such as the transformational trend *digital health platform* for hospitals. The concept *digital health platform* as the architecture enabling the composable enterprise for health care providers supports the evolution from electronic medical records to a *hospital without walls* by enabling these organizations to rapidly adapt to changing patient demand, partner capabilities, and industry trends. This concept can be realized by leveraging packaged business capabilities and is anticipated to become mainstream in 2025-2030. Further major industry directions are health data curation and enrichment, AI-enabled diagnostic imaging interpretation, and particularly an app marketplace for health care providers [54]. All these are included or could easily be enabled by a platform such as OP 4.1. Natural extensions of the OP 4.1 platform could focus on digital twins in health care, precision health, and medicine, as well as AI health care advisors. Furthermore, there is a strong potential to not only promote IoT in health care, a growing area of research and another transformational trend with the opportunity for health care systems to predict health issues and to monitor patients in

various institutions [55], but to also help to make it commercially viable through the concept of a multisided market that we introduced with OP 4.1.

Limitations

This study includes certain limitations. In the proof-of-concept project OP 4.1, we developed a prototype. During the development of the OP 4.1 platform for implementation at different institutions, further challenges and tasks will arise. The OP 4.1 platform prototype would need to be adapted to hospital-specific infrastructures and requirements; therefore, it would need to be delivered as a customer-specific solution.

In the OP 4.1 project, 4 apps in the field of urology for the medical use case *kidney tumor* were developed. These apps demonstrated as a proof of concept the feasibility of developing and integrating apps into the OP 4.1 platform prototype. The verification of the superiority of these apps compared with standard procedures, that is, through validation with real-world data, is pending. In this proof-of-concept project, we did not present quantitative data regarding an actual implementation on the OP 4.1 platform. After having proven the functionality of integrating medical research as apps into the OP 4.1 platform, the next step will be to prospectively validate the individual apps in clinical routine according to their cost-benefit ratio, patient safety, and improved clinical outcomes.

Going forward, to expand the scope for other medical disciplines, more apps need to be integrated into the OP 4.1 platform. The number of valuable apps needs to be expanded and adapted to the different medical disciplines and customer needs.

It should also be noted that before apps can be released for distribution through a platform such as OP 4.1, they would need to be certified and approved for use. It is important to ensure that the respective rules of medical device regulations are strictly observed. Personal data security and protection compliance as well as medical device certifications need to be fulfilled by each software developer individually because only they know which data are generated, stored, processed, and so on.

Conclusions

In the proof-of-concept project OP 4.1, the prototype of a user-centric, open, and extensible platform for the intelligent support of processes in the operating room was developed. By connecting data sources and medical devices from different manufacturers, the technology and business platform creates a multisided market for the successful development, implementation, and accounting of innovative software solutions in health care. Consequently, software-based medical innovation can be translated into clinical routine quickly, efficiently, and cost-effectively, optimizing the treatment of patients through smartly assisted procedures.

Acknowledgments

The OP 4.1 consortium was supported by the German Federal Ministry for Economic Affairs and Energy with a grant for economic innovation to push forward the cloud-based technology and business platform concept OP 4.1. The project OP 4.1 was an official part of the technology program *Smart Service World II*. The project governance body, Deutsche Gesellschaft für Luft- und Raumfahrt (German Aerospace Center), monitored the project and provided administrative, project management, and review support. The authors especially want to thank Ms Gudrun Müller from the Deutsche Gesellschaft für Luft- und Raumfahrt for her valuable advice and support.

Authors' Contributions

MG, MB, MR, JF, MM, MVKD, DT, PS, LMH, SD, and MH conceptualized and led the project. MG, MR, VS, PR, CG, TS, KM, AS, MN, TR, DMF, DM, JM, SO, SS, and LM performed data collection, prototype design, development, training, optimization, and validation. MH, MB, JF, MM, and LMH supervised the prototype development and validation. MG and MB wrote the draft of the manuscript in close collaboration with SD and MH. All authors critically reviewed the paper. All authors approved the final submitted research manuscript and agree to be personally accountable for their contribution and for the academic integrity of the work.

Conflicts of Interest

MB is employed at SAP SE (Walldorf, Germany). LM and JF are employed at KARL STORZ SE & Co. KG (Tuttlingen, Germany). MM is chief executive officer at mbits imaging GmbH (Heidelberg, Germany). PS is employed at Siemens Healthineers AG (Forchheim, Germany).

References

1. Simpfendörfer T, Gasch C, Hatiboglu G, Müller M, Maier-Hein L, Hohenfellner M, et al. Intraoperative computed tomography imaging for navigated laparoscopic renal surgery: first clinical experience. J Endourol 2016 Oct;30(10):1105-1111. [doi: 10.1089/end.2016.0385] [Medline: 27530774]
2. Porpiglia F, Checcucci E, Amparore D, Manfredi M, Massa F, Piazzolla P, et al. Three-dimensional elastic augmented-reality robot-assisted radical prostatectomy using hyperaccuracy three-dimensional reconstruction technology: a step further in the identification of capsular involvement. Eur Urol 2019 Oct;76(4):505-514. [doi: 10.1016/j.eururo.2019.03.037] [Medline: 30979636]

3. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2:79 [FREE Full text] [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
4. Lee DH, Reasoner K, Stewart A. From concept to counter: a review of bringing an orthopaedic implant to market. J Am Acad Orthop Surg 2020 Jul 15;28(14):604-611. [doi: 10.5435/JAAOS-D-20-00017] [Medline: 32692096]
5. Hao A, Wang L. Medical device integration model based on the internet of things. Open Biomed Eng J 2015;9:256-261 [FREE Full text] [doi: 10.2174/1874120701509010256] [Medline: 26628938]
6. Herrmann M, Boehme P, Mondritzki T, Ehlers JP, Kavadias S, Truebel H. Digital transformation and disruption of the health care sector: internet-based observational study. J Med Internet Res 2018 Mar 27;20(3):e104 [FREE Full text] [doi: 10.2196/jmir.9498] [Medline: 29588274]
7. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
8. Software-defined vehicles – a forthcoming industrial evolution. Deloitte. URL: https://www2.deloitte.com/cn/en/pages/consumer-business/articles/software-defined-cars-industrial-revolution-on-the-arrow.html [accessed 2021-12-30]
9. Jararweh Y, Al-Ayyoub M, Darabseh A, Benkhelifa E, Vouk M, Rindos A. Software defined cloud: survey, system and evaluation. Future Gen Comput Syst 2016 May;58:56-74 [FREE Full text] [doi: 10.1016/j.future.2015.10.015]
10. Wild E, Teber D, Schmid D, Simpfendörfer T, Müller M, Baranski A, et al. Robust augmented reality guidance with fluorescent markers in laparoscopic surgery. Int J Comput Assist Radiol Surg 2016 Jun;11(6):899-907. [doi: 10.1007/s11548-016-1385-4] [Medline: 27177759]
11. Wirkert SJ, Kenngott H, Mayer B, Mietkowski P, Wagner M, Sauer P, et al. Robust near real-time estimation of physiological parameters from megapixel multispectral images with inverse Monte Carlo and random forest regression. Int J Comput Assist Radiol Surg 2016 Jun;11(6):909-917 [FREE Full text] [doi: 10.1007/s11548-016-1376-5] [Medline: 27142459]
12. Seitel A, Bellemann N, Hafezi M, Franz AM, Servatius M, Saffari A, et al. Towards markerless navigation for percutaneous needle insertions. Int J Comput Assist Radiol Surg 2016 Jan;11(1):107-117. [doi: 10.1007/s11548-015-1156-7] [Medline: 26018847]
13. Cigaina M. Representing and designing platform businesses. SAP Community. 2017. URL: https://blogs.sap.com/2017/09/05/representing-and-designing-platform-businesses/ [accessed 2021-12-30]
14. Cloud Platform. SAP. URL: https://www.sap.com/sea/products/cloud-platform.html [accessed 2021-12-30]
15. Cloud foundry, the proven development platform for cloud-native applications. Cloud Foundry. URL: https://www.cloudfoundry.org/ [accessed 2021-12-30]
16. SAP subscription billing. SAP. URL: https://www.sap.com/products/subscription-billing.html [accessed 2021-12-30]
17. SAP enterprise consent and preference management. SAP. URL: https://www.sap.com/products/customer-consent.html [accessed 2021-12-30]
18. SAP credential store. SAP Discovery Center. URL: https://discovery-center.cloud.sap/serviceCatalog/credential-store?region=all [accessed 2021-12-30]
19. Sinch digital interconnect LiveLink 365. SAP API Business Hub. URL: https://api.sap.com/api/MobileMessaging/resource [accessed 2021-12-30]
20. Irion KM, Novak P. Systems workplace for endoscopic surgery. Minim Invasive Ther Allied Technol 2000;9(3-4):193-197. [doi: 10.1080/13645700009169647] [Medline: 20156014]
21. IEEE 11073-20701-2018 - IEEE Standard - Health informatics--Point-of-care medical device communication - Part 20701: service-oriented medical device exchange architecture and protocol binding. IEEE Standards Association. URL: https://standards.ieee.org/standard/11073-20701-2018.html [accessed 2021-12-30]
22. Kasparick M, Schlichting S, Golatowski F, Timmermann D. New IEEE 11073 standards for interoperable, networked point-of-care medical devices. Annu Int Conf IEEE Eng Med Biol Soc 2015 Aug;2015:1721-1724. [doi: 10.1109/EMBC.2015.7318709] [Medline: 26736609]
23. Kasparick M, Schmitz M, Andersen B, Rockstroh M, Franke S, Schlichting S, et al. OR.NET: a service-oriented architecture for safe and dynamic medical device interoperability. Biomed Tech (Berl) 2018 Mar 23;63(1):11-30. [doi: 10.1515/bmt-2017-0020] [Medline: 29346114]
24. Welcome to FHIR. HL7 FHIR. URL: https://www.hl7.org/fhir/ [accessed 2021-12-30]
25. IHE patient care device technical framework. IHE International. URL: https://www.ihe.net/uploadedFiles/Documents/PCD/IHE_PCD_TF_Vol1.pdf [accessed 2021-12-30]
26. Current edition. The DICOM Standard. URL: https://www.dicomstandard.org/current [accessed 2021-12-30]
27. Apache Kafka. Kafka. URL: https://kafka.apache.org/ [accessed 2021-12-30]
28. Interoperability in healthcare. HIMSS. URL: https://www.himss.org/resources/interoperability-healthcare [accessed 2021-12-30]
29. Gaudet-Blavignac C, Foufi V, Bjelogrlic M, Lovis C. Use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for processing free text in health care: systematic scoping review. J Med Internet Res 2021 Jan 26;23(1):e24594 [FREE Full text] [doi: 10.2196/24594] [Medline: 33496673]
30. SAP edge services. SAP. URL: https://www.sap.com/products/edge-services.html [accessed 2021-12-30]

31. Taitsman JK, Grimm CM, Agrawal S. Protecting patient privacy and data security. N Engl J Med 2013 Mar 14;368(11):977-979. [doi: 10.1056/NEJMp1215258] [Medline: 23444980]
32. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). EUR-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504 [accessed 2021-12-30]
33. OAuth 2.0. URL: https://oauth.net/2/ [accessed 2021-12-30]
34. SAP business technology platform security. SAP Community. URL: https://community.sap.com/topics/btp-security [accessed 2021-12-30]
35. SAP cloud identity services - identity authentication. SAP. URL: https://help.sap.com/viewer/product/IDENTITY_AUTHENTICATION/Cloud/en-US [accessed 2021-12-30]
36. SAP trust center. SAP. URL: https://www.sap.com/about/trust-center/certification-compliance.html [accessed 2021-12-30]
37. Yezdani M, Yu S, Lee DI. Selective arterial clamping versus hilar clamping for minimally invasive partial nephrectomy. Curr Urol Rep 2016 May;17(5):40. [doi: 10.1007/s11934-016-0596-0] [Medline: 26968420]
38. Chang AY, Skirbekk VF, Tyrovolas S, Kassebaum NJ, Dieleman JL. Measuring population ageing: an analysis of the Global Burden of Disease Study 2017. Lancet Public Health 2019 Mar;4(3):e159-e167 [FREE Full text] [doi: 10.1016/S2468-2667(19)30019-2] [Medline: 30851869]
39. Fani Marvasti F, Stafford RS. From sick care to health care--reengineering prevention into the U.S. system. N Engl J Med 2012 Sep 06;367(10):889-891 [FREE Full text] [doi: 10.1056/NEJMp1206230] [Medline: 22931257]
40. Jagadeeswari V, Subramaniyaswamy V, Logesh R, Vijayakumar V. A study on medical Internet of Things and Big Data in personalized healthcare system. Health Inf Sci Syst 2018 Dec;6(1):14 [FREE Full text] [doi: 10.1007/s13755-018-0049-x] [Medline: 30279984]
41. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 2011 Mar;8(3):184-187. [doi: 10.1038/nrclinonc.2010.227] [Medline: 21364692]
42. Sharon T. The Googlization of health research: from disruptive innovation to disruptive ethics. Per Med 2016 Nov;13(6):563-574. [doi: 10.2217/pme-2016-0057] [Medline: 29754546]
43. Mandl KD, Mandel JC, Murphy SN, Bernstam EV, Ramoni RL, Kreda DA, et al. The SMART Platform: early experience enabling substitutable applications for electronic health records. J Am Med Inform Assoc 2012;19(4):597-603 [FREE Full text] [doi: 10.1136/amiajnl-2011-000622] [Medline: 22427539]
44. Goldman JM, Weininger S, Jaffe MB. Applying medical device informatics to enable safe and secure interoperable systems: medical device interface data sheets. Anesth Analg 2020 Sep;131(3):969-976. [doi: 10.1213/ANE.0000000000004251] [Medline: 31804406]
45. InnOPlan - Innovative, datengetriebene Effizienz OP-übergreifender Prozesslandschaften. InnOPlan. URL: https://innoplan.uni-hohenheim.de/ [accessed 2021-12-30]
46. Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed - an open platform approach to enhance care and research across institutional boundaries. Methods Inf Med 2018 Jul;57(S 01):66-81 [FREE Full text] [doi: 10.3414/ME18-02-0002] [Medline: 30016813]
47. About the initiative. Medical Informatics Initiative Germany. URL: https://www.medizininformatik-initiative.de/en/about-initiative [accessed 2021-12-30]
48. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
49. Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, et al. Deep learning intervention for health care challenges: some biomedical domain considerations. JMIR Mhealth Uhealth 2019 Aug 02;7(8):e11966 [FREE Full text] [doi: 10.2196/11966] [Medline: 31376272]
50. Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics 2015 Jun 27;8:33 [FREE Full text] [doi: 10.1186/s12920-015-0108-y] [Medline: 26112054]
51. Carter RE, Attia ZI, Lopez-Jimenez F, Friedman PA. Pragmatic considerations for fostering reproducible research in artificial intelligence. NPJ Digit Med 2019;2:42 [FREE Full text] [doi: 10.1038/s41746-019-0120-2] [Medline: 31304388]
52. Lovis C. Unlocking the power of artificial intelligence and big data in medicine. J Med Internet Res 2019 Nov 08;21(11):e16607 [FREE Full text] [doi: 10.2196/16607] [Medline: 31702565]
53. Lacombe D, Liu L, Meunier F, Golfinopoulos V. Precision medicine: from "Omics" to economics towards data-driven healthcare - time for European transformation. Biomed Hub 2017;2(Suppl 1):212-221 [FREE Full text] [doi: 10.1159/000480117] [Medline: 31988951]
54. Hype cycle for healthcare providers. Gartner. 2020. URL: https://www.gartner.com/en/documents/3988462-hype-cycle-for-healthcare-providers-2020 [accessed 2021-12-30]
55. Kelly JT, Campbell KL, Gong E, Scuffham P. The internet of things: impact and implications for health care delivery. J Med Internet Res 2020 Nov 10;22(11):e20135 [FREE Full text] [doi: 10.2196/20135] [Medline: 33170132]

Abbreviations

AI: artificial intelligence
API: application programming interface
DKFZ: Deutsches Krebsforschungszentrum (German Cancer Research Center)
HIS: hospital information system
IoT: internet of things
IT: information technology

Edited by G Eysenbach; submitted 21.03.21; peer-reviewed by M Kasparick, C Turcu, C Reis; comments to author 30.04.21; revised version received 25.06.21; accepted 21.11.21; published 20.01.22.

Please cite as:
Görtz M, Byczkowski M, Rath M, Schütz V, Reimold P, Gasch C, Simpfendörfer T, März K, Seitel A, Nolden M, Ross T, Mindroc-Filimon D, Michael D, Metzger J, Onogur S, Speidel S, Mündermann L, Fallert J, Müller M, von Knebel Doeberitz M, Teber D, Seitz P, Maier-Hein L, Duensing S, Hohenfellner M
A Platform and Multisided Market for Translational, Software-Defined Medical Procedures in the Operating Room (OP 4.1): Proof-of-Concept Study
JMIR Med Inform 2022;10(1):e27743
URL: https://medinform.jmir.org/2022/1/e27743
doi:10.2196/27743
PMID:35049510

Original Paper

Prediction of Physical Frailty in Orthogeriatric Patients Using Sensor Insole–Based Gait Analysis and Machine Learning Algorithms: Cross-sectional Study

Moritz Kraus[1]; Maximilian Michael Saller[1], Dr rer nat; Sebastian Felix Baumbach[1], Dr med; Carl Neuerburg[1], Dr med, PD; Ulla Cordula Stumpf[1], Dr med; Wolfgang Böcker[1], Prof Dr med; Alexander Martin Keppler[1], Dr med

Department of Orthopaedics and Trauma Surgery, Musculoskeletal University Center Munich, Ludwig-Maximilians Universität Munich, Munich, Germany

Corresponding Author:
Alexander Martin Keppler, Dr med
Department of Orthopaedics and Trauma Surgery
Musculoskeletal University Center Munich
Ludwig-Maximilians Universität Munich
Marchioninistr. 15
Munich, 81377
Germany
Phone: 49 89 4400 0
Email: alexander.keppler@med.uni-muenchen.de

Abstract

Background: Assessment of the physical frailty of older patients is of great importance in many medical disciplines to be able to implement individualized therapies. For physical tests, time is usually used as the only objective measure. To record other objective factors, modern wearables offer great potential for generating valid data and integrating the data into medical decision-making.

Objective: The aim of this study was to compare the predictive value of insole data, which were collected during the Timed-Up-and-Go (TUG) test, to the benchmark standard questionnaire for sarcopenia (SARC-F: strength, assistance with walking, rising from a chair, climbing stairs, and falls) and physical assessment (TUG test) for evaluating physical frailty, defined by the Short Physical Performance Battery (SPPB), using machine learning algorithms.

Methods: This cross-sectional study included patients aged >60 years with independent ambulation and no mental or neurological impairment. A comprehensive set of parameters associated with physical frailty were assessed, including body composition, questionnaires (European Quality of Life 5-dimension [EQ 5D 5L], SARC-F), and physical performance tests (SPPB, TUG), along with digital sensor insole gait parameters collected during the TUG test. Physical frailty was defined as an SPPB score≤8. Advanced statistics, including random forest (RF) feature selection and machine learning algorithms (K-nearest neighbor [KNN] and RF) were used to compare the diagnostic value of these parameters to identify patients with physical frailty.

Results: Classified by the SPPB, 23 of the 57 eligible patients were defined as having physical frailty. Several gait parameters were significantly different between the two groups (with and without physical frailty). The area under the receiver operating characteristic curve (AUROC) of the TUG test was superior to that of the SARC-F (0.862 vs 0.639). The recursive feature elimination algorithm identified 9 parameters, 8 of which were digital insole gait parameters. Both the KNN and RF algorithms trained with these parameters resulted in excellent results (AUROC of 0.801 and 0.919, respectively).

Conclusions: A gait analysis based on machine learning algorithms using sensor soles is superior to the SARC-F and the TUG test to identify physical frailty in orthogeriatric patients.

(JMIR Med Inform 2022;10(1):e32724) doi:10.2196/32724

KEYWORDS

wearables; insole sensors; orthogeriatric; artificial intelligence; prediction models; machine learning; gait analysis; digital sensors; digital health; aging; prediction algorithms; geriatric; mobile health; mobile insoles

Introduction

The physiological process of aging is inevitably connected to a decrease in physical performance [1]. It has been estimated that approximately 30% of the US population above the age of 55 years suffer from moderate to severe physical limitations [2]. In an orthogeriatric patient population, the assessment of physical frailty is of particular importance, as it is not only strongly associated with falls but also to an inferior outcome following surgery [3]. Consequently, it is of upmost importance to test for and thereby objectify physical impairment (ie, frailty).

Various individual parameters have been proposed to assess physical performance, including handgrip strength, daily step count, and gait speed. However, all of these have considerable interindividual variation [4]. Along with individual physiologic parameters, a variety of questionnaires such as the Barthel index [5], De-Morton Mobility index [6], or FRAIL scale [7] have been developed to quantify frailty. However, these questionnaires have proven to be inferior to the more complex physical assessments [8]. The Short Physical Performance Battery (SPPB) [9] is often considered one of the benchmark tests to assess frailty [8]. The SPPB combines multiple physical assessments, including gait, balance, and strength [10]. There is a consensus that screening for physical frailty is not only the prerequisite for successful individual patient care but also for cost-effectiveness [11]. Nonetheless, an international consensus on the most appropriate screening method is still missing [12].

As outlined above, comprehensive physical stance and gait assessments might be the most effective approach to quantify frailty. A new approach to assess physical activity and gait parameters includes the use of wearables and physical activity monitors [13]. These devices enable physicians and researchers to assess physical activity comprehensively under real-life conditions, and they have already been successfully applied to assist in the diagnosis of musculoskeletal diseases and to monitor rehabilitation [14-17]. A more recent development is sensor insoles with pressure and gyroscope sensors. These insoles can be easily inserted into any shoe and allow for the assessment of several gait parameters in an outpatient setting and also during various daily activities. This might provide a more feasible alternative to time-consuming assessments in specialized gait laboratories.

Although sensor insoles might help in the assessment of frailty, the large number of data points generated necessitates advanced statistical analysis. The random forest (RF) based on decision trees or the K-nearest neighbor (KNN) based on the Euclidean distance between points in high-dimensional space are two suitable strategies to develop clinical decision algorithms [18].

The aim of this study was to compare the classification capability of insole data collected during the Timed-Up-and-Go (TUG) test—a clinical gait test to assess a patient's mobility and risk of falling—to SARC-F (a five-item questionnaire for the quick assessment of the risk of sarcopenia, assessing strength, assistance with walking, rising from a chair, climbing stairs, and falls) and the TUG test to assess physical frailty, defined by the SPPB, using machine learning algorithms.

Methods

Patient Selection

Patients presenting to our orthogeriatric outpatient clinic for an osteoporosis diagnosis or therapy between December 2020 and March 2021 were invited to participate in this study. Inclusion criteria were aged >60 years, independent ambulation without any walking aids, and no mental or neurological impairment. Patients were informed of the study details, including the anonymized evaluation of the collected data, and then provided written consent. This cross-sectional study was approved by the local ethics committee (#19-177).

General Data Assessment

All data were collected in a standardized fashion by a unique, specially trained investigator. Demographic data included age, weight, height, BMI, body composition, general health-related quality of life assessed by the European Quality of Life 5-dimension (EQ-5D-5L) questionnaire [19], and the sarcopenia and physical frailty screening questionnaire SARC-F [20]. All questionnaires were completed together with the patients to obtain the highest possible data quality. Body composition (ie, body fat and muscle percentages) was measured using a clinically validated body composition monitor (BF511, Omron-Healthcare, Kyoto, Japan).

Assessment of Physical Frailty

Physical frailty was assessed by three different means: the SPPB, the TUG test, and digital insole gait parameters assessed during the TUG test using sensor insoles (Science3, Moticon, Munich, Germany).

The SPPB [9] is considered the benchmark test to assess physical frailty and was therefore used as the primary outcome parameter [8]. The SPPB is comprised of multiple tests for gait and stance safety, as well as lower-extremity strength and performance [10]. This tool scores the ability to stand in three different positions for 10 seconds, the time required to walk 3 meters, and the time it takes to rise from and sit down on a chair 5 times. Points are awarded for each subtest according to the time achieved, with a maximum score of 12 and a minimum score of 0. Patients with SPPB scores≤8 are considered to be physically frail [21,22]. The binary SPPB score (not physically frail vs physically frail) was used as the classification label for the machine learning models applied in this study.

The TUG test measures the time a patient takes to rise from a chair (height 46 centimeters), walk 3 meters, turn 180 degrees, and return to their initial seating position [23]. A duration of 12 seconds or longer has been associated with a higher probability of physical frailty [24]. Therefore, a cut-off value of 12 seconds was chosen to classify patients into physically frail and not physically frail groups.

The gait parameters were assessed by Science3 digital sensor insoles during the TUG test. Each of these insoles has 19 pressure sensors and a 3D gyroscope sensor to measure a variety of temporal, spatial, and local gait parameters, including gait speed and pressure distribution [25,26]. The parameters assessed are outlined in detail in Table 1.

Table 1. Overview of all insole gait parameters assessed.

Parameter	Unit
TUG[a] test time	seconds
Steps	number
Mean length of gait line	millimeters
Standard deviation x/y of gait line	meters
Mean total force during stance	Newtons
Mean gait cycle time	seconds
Mean gait cadence	strides/minute
Mean double support time	seconds
Mean acceleration over gait cycle (x/y/z)	*g*
Mean stride length	meters
Mean fraction of stance phase	%
Mean fraction of swing phase	%
Walking distance	meters
Mean walking speed	meters/second
COP[b] variability (left/right)	meters
COP trace length (left/right)	meters

[a]TUG: Timed-Up-and-Go.

[b]COP: center of pressure.

General Statistical Analysis

Unpaired *t* tests were used with α adjustment according to the Benjamini and Hochberg method [27] to compare interval-scaled, normally distributed variables (demographics, questionnaires, and gait parameters) between patients with and without physical frailty. Data are expressed as mean (SD). The effect size is expressed as the standardized mean difference.

Prediction Algorithms

To train the prediction algorithms, all collected performance- and nonperformance-related variables were used to train a recursive feature elimination algorithm that can identify the most relevant parameters for distinguishing patients with (SPPB score≤8) and without (SPPB score>8) physical frailty. For this purpose, the feature elimination algorithm was used to choose the best suitable variables based on an RF algorithm from the ranger package [28]. Gini impurity was used to rank the variables in order of their importance, as this measure is particularly suited to assess how well certain variables divide up a data set [29]. Based on this ordering of the variables, the variables were gradually removed until the lowest possible classification error was achieved. The classification error was chosen as the performance measure for the recursive feature selection, since the main focus was on maximizing the accuracy of the models developed later.

Two supervised machine learning algorithms, KNN [30] and RF, were used for further analysis using the previously selected variables. Both algorithms rely on being trained with labeled training data with a subsequent performance evaluation using test data. Prior to the training and tuning processes, the data were split into a training and a testing data set at a 70:30 ratio. The training process included an internal 3-fold cross-validation step. As hyperparameter tuning is essential for supervised machine learning algorithms to increase the accuracy of the classification [31], both algorithms were subjected to a tuning process that optimizes all variables to be tuned simultaneously, exclusively using the training data set. For the KNN, the tuning range for the number of neighbors was set from 1 to 22. For the type of kernels, the four variants rectangular, Gaussian, rank, and optimal were tested. For the unit of measurement of the distance, the options Euclidean distance, absolute distance, and Minkowski distance were available. For the RF, the number of variables considered as split candidates within a tree was tuned in the range of 1 to 7, the maximum number of branches in a tree was in the range of 2 to 10, and the number of trees in the RF was set from 100 to 1000. The nested resampling technique was used to enable better estimation of the true model performance on unseen data [32]. The 30% of the data not seen by the model were used to compare the performance of the different models subsequently.

To compare the generated algorithms to the classification properties of the TUG and SARC-F, confusion matrices and receiver operating characteristic (ROC) curves were created based on a logistic regression for the SARC-F using solely the score achieved and for the TUG using only the time taken to complete the test so as to compare the different prediction strategies. All data were collected in a REDCap study database [33] and analyzed in a standardized manner with RStudio software (version 1.3.1093), R (version 4.0.3), using the packages dplyr (version 1.0.2), Hmisc (version 4.6-0), ggplot2 (version 3.3.2), caret (version 6.0-86), and mlr3 (version 6.0-86)

[34]. The code used to create and compare the models to the established tests has been made publicly available on GitHub [35].

Results

All of the 57 eligible consecutive orthogeriatric patients were included in the final analysis. The patients' mean age was 77 (SD 6) years and 93% were women. Classified by the SPPB, 23 patients (40%) had physical frailty. Table 2 shows the comparison of all assessed general parameters between the patients with and without physical frailty. Only age, EQ-5D-5L index, and SARC-F score differed significantly between the two groups. It should be emphasized that the average age of the patients with physical frailty was more than 5 years above the average age of the patients without physical frailty. In parallel, the mean health index of the patients with physical frailty determined by the EQ-5D-5L was almost 0.2 points below that of the patients without physical frailty. All other collected demographic data such as weight, height, BMI, body fat, and muscle mass did not differ significantly between the two groups.

The between-groups comparison of the digital gait analysis is presented in Table 3. The two groups differed significantly for all insole-generated gait parameters (all $P<.05$).

The classification errors of the TUG test and SARC-F to identify patients with physical frailty were 0.333 and 0.316, respectively. However, the area under the ROC curve (AUROC) for the TUG test was higher when compared with that of the SARC-F (0.862 vs 0.639; Figure 1A, Figure 1B).

The RF-based recursive feature elimination algorithm was trained to extract the most important features for classifying physical frailty using all parameters collected, except the SPPB, TUG test, and SARC-F, as they either define the result or represent the classification methods to be compared.

Based on the defined criteria, the 9 parameters outlined in Figure 2 were included. Notably, 8 out of the 9 parameters selected were gait parameters collected by the insoles (Figure 2). The number of steps and the step length were the most decisive factors for the identification of physical frailty by the algorithm. The gait speed followed in third place. Of the variables selected, double support seemed to have the least effect on classification.

These variables were then used to train the two classification algorithms KNN and RF. The tuning process resulted in an optimal combination of hyperparameters for the KNN as follows: k=15, a "rank" kernel, and the Minkowski distance. The optimal combination for the RF was 7 split variables, 6 branches, and 550 trees.

To compare the classification abilities of the TUG and the SARC-F with the algorithms created, a logistic regression was carried out on the SARC-F score and the TUG time on the dependent variable physical frailty and the ROC curve was drawn (Figure 1A-D). Table 4 summarizes the prediction accuracy of the four classifiers. Both classical approaches were outperformed by the machine learning–based models in terms of classification error (KNN=0.246, Figure 1D; RF=0.281, Figure 1C). The AUROC for the RF was slightly superior to that of the KNN (Table 4). Overall, the KNN showed the lowest error rate in classification at 24.6% (Figure 1). RF showed the largest AUROC value and thus appears to be the most suitable for classification. In the conventional tests, the TUG test was far superior to the SARC-F in terms of area under the ROC curve and classification error. The KNN showed the lowest classification error rate, but had a slightly smaller AUROC value than those of the RF and the TUG test.

Table 2. Comparison of demographics, body composition, physical activity, physical performance, and health questionnaire scores between patients with and without physical frailty.

Variable	No physical frailty (n=34)	Physical frailty (n=23)	*P* value	SMD[a]
Age (years), mean (SD)	74.76 (5.92)	80.00 (5.82)	.002	0.892
BMI (kg/m^2), mean (SD)	24.42 (4.81)	24.66 (3.79)	.84	0.055
Height (cm), mean (SD)	160.94 (6.37)	160.56 (7.84)	.85	0.053
Weight (kg), mean (SD)	62.77 (9.72)	63.45 (9.61)	.80	0.070
Body fat (%), mean (SD)	30.15 (8.55)	32.14 (7.86)	.37	0.243
Visceral fat (%), mean (SD)	7.95 (3.21)	8.71 (2.72)	.34	0.254
Muscle mass (%), mean (SD)	30.26 (4.20)	28.52 (3.29)	.09	0.460
Resting metabolism (kcal), mean (SD)	1345.32 (110.40)	1341.29 (123.22)	.90	0.034
Calf circumference, mean (SD)	35.04 (3.12)	34.31 (3.30)	.41	0.228
EQ-5D-5L[b] index, mean (SD)	0.84 (0.16)	0.65 (0.27)	.007	0.818
SPPB[c] score (points), mean (SD)	11.30 (0.79)	6.44 (2.06)	<.001	–3.106
SPPB score≤8, n (%)	0 (0)	23 (40)	<.001	
SARC-F[d] score, n (%)			.01	1.002
0	22 (65)	6 (26)		
1	8 (24)	7 (30)		
2	2 (6)	3 (13)		
3	0 (0)	4 (17)		
4	2 (6)	3 (13)		
Number of falls in past year, n (%)			.31	0.422
0	24 (71)	12 (52)		
1-3	7 (21)	9 (39)		
>3	3 (9)	2 (9)		
BMD[e] femoral neck (g/cm^3), mean (SD)	0.61 (0.06)	0.59 (0.06)	.27	0.303
BMD lumbar spine (g/cm^3), mean (SD)	0.85 (0.12)	0.91 (0.16)	.17	0.391
Smoking, n (%)			>.99	0.005
No	31 (91)	21 (91)		
Yes	3 (9)	2 (9)		
Self-sustaining, n (%)			.74	0.103
No	6 (18)	5 (22)		
Yes	28 (82)	18 (78)		
Daily leaving apartment, n (%)			.05	0.566
No	4 (12)	8 (35)		
Yes	30 (88)	15 (65)		
Weekly sports activity (>3 h), n (%)			.06	0.569
No	10 (29)	13 (57)		
Yes	24 (71)	10 (43)		

[a]SMD: standardized mean difference.

[b]EQ-5D-5L: European Quality of Life 5-dimension questionnaire.

[c]SPPB: Short Physical Performance Battery.

[d]SARC-F: sarcopenia test (strength, assistance with walking, rising from a chair, climbing stairs, and fall).

[e]BMD: bone mineral density.

Table 3. Comparison of gait parameters between patients with and without physical frailty.

Variable	No physical frailty, mean (SD)	Physical frailty, mean (SD)	*P* value	SMD[a]
Mean gait speed (m/s)	1.09 (0.28)	0.69 (0.19)	<.001	–1.637
TUG[b] time (s)	8.52 (1.93)	15.79 (5.50)	<.001	1.765
Mean stride length (m)	1.12 (0.19)	0.85 (0.17)	<.001	–1.450
Mean gait cadence (strides/min)	59.72 (8.83)	49.37 (8.21)	<.001	–1.214
Mean gait cycle time (s)	1.05 (0.16)	1.27 (0.20)	<.001	1.199
Mean double support time (s)	0.40 (0.13)	0.51 (0.14)	.003	0.843
Number of steps (n)	15.32 (6.05)	20.04 (5.67)	.005	0.804
Mean acceleration over gait cycle right (*g*)	0.03 (0.89)	0.59 (0.74)	.02	0.695
COP[c] trace length right (m)	5.25 (1.96)	7.06 (3.22)	.02	0.680
Mean acceleration over gait cycle right (*g*)	–2.36 (1.32)	–1.39 (1.54)	.02	0.672
Mean length width of gait line right (mm)	131.10 (21.20)	142.66 (19.05)	.04	0.574
Variance of acceleration over gait cycle (m/s^2)	1.66 (0.86)	1.21 (0.78)	.05	–0.552

[a]SMD: standardized mean difference.

[b]TUG: Timed-Up-and-Go.

[c]COP: center of pressure.

Figure 1. Comparison of the receiver operating characteristic (ROC) curves of the classification properties of the sarcopenia index SARC-F (A), Timed-Up-and-Go (TUG) test (B), and the random forest (C) and k-nearest neighbor (D) algorithms. AUC: area under the ROC curve.

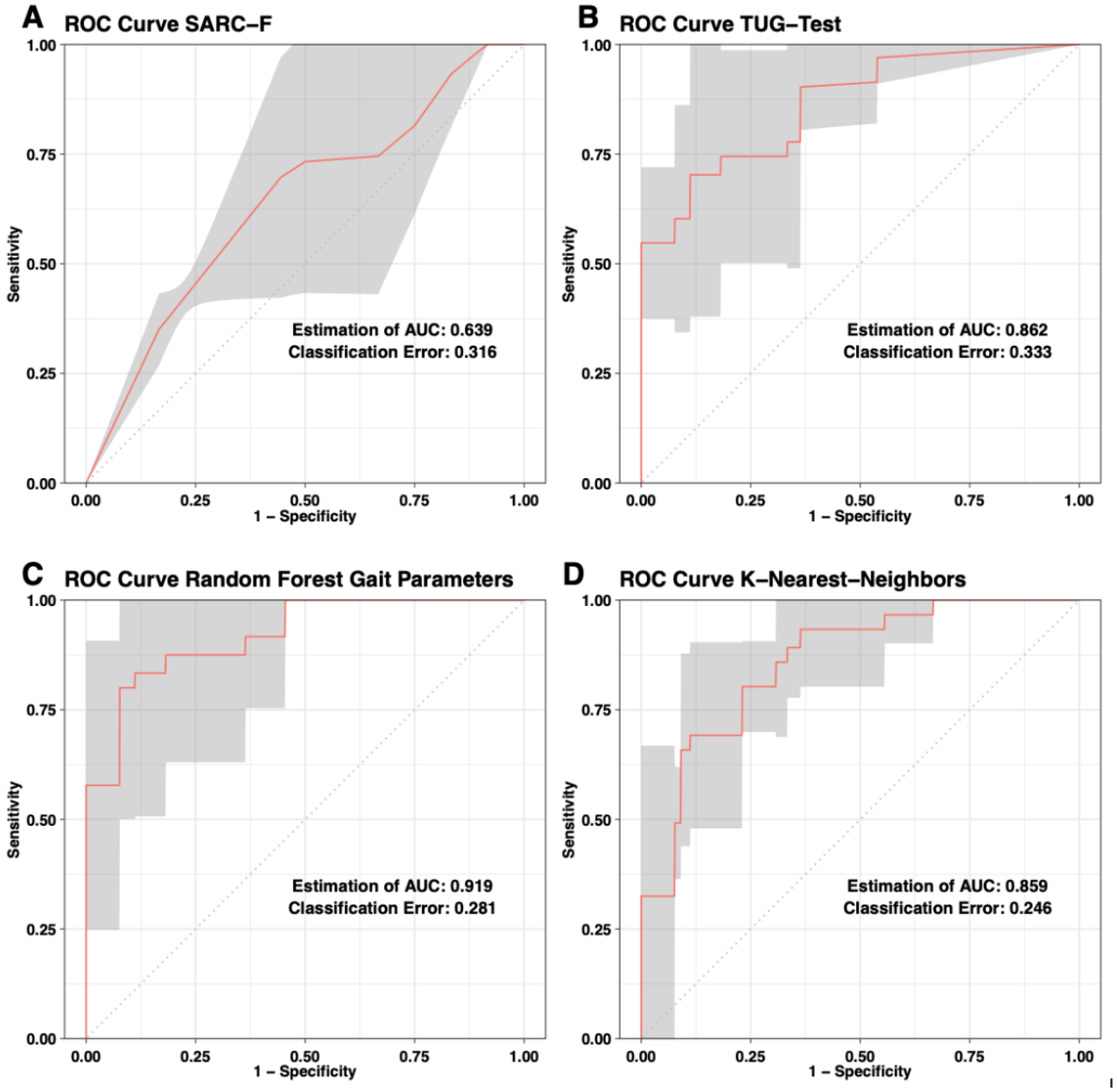


Figure 2. Selected parameters based on the recursive feature elimination algorithm, ordered by their importance for reduction of classification error ranked by Gini-Impurity [29].

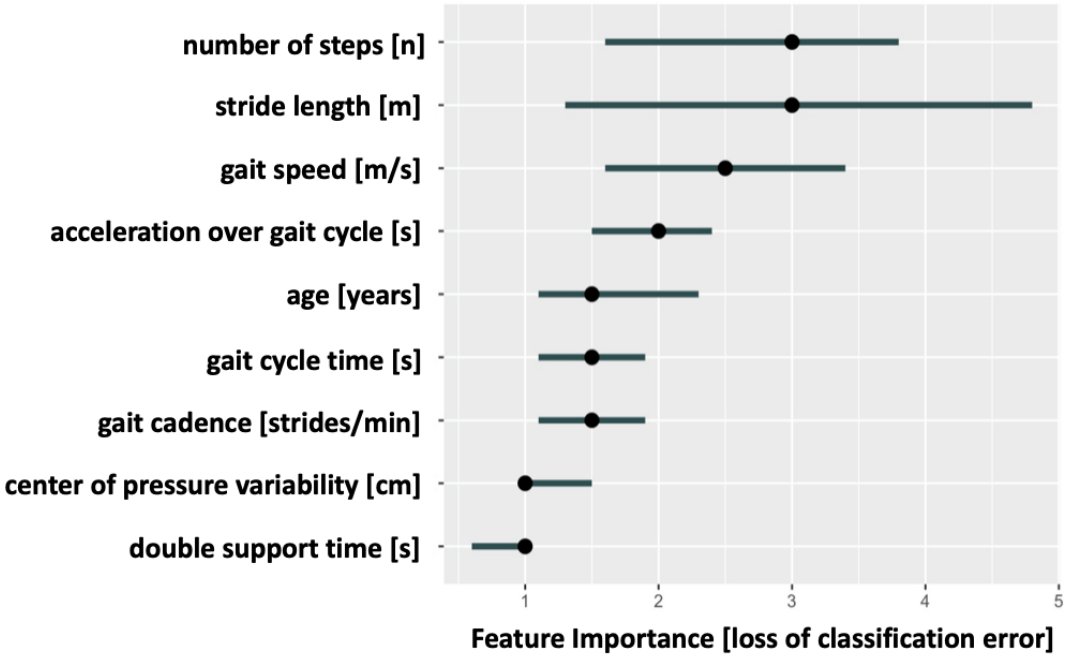


Table 4. Comparison of physical frailty prediction methods.

Performance metric	SARC-F[a] LR[b]	TUG[c] test LR	KNN[d] classifier	RF[e] classifier
Accuracy	0.684	0.667	0.719	0.724
AUROC[f]	0.639	0.862	0.919	0.859

[a]SARC-F: sarcopenia test (strength, assistance with walking, rising from a chair, climbing stairs, and fall).

[b]LR: logistic regression.

[c]TUG: Timed-Up-and-Go.

[d]KNN: K-nearest neighbor.

[e]RF: random forest.

[f]AUROC: area under the receiver operating characteristic curve.

Discussion

Principal Findings

Based on a sample of 57 patients and advanced statistics, this study shows that gait parameters assessed by digital insoles during the TUG test outperformed both the benchmark tests (the TUG physical assessment and SARC-F questionnaire) to identify patients with physical frailty.

Patients identified as physically frail classified by their SPPB scores (≤8) were on average 5 years older than patients that were not classified as physically frail, with no significant difference in BMI or body composition. By contrast, previous studies have reported a decreased muscle mass and increased fat percentage in patients with physical frailty [36]. Despite the considerable amount of physical frailty–related data collected (Tables 1 and 2), the vast majority (8 out of 9) of the parameters selected by the recursive feature elimination algorithm were insole gait parameters collected during the TUG test. Although the temporal gait variables such as gait speed, double support time, and gait cadence can be considered dependent variables, they all reflect different aspects of gait. For this reason, it makes sense to integrate several of these aspects into the machine learning algorithms to better map the gait pattern of an individual patient and derive the best possible classification.

Previous studies have proposed that gait speed is the most relevant parameter to identify patients with physical frailty [4]. It has been shown that a slow gait speed is associated with an increased fall risk [37], as well as a higher mortality rate [38]. Interestingly, the advanced modeling used in this study weighted stride length equally important as gait speed to differentiate between physical frailty and no physical frailty in patients, in terms of their classification importance measured by the Gini impurity (Figure 2). Although gait speed is easily assessed, it might be biased by patients' motivation. One can hypothesize a "white coat effect," in this case a higher level of motivation during medical gait speed examinations. Stride length might be a more robust (ie, harder to influence consciously) parameter in such settings, which might explain its superiority in the herein

applied modeling. Espy et al [39] provided a possible explanation for the higher robustness of stride length compared to gait speed. They were able to show that a slow gait leads to instability, which again is compensated for by a small-stepped gait pattern [39]. It appears reasonable that patients with physical frailty would therefore compensate for their unstable gait pattern by a reduction of their stride length [39]. Overall, stride length and gait speed were found to be the two most relevant parameters for the model (Figure 2), and could only be slightly increased by adding additional gait parameters such as cadence, double support time, and acceleration over gait cycle. Consequently, stride length in addition to gait speed might be a valuable clinical parameter to identify patients with physical frailty. Their early identification is essential to reduce the number of falls [37] and possibly mortality rates [38], as well as to increase further health outcomes [40]. These considerable implications are not only important in an orthogeriatric setting but also for almost all medical specialties.

In line with previous studies, the SARC-F as well as the TUG test were found to be suitable for estimating the physical frailty status [41]. The slightly better results for the TUG test compared with the SARC-F might be explained by their different natures. The SARC-F is a patient-reported outcome measure, whereas the TUG test is a more objective score. Older patients have been shown to overestimate their physical abilities [42,43], which might result in false negative SARF-F scores. Complementing the SARC-F by an objective measurement such as the TUG test, handgrip strength, or a gait analysis might increase its accuracy and therefore screening value.

Nevertheless, the combination of machine learning algorithms and digital gait analysis outperformed the TUG test and SARC-F in the detection of physical frailty. The digital insoles used in this study can easily be applied and have proven to be reliable [25]. Furthermore, they could be integrated into health assessment apps, such as on a smartphone. This can facilitate both the collection of longitudinal data and remote monitoring of at-risk patients, and potentially even guide rehabilitation. Consequently, gait analysis by digital insoles might become another valuable part of the growing body of digital health devices.

Limitations and Strengths

An obvious limitation of this study is the limited number of patients. The smaller the number of patients the algorithm is trained on, the more limited is its generalizability. Therefore, the herein proposed algorithm must be validated in a larger cohort. In the setting of a longitudinal, multicenter trial, the applied statistics could be extended to deep learning methods such as neural networks, which could further increase the accuracy of the predictions. Another limitation is the definition of physical frailty. Due to the current setup, it was only possible to define physical frailty by the SPPB. Although the SPBB is considered one of the benchmark tests for physical frailty [44], it would be even more meaningful to directly assess the occurrence of various health impairments such as falls, fractures, progression to impaired ambulation, or death. Nonetheless, these parameters can only be assessed in a longitudinal study setup.

Despite these limitations, several strengths of this study are noteworthy. First, the combined use of modern wearables and data analysis strategies from the field of data science to complement the classic statistical analysis is an advantage of this study. Due to the increasing amount of data points collected by digital devices, advanced statistics will become the primary working horse to analyze the data. Second, the meta-modeling approach applied represents a pessimistic estimation of the models' performance in a larger cohort. Nevertheless, the resulting AUROC values of 0.801 and 0.841 can be judged as excellent [45]. These excellent results argue for the value of digital insole gait parameters. For application in clinical practice, it is conceivable that a doctor will receive an analysis on their terminal device in real time during the test, which can provide time-efficient support in clinical decision-making for or against prescribing fall prevention training, certain medications, or other therapeutic interventions. Finally, this study also indicates that gait parameters might be a promising target for physical frailty therapies. It can by hypothesized that focused physiotherapy or fall risk minimization counseling could counteract physical frailty and thereby increase the patient's health-related quality of life.

Conclusion

Machine learning algorithms–based gait analysis using mobile insoles appears to be a promising approach to screen for physical frailty in an outpatient setting. Due to the increasing amount of data collected, high-performance data processing will become increasingly important. Future large-scale, longitudinal, and multicenter screening trials should collect as many data points as possible, including from digital devices such as wearables, and apply advanced statistics to increase the diagnostic sensitivity and accuracy of physical frailty diagnosis.

Conflicts of Interest

None declared.

References

1. Cruz-Jimenez M. Normal changes in gait and mobility problems in the elderly. Phys Med Rehabil Clin N Am 2017 Nov;28(4):713-725. [doi: 10.1016/j.pmr.2017.06.005] [Medline: 29031338]
2. Tieland M, Trouwborst I, Clark BC. Skeletal muscle performance and ageing. J Cachexia Sarcopenia Muscle 2018 Feb;9(1):3-19. [doi: 10.1002/jcsm.12238] [Medline: 29151281]
3. Marquez A, Queirós C. Frailty, sarcopenia and falls. In: Hertz K, Santy-Tomlinson J, editors. Fragility fracture nursing: holistic care and management of the orthogeriatric patient. Cham: Springer; 2018:15-26.

4. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait speed and survival in older adults. JAMA 2011 Jan 05;305(1):50-58 [FREE Full text] [doi: 10.1001/jama.2010.1923] [Medline: 21205966]
5. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965 Feb;14:61-65. [Medline: 14258950]
6. de Morton NA, Davidson M, Keating JL. The de Morton Mobility Index (DEMMI): an essential health index for an ageing world. Health Qual Life Outcomes 2008 Aug 19;6:63 [FREE Full text] [doi: 10.1186/1477-7525-6-63] [Medline: 18713451]
7. Thompson MQ, Theou O, Tucker GR, Adams RJ, Visvanathan R. FRAIL scale: Predictive validity and diagnostic test accuracy. Australas J Ageing 2020 Dec;39(4):e529-e536. [doi: 10.1111/ajag.12829] [Medline: 32748992]
8. Freiberger E, de Vreede P, Schoene D, Rydwik E, Mueller V, Frändin K, et al. Performance-based physical function in older community-dwelling persons: a systematic review of instruments. Age Ageing 2012 Nov;41(6):712-721. [doi: 10.1093/ageing/afs099] [Medline: 22885845]
9. Treacy D, Hassett L. The Short Physical Performance Battery. J Physiother 2018 Jan;64(1):61 [FREE Full text] [doi: 10.1016/j.jphys.2017.04.002] [Medline: 28645532]
10. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994 Mar;49(2):M85-M94. [doi: 10.1093/geronj/49.2.m85] [Medline: 8126356]
11. Bleijenberg N, Drubbel I, Neslo RE, Schuurmans MJ, Ten Dam VH, Numans ME, et al. Cost-effectiveness of a proactive primary care program for frail older people: a cluster-randomized controlled trial. J Am Med Dir Assoc 2017 Dec 01;18(12):1029-1036.e3 [FREE Full text] [doi: 10.1016/j.jamda.2017.06.023] [Medline: 28801235]
12. Dent E, Martin FC, Bergman H, Woo J, Romero-Ortuno R, Walston JD. Management of frailty: opportunities, challenges, and future directions. Lancet 2019 Oct 12;394(10206):1376-1386. [doi: 10.1016/S0140-6736(19)31785-4] [Medline: 31609229]
13. De Luca V, Muaremi A, Giggins O, Walsh L, Clay I. Towards fully instrumented and automated assessment of motor function tests. 2018 Mar 01 Presented at: IEEE EMBS Int Conf Biomed Heal Informatics; 2018; Las Vegas p. 83. [doi: 10.1109/bhi.2018.8333375]
14. Keppler AM, Nuritidinow T, Mueller A, Hoefling H, Schieker M, Clay I, et al. Validity of accelerometry in step detection and gait speed measurement in orthogeriatric patients. PLoS One 2019;14(8):e0221732 [FREE Full text] [doi: 10.1371/journal.pone.0221732] [Medline: 31469864]
15. Pfeufer D, Becker CA, Faust L, Keppler AM, Stagg M, Kammerlander C, et al. Load-bearing detection with insole-force sensors provides new treatment insights in fragility fractures of the pelvis. J Clin Med 2020 Aug 06;9(8):2551 [FREE Full text] [doi: 10.3390/jcm9082551] [Medline: 32781675]
16. Keppler AM, Holzschuh J, Pfeufer D, Neuerburg C, Kammerlander C, Böcker W, et al. Postoperative physical activity in orthogeriatric patients - new insights with continuous monitoring. Injury 2020 Mar;51(3):628-632. [doi: 10.1016/j.injury.2020.01.041] [Medline: 32033808]
17. Mueller A, Hoefling HA, Muaremi A, Praestgaard J, Walsh LC, Bunte O, et al. Continuous digital monitoring of walking speed in frail elderly patients: noninterventional validation study and longitudinal clinical trial. JMIR Mhealth Uhealth 2019 Nov 27;7(11):e15191 [FREE Full text] [doi: 10.2196/15191] [Medline: 31774406]
18. Garg A, Mago V. Role of machine learning in medical research: a survey. Comput Sci Rev 2021 May;40:100370. [doi: 10.1016/j.cosrev.2021.100370]
19. Hernandez G, Garin O, Dima AL, Pont A, Martí Pastor M, Alonso J, ASTRO-LAB Group. EuroQol (EQ-5D-5L) validity in assessing the quality of life in adults with asthma: cross-sectional study. J Med Internet Res 2019 Jan 23;21(1):e10178 [FREE Full text] [doi: 10.2196/10178] [Medline: 30672744]
20. Hax V, do Espírito Santo RC, Dos Santos LP, Farinon M, de Oliveira MS, Três GL, et al. Practical screening tools for sarcopenia in patients with systemic sclerosis. PLoS One 2021;16(1):e0245683 [FREE Full text] [doi: 10.1371/journal.pone.0245683] [Medline: 33481872]
21. Yuguchi S, Saitoh M, Oura K, Tahara M, Kamisaka K, Kawamura T, et al. Impact of preoperative frailty on regaining walking ability in patients after cardiac surgery: Multicenter cohort study in Japan. Arch Gerontol Geriatr 2019;83:204-210. [doi: 10.1016/j.archger.2019.04.003] [Medline: 31082565]
22. Lim YJ, Ng YS, Sultana R, Tay EL, Mah SM, Chan CHN, et al. Frailty assessment in community-dwelling older adults: a comparison of 3 diagnostic instruments. J Nutr Health Aging 2020;24(6):582-590. [doi: 10.1007/s12603-020-1396-2] [Medline: 32510110]
23. Miyoshi S, Morita T, Kadoi Y, Goto F. Analysis of the factors related to a decrease in jugular venous oxygen saturation in patients with diabetes mellitus during normothermic cardiopulmonary bypass. Surg Today 2005;35(7):530-534. [doi: 10.1007/s00595-004-2977-0] [Medline: 15976948]
24. Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/steadi/pdf/TUG_Test-print.pdf [accessed 2021-05-10]
25. Braun BJ, Veith NT, Hell R, Döbele S, Roland M, Rollmann M, et al. Validation and reliability testing of a new, fully integrated gait analysis insole. J Foot Ankle Res 2015;8(1):54 [FREE Full text] [doi: 10.1186/s13047-015-0111-8] [Medline: 26396594]

26. Braun BJ, Bushuven E, Hell R, Veith NT, Buschbaum J, Holstein JH, et al. A novel tool for continuous fracture aftercare - Clinical feasibility and first results of a new telemetric gait analysis insole. Injury 2016 Feb;47(2):490-494. [doi: 10.1016/j.injury.2015.11.004] [Medline: 26626806]
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
28. Wright MN, Ziegler A. ranger: a fast implementation of random forests for high dimensional data in C++ and R and. J Stat Soft 2017 Mar 01;77(1):1-17. [doi: 10.18637/jss.v077.i01]
29. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. Comput Stat Data Anal 2008 Jan;52(4):2249-2260. [doi: 10.1016/j.csda.2007.08.015]
30. Keller JM, Gray MR, Givens JA. A fuzzy K-nearest neighbor algorithm. IEEE Trans Syst Man Cybern 1985 Jul;SMC-15(4):580-585. [doi: 10.1109/tsmc.1985.6313426]
31. van Rijn J, Hutter F. Hyperparameter importance across datasets. : ACM; 2018 Aug 19 Presented at: 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; August 19-23, 2018; London URL: https://dl.acm.org/doi/10.1145/3219819.3220058 [doi: 10.1145/3219819.3220058]
32. Bischl B, Mersmann O, Trautmann H, Weihs C. Resampling methods for meta-model validation with recommendations for evolutionary computation. Evol Comput 2012 Jun;20(2):249-275. [doi: 10.1162/evco_a_00069]
33. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
34. Lang M, Binder M, Richter J, Schratz P, Pfisterer F, Coors S, et al. mlr3: A modern object-oriented machine learning framework in R. JOSS 2019 Dec;4(44):1903. [doi: 10.21105/joss.01903]
35. Kraus M. Models for gait analysis evaluation Internet. GitHub. URL: https://github.com/KrausMoritz/Gait_Analysis [accessed 2021-10-01]
36. Cesari M, Leeuwenburgh C, Lauretani F, Onder G, Bandinelli S, Maraldi C, et al. Frailty syndrome and skeletal muscle: results from the Invecchiare in Chianti study. Am J Clin Nutr 2006 May 01;83(5):1142-1148 [FREE Full text] [doi: 10.1093/ajcn/83.5.1142] [Medline: 16685058]
37. Verghese J, Holtzer R, Lipton RB, Wang C. Quantitative gait markers and incident fall risk in older adults. J Gerontol A Biol Sci Med Sci 2009 Aug 06;64(8):896-901 [FREE Full text] [doi: 10.1093/gerona/glp033] [Medline: 19349593]
38. Stanaway FF, Gnjidic D, Blyth FM, Le Couteur DG, Naganathan V, Waite L, et al. How fast does the Grim Reaper walk? Receiver operating characteristics curve analysis in healthy men aged 70 and over. BMJ 2011 Dec 15;343(dec15 1):d7679-d7679 [FREE Full text] [doi: 10.1136/bmj.d7679] [Medline: 22174324]
39. Espy D, Yang F, Bhatt T, Pai Y. Independent influence of gait speed and step length on stability and fall risk. Gait Posture 2010 Jul;32(3):378-382 [FREE Full text] [doi: 10.1016/j.gaitpost.2010.06.013] [Medline: 20655750]
40. Pandey A, Kitzman D, Whellan DJ, Duncan PW, Mentz RJ, Pastva AM, et al. Frailty among older decompensated heart failure patients: prevalence, association with patient-centered outcomes, and efficient detection methods. JACC Heart Fail 2019 Dec;7(12):1079-1088 [FREE Full text] [doi: 10.1016/j.jchf.2019.10.003] [Medline: 31779931]
41. Si H, Jin Y, Qiao X, Tian X, Liu X, Wang C. Comparison of 6 frailty screening tools in diagnostic properties among Chinese community-dwelling older people. Geriatr Nurs 2021 Jan;42(1):276-282. [doi: 10.1016/j.gerinurse.2020.08.017] [Medline: 32948340]
42. Kawasaki T, Tozawa R. Motor function relating to the accuracy of self-overestimation error in community-dwelling older adults. Front Neurol 2020;11:599787. [doi: 10.3389/fneur.2020.599787] [Medline: 33329358]
43. Martin P, Keppler AM, Alberton P, Neuerburg C, Drey M, Böcker W, et al. Self-Assessment of mobility of people over 65 years of age. Medicina (Kaunas) 2021 Sep 17;57(9):980 [FREE Full text] [doi: 10.3390/medicina57090980] [Medline: 34577903]
44. Chen J, Liang C, Chang Q. Comparison of fallers and nonfallers on four physical performance tests: a prospective cohort study of community-dwelling older indigenous Taiwanese women. Int J Geront 2018 Mar;12(1):22-26. [doi: 10.1016/j.ijge.2017.04.006]
45. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010 Sep;5(9):1315-1316 [FREE Full text] [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]

Abbreviations

AUROC: area under the receiver operating characteristic curve
EQ-5D-5L: European Quality of Life 5-dimension
KNN: K-nearest neighbor
RF: random forest
ROC: receiver operating characteristic
SARC-F: sarcopenia questionnaire (strength, assistance with walking, rising from a chair, climbing stairs, and falls)

SPPB: Short Physical Performance Battery
TUG: Timed Up and Go

Edited by C Lovis, J Hefner; submitted 07.08.21; peer-reviewed by I Clay, SD Boie; comments to author 19.09.21; revised version received 29.10.21; accepted 10.11.21; published 05.01.22.

Please cite as:
Kraus M, Saller MM, Baumbach SF, Neuerburg C, Stumpf UC, Böcker W, Keppler AM
Prediction of Physical Frailty in Orthogeriatric Patients Using Sensor Insole–Based Gait Analysis and Machine Learning Algorithms: Cross-sectional Study
JMIR Med Inform 2022;10(1):e32724
URL: https://medinform.jmir.org/2022/1/e32724
doi:10.2196/32724
PMID:34989684

Original Paper

Patient Representation Learning From Heterogeneous Data Sources and Knowledge Graphs Using Deep Collective Matrix Factorization: Evaluation Study

Sajit Kumar[1], MBA; Alicia Nanelia[2], BSc; Ragunathan Mariappan[2], MSc; Adithya Rajagopal[3], BTech; Vaibhav Rajan[2], PhD

[1]Great Learning, Bengaluru, India

[2]Department of Information Systems and Analytics, National University of Singapore, Singapore, Singapore

[3]National Institute of Technology, Thiruchirappalli, India

Corresponding Author:
Vaibhav Rajan, PhD
Department of Information Systems and Analytics
National University of Singapore
13 Computing Drive
Singapore, 117417
Singapore
Phone: 65 65166737
Email: vaibhav.rajan@nus.edu.sg

Abstract

Background: Patient representation learning aims to learn features, also called representations, from input sources automatically, often in an unsupervised manner, for use in predictive models. This obviates the need for cumbersome, time- and resource-intensive manual feature engineering, especially from unstructured data such as text, images, or graphs. Most previous techniques have used neural network–based autoencoders to learn patient representations, primarily from clinical notes in electronic medical records (EMRs). Knowledge graphs (KGs), with clinical entities as nodes and their relations as edges, can be extracted automatically from biomedical literature and provide complementary information to EMR data that have been found to provide valuable predictive signals.

Objective: This study aims to evaluate the efficacy of collective matrix factorization (CMF), both the classical variant and a recent neural architecture called deep CMF (DCMF), in integrating heterogeneous data sources from EMR and KG to obtain patient representations for clinical decision support tasks.

Methods: Using a recent formulation for obtaining graph representations through matrix factorization within the context of CMF, we infused auxiliary information during patient representation learning. We also extended the DCMF architecture to create a task-specific end-to-end model that learns to simultaneously find effective patient representations and predictions. We compared the efficacy of such a model to that of first learning unsupervised representations and then independently learning a predictive model. We evaluated patient representation learning using CMF-based methods and autoencoders for 2 clinical decision support tasks on a large EMR data set.

Results: Our experiments show that DCMF provides a seamless way for integrating multiple sources of data to obtain patient representations, both in unsupervised and supervised settings. Its performance in single-source settings is comparable with that of previous autoencoder-based representation learning methods. When DCMF is used to obtain representations from a combination of EMR and KG, where most previous autoencoder-based methods cannot be used directly, its performance is superior to that of previous nonneural methods for CMF. Infusing information from KGs into patient representations using DCMF was found to improve downstream predictive performance.

Conclusions: Our experiments indicate that DCMF is a versatile model that can be used to obtain representations from single and multiple data sources and combine information from EMR data and KGs. Furthermore, DCMF can be used to learn representations in both supervised and unsupervised settings. Thus, DCMF offers an effective way of integrating heterogeneous data sources and infusing auxiliary knowledge into patient representations.

(JMIR Med Inform 2022;10(1):e28842) doi:10.2196/28842

KEYWORDS

representation learning; deep collective matrix factorization; electronic medical records; knowledge graphs; multiview learning; graph embeddings; clinical decision support

Introduction

Background

Machine learning–based predictive models have been found to be highly accurate in many clinical decision support tasks. Examples include predictions of unforeseen complications [1], patient severity assessment through mortality predictors [2] and automated coding for billing [3], and prediction of patient outcomes [4], to name a few. The key ingredients of these models are the features used to describe patients for whom predictions are required. The traditional approach for building these features is to handcraft them typically in collaboration with a domain expert. However, with the growing amount, complexity, and diversity of clinical information sources, such manual feature engineering is practically infeasible. For instance, in electronic medical records (EMRs), patient information may be distributed among laboratory tests, nursing notes, radiology images and reports, genomic data, and other data sources.

Representation learning aims to learn features or representations from the given input sources automatically, often in an unsupervised manner. This obviates the need for manual feature engineering and is particularly useful with unstructured data sources such as clinical notes. These real-valued vectorial representations can be used as features directly in machine learning models for various downstream tasks such as prediction or cluster detection. Such representation learning has been found to be effective in several predictive models, for example, disease category prediction [5] and mortality prediction [6].

Previous studies have primarily used clinical notes to learn patient representations. Clinical notes are a rich source of information containing detailed subjective and objective evaluations of patient conditions during the hospital stay. Some previous studies have also combined other structured tables from EMR with features extracted from notes to obtain patient representations [1,5] or to mine clinical information such as drug mentions [7]. Many of these studies have used variants of deep neural architecture based on autoencoders to obtain unsupervised patient representations.

When information from multiple heterogeneous sources is available, predictive models benefit from latent representations that systematically model correlated shared structures. The aim of multi-view learning is to effectively build such latent representations, where views refer to measurements for the same subjects that differ in source, datatype, or modality; heterogeneous data sources within EMR provide such multiple views of patients. A general technique for multi-view representation learning from arbitrary collections of heterogeneous data sources is collective matrix factorization (CMF) [8]. CMF can be used to obtain patient representations from multi-view EMR data and can also be used to seamlessly integrate auxiliary information from external sources.

One such auxiliary source of information is a clinical knowledge graph (KG) that has been found to be valuable for improving both the accuracy and interpretability of predictive models. These KGs have clinical entities (eg, diseases, drugs, and biomolecules) as nodes and different kinds of relations (eg, treats, predisposes, and causes) as edges. They can be automatically created from various sources such as biomedical literature and web-based health portals. Representation learning methods have also been developed for graph inputs that can automatically learn vectorial representations of nodes to incorporate the global structural and semantic properties of the graph. These node representations can then be used in machine learning models for graph analytics such as community detection or node classification. Owing to its wide applicability, a large number of graph representation learning techniques have been developed for various classes of graphs, including KGs.

In this paper, we analyze patient representation learning in light of 2 recent advances in CMF and KG representation learning. A deep autoencoder-based architecture, called deep CMF (DCMF), was developed for CMF, which was found to outperform classical nonneural variants of CMF in several tasks [9]. Using DCMF, which provides a seamless way of integrating heterogeneous data, we evaluate the effectiveness of patient representations when the input data are augmented with additional information from literature-derived KGs. The generality of DCMF allows many different ways of using KG as inputs; however, not all of them are equally effective. Recently, it has been shown that many graph representation learning methods can be reformulated as a matrix factorization problem. Leveraging this formulation within the context of CMF and DCMF, we infuse auxiliary information during patient representation learning. To our knowledge, this is the first study to use this technique to obtain clinical KG representations and use it within the DCMF framework to obtain patient representations.

Furthermore, the DCMF architecture can easily be extended to create a task-specific end-to-end model that learns to simultaneously find effective patient representations and predictions. We also compare the efficacy of such a model to that of a 2-stage process of first learning unsupervised representations and then independently learning a predictive model.

We rigorously evaluate patient representation learning using DCMF-based methods and autoencoders for 2 clinical decision support tasks on EMR data comprising 28,563 patient episodes. The first task is that of primary diagnosis category prediction, which is performed during coding from discharge summaries when a patient is discharged from the hospital for billing and reimbursement purposes. The second task is that of mortality (risk of death) prediction, which can be used to identify high-risk patients and prioritize their care.

The utility of DCMF-based patient representations, obtained from only EMR data and a combination of KGs and EMR data in these 2 tasks, is empirically analyzed and discussed.

Related Work

Representation Learning

Statistical machine learning models typically assume inputs as feature vectors. To obviate the need for cumbersome, time- and resource-intensive manual feature engineering, especially from unstructured data such as text, images, or graphs, representation learning aims to learn features or representations from the input directly, often in an unsupervised manner. These real-valued vectorial representations can be used as features directly in machine learning models for various downstream tasks such as prediction or cluster detection.

Representation learning has been successfully used in many domains, such as natural language processing (NLP) [10,11], multimodal learning [12], social network analysis [13], and bioinformatics [14]. In addition, representation learning has been applied within medical informatics to learn patient representations from clinical notes [6], EMR data [1,5], clinical time series [15], and clinical KGs [16,17].

Autoencoder-based neural architectures have been used in most methods to learn patient representations. Miotto et al [5] used stacked denoising autoencoders (SDAE) to learn patient representations from both structured EMR data and topics extracted from clinical notes. Dubois et al [18] obtained note-level representations from clinical notes and combined them to form patient representations. Suresh et al [19] evaluated different autoencoder architectures to find patient phenotypes. Sushil et al [6] evaluated SDAE and Doc2vec representations, both independently and together, to obtain patient representations from clinical notes.

An autoencoder is a simple feedforward neural network that learns to reconstruct its input; it does so by first encoding the input into a dense, low-dimensional vector, also called bottleneck (which is used as the representation after training), and then decoding the bottleneck into the output. The network is trained to make the output as close as possible to the input. Both the encoder and decoder are implemented using neural networks. When there are multiple sources of patient information, such as demographic data, laboratories, and medications, they can be concatenated and provided as input to an autoencoder. A denoising autoencoder uses corrupted versions of inputs and is trained to reconstruct the uncorrupted version. SDAE is a variant based on stacking layers of denoising autoencoders, which are trained locally to denoise corrupted versions of their inputs [20].

In a different approach for combining multiple data sources, patient representations based on CMF were used in the study by Huddar et al [1] to combine multiple EMR matrices with features extracted from clinical notes. These representations were found to be effective in predicting postoperative acute respiratory failure in intensive care unit (ICU) patients.

DCMF Architecture

In multi-view learning, views refer to measurements for the same subjects that differ in source, datatype, or modality. CMF is a general technique for learning shared representations from arbitrary collections of heterogeneous data sources [8].

For a single matrix $X_{m\times n}$ containing m rows and n columns, low-rank factorization aims to obtain latent factors $U_{m\times k}$, and $V_{n\times k}$, such that $X\approx UV^T$, where the latent dimension $k<\min(m,n)$. The latent factors can be viewed as low-dimensional representations of the row and column entities. For example, if X is a matrix containing diagnoses of m patients, where each patient can have $n\geq 1$ diagnoses, the factors provide k-dimensional representations of patients (in U) and diseases (in V). The factors are typically learned by solving the optimization problem:  , where *l* denotes a loss function.

CMF generalizes this idea of single matrix factorization for an arbitrary collection of matrices. The input to the CMF is a collection of matrices, where each matrix, representing a view, has a relationship between 2 entity types along each matrix dimension, and entity types may be involved in multiple views. CMF collectively factorizes the input set of matrices to learn a low-rank latent representation for each entity type from all the views in which the entity type is present. As the CMF models arbitrary collections of matrices, this setting is also referred to as *augmented multi-view learning*.

A model for CMF based on deep learning was developed by Mariappan and Rajan [9], which is briefly described next. Given M matrices (indexed by m) that describe the relationships between E entities (indexed by e), each with dimension d_e, DCMF jointly obtains latent representations of each entity U_e and low-rank factorizations of each matrix  such that $U^e=f_\theta([C]^{(e)})$, where f_θ is an entity-specific nonlinear transformation, obtained through a neural network–based encoder with weights θ and $[C]^{(e)}$ denotes all matrices in the collection that contain a relationship of entity e. The entities corresponding to the rows and columns of the m^{th} matrix are denoted by indices r_m and c_m, respectively.

There are 2 steps in DCMF model construction:

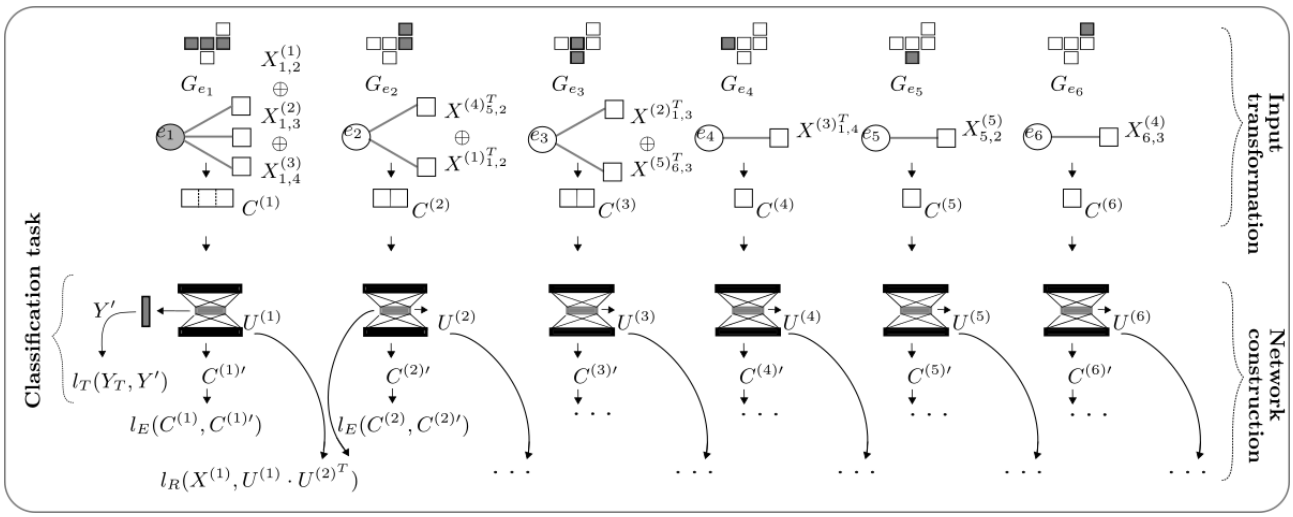
1. Input transformation: For each entity e, we create a new matrix $C^{(e)}$, which we call a concatenated matrix, by concatenating all the matrices containing entity e.
2. Network construction: We then use E (dependent) autoencoders to obtain the latent factors U_e from the concatenated matrices $C^{(e)}$. For each entity e, our network has an autoencoder whose input is $C^{(e)}$, and the decoding is represented by $C^{(e)'}$. The bottleneck or encoding of each autoencoder, after training, forms the latent factor U_e.

The latent factors are learned by training all the autoencoders together by solving the following equation:

where l_E is the reconstruction loss between the autoencoder's input $C^{(e)}$ and the decoding $C^{(e)'}$; l_R is the matrix reconstruction loss, where the reconstructed matrix  of the m^{th} view is obtained by multiplying the associated row and column entity representations  and  . Figure 1 shows a schematic of the model construction steps for an example comprising 5 matrices.

Collective training of all autoencoders induces dependencies between the autoencoder networks, which may result in simultaneous underfitting in some networks and overfitting in other networks. This makes collective learning of all latent representations challenging and, to scale to arbitrary collections of matrices, necessitates automatic hyperparameter selection. We address these optimization challenges through multitask Bayesian optimization (details can be found in the study by Mariappan and Rajan [9]).

Figure 1. Schematic of supervised deep collective matrix factorization architecture for an example input of 5 matrices, 6 entities. Top: input matrices and a graph showing the entities present in each matrix. Bottom: for each entity, matrices containing that entity (as row or column) are concatenated (shaded) and then given as input to the autoencoder. All autoencoders are trained collectively.



Graph Embeddings

Representation learning from graphs aims to learn low-dimensional real-valued features of its nodes, also called graph embeddings, to capture the global structural information and semantic properties in the graph. Many representation learning methods have been proposed for homogeneous graphs, where nodes and edges are both of a single type, for example, DeepWalk [21] and Node2Vec [22]. Many real-world interactions, including those found in clinical KGs, give rise to heterogeneous information networks (HINs) where nodes and edges can be of different types. Representation learning methods for such graphs have also been developed, for example, Metapath2vec [23] and Heterogeneous Graph Neural Network [24]. Cui et al [25] and Cai et al [26] described general surveys, Yang et al [27] described a survey on HIN embeddings, and Wang et al [28] described a survey on representation learning of KGs.

The key underlying idea of many of these techniques is to learn the similarities or correlations between nodes in the input network and approximate them at the latent level in the embeddings. Many network embedding techniques are equivalent to the factorization of a node similarity matrix with suitable definitions of similarities [29].

Knowledge Graphs

Knowledge bases and ontologies systematically organize the wealth of available biomedical knowledge. For instance, the Unified Medical Language System (UMLS) Metathesaurus [30] contains >5 million clinical concepts, identified by controlled unique identifiers (CUIs) and organized into several structured ontologies. Biomedical knowledge is growing at a rapid rate—MEDLINE, the largest index of medical literature, contains >24 million articles with >1.8 million new articles published annually [31]. One cannot possibly assimilate all the knowledge, even in a narrow domain that is growing at such a tremendous pace, let alone find novel connections. To facilitate automated knowledge discovery, hypothesis generation, and predictive modeling from such an enormous and rapidly growing source, automated techniques to extract and organize knowledge into KGs have been developed.

These KGs contain clinical entities as nodes and the relations between entities as edges. As there are different kinds of clinical entities (eg, diseases, drugs, and biomolecules) and different kinds of relations (eg, treats, predisposes, and causes), such KGs are essentially HINs. Examples include Hetionet [32], which comprises 47,031 nodes of 11 types and 2,250,197 relationships of 24 types; KnowLife [33], which contains >500,000 relations for 13 node types, covering genes, organs, diseases, symptoms, and treatments, as well as environmental and lifestyle risk factors; and Semantic Medline Database (SemMedDB) [34], which contains approximately 94 million relations automatically extracted from approximately 27.9 million PubMed abstracts.

In this study, we used the SemMedDB, which, through the use of NLP techniques, automatically creates a KG from biomedical literature. In SemMedDB, clinical concepts are identified in PubMed abstracts through entity recognition algorithms and then mapped to their CUIs. Various heuristics are used to infer

the relations between concepts [35]. SemMedDB infers 30 different kinds of relations that are organized into *subject-predicate-object* triplets (eg, drugA–TREATS–diseaseB), where both the subject and object are clinical concepts, and the predicate is a relation. These triplets form an HIN comprising multiple vertex types (clinical concepts) and multiple edge types (predicates).

Biomedical knowledge, in various forms, including KGs, has been used in clinical predictive models. For instance, the International Classification of Diseases (ICD) hierarchy, which represents relationships across diseases, has been used for diagnosis prediction [36-38]. Recently, domain knowledge–guided recurrent neural network, a recurrent neural network architecture, was proposed [39], where embeddings from a general KG were used internally for initialization. Most of these approaches have specialized architectures for predictive tasks and are not designed to obtain patient representations from heterogeneous collections of data.

Methods

Supervised DCMF

We extended the unsupervised DCMF model to incorporate task-specific supervision. This allowed us to learn entity representations that are influenced by the target variables provided for the predictive task. Furthermore, this creates a predictive model that can seamlessly learn from arbitrary collections of matrices. We assumed that the predictive task, for example, regression or classification, is with respect to one entity only. In the case of clinical tasks, this entity is most often patients. All other data, such as EMRs and KGs, can be used as inputs from which a predictive model for patients can be built. Examples include predicting the length of stay (regression) or the risk of an unforeseen complication (classification).

The DCMF architecture is extended by adding an additional task-specific layer that takes as input the latent representation of the entity for which labels are provided. This layer is provided with labels during training and is trained along with the rest of the network. Let e_p be the specific entity (eg, patients) for which task-specific labels y_T are provided for a task T. Let  be the bottleneck of the autoencoder corresponding to the entity e_p. The network is constructed as described above with the addition of a single network layer that takes  as input and has an activation layer depending on the task and loss function (eg, sigmoid for classification and linear for regression). There is a task-specific loss $l_T(y_T,y')$ associated with this layer that is also task dependent (eg, cross-entropy for classification and mean-squared error for regression), where y' denotes the network's predictions. The supervised latent representations are now learned by solving the following equation:

Collective training of all autoencoders is performed in exactly the same way as in DCMF but with the new loss function as given above. During prediction, new inputs for entity e_p may be given along with all other auxiliary data, and the additional layer's outputs can be used as predictions.

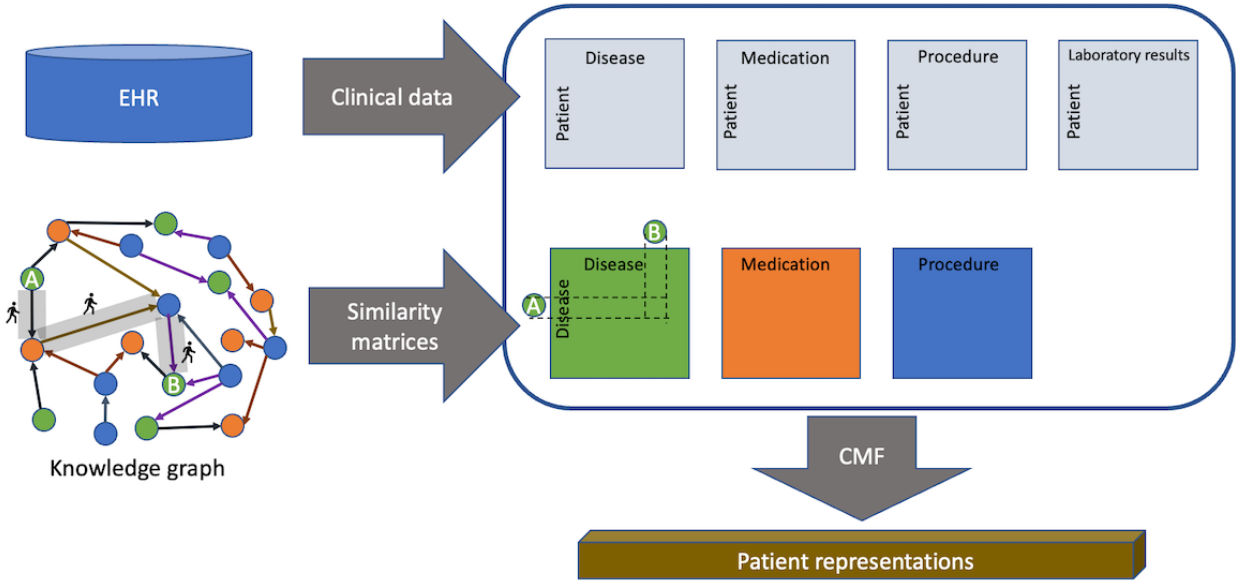
Figure 1 shows a schematic of the model. There are 5 input matrices containing pairwise relations across 6 entities. The graph at the top shows the associations between entities and matrices. One of the entities (shaded) is associated with the labels for a classification task. The network comprises 6 autoencoders, as shown at the bottom, 1 for each entity. The input to the autoencoders is from the concatenated matrix corresponding to each entity (shown in the input transformation part). The bottleneck layer from the first autoencoder is used as input to a network layer that uses the provided labels during training. Note that this illustration shows a specific example of 5 matrices; however, the DCMF model can be used with any collection of input matrices.

Combined Data-Driven and Knowledge-Based Representation Learning Using DCMF

Any graph may be represented by its adjacency matrix. However, factorization of this adjacency matrix may not yield effective representations. We also observed this empirically in our experiments. Another way of using KGs is to first obtain graph embeddings and then use the embeddings within the CMF. We experimented with TransE [40] and found that this did not yield effective representations. To obtain good representations, we used the technique used previously by Liu et al [29]. The key idea was to compute the similarities between the nodes in the graphs and obtain representations by factorizing the similarity matrices.

The global resource allocation (GRA) similarity, between 2 nodes in a graph, was proposed by Liu et al [29] with the aim of having similar embeddings for similar nodes and generalizing previous metrics. We found similarities between diseases, medications, and procedures (separately) from the SemMedDB KG using the GRA similarity. These similarity matrices are provided as input to CMF-based methods that internally factorize all the matrices collectively, as shown in Figure 2.

Figure 2. Schematic of combined data-driven knowledge-based representation learning. Pairwise Global Resource Allocation similarities among clinical entities are computed from the knowledge graph. Patient representations are learnt from these similarity matrices and the input electronic health record data collectively using Collective Matrix Factorization-based methods. CMF: Collective Matrix Factorization; EHR: electronic health record.



We now provide an intuitive explanation of GRA similarity and explain why it is a good measure for clinical KGs; a more technical description can be found in the study by Liu et al [29]. The similarity between 2 nodes *i* and *j* is computed based on the paths that exist between them. Such a global measure can be applied to any 2 nodes in the graph, irrespective of their distance within the graph. In contrast, local measures, such as the number of common neighbors, often yield ineffective embeddings as many node pairs may have the same scores. This is particularly true for dense clinical KGs.

The similarity score depends on (1) the number of paths, (2) the length of the paths, and (3) the node degrees of the intermediate nodes in each path. For each path between i and j, its contribution is equal to the reciprocal of the product of the degrees of the intermediate nodes of the path. Let $p^l(i,j)$ be a path of length l between nodes i and j, and let the intermediate nodes be $i_1,i_2,...i_{\{l-2\}}$. Let k(i) denote the degree of node i, that is, the number of edges incoming to or outgoing from *i*. The contribution of a path $c(p^l)$ is defined as follows:

In this manner, paths that contain high-degree nodes have higher denominators, and their contributions are decreased. This is justified as high-degree nodes connect many different nodes and thus affect many paths. Therefore, paths that do not contain such high-degree nodes should contribute to the higher similarity between the nodes. The final GRA similarity is the sum of the contributions over all paths weighted by a factor that decays exponentially with path length:

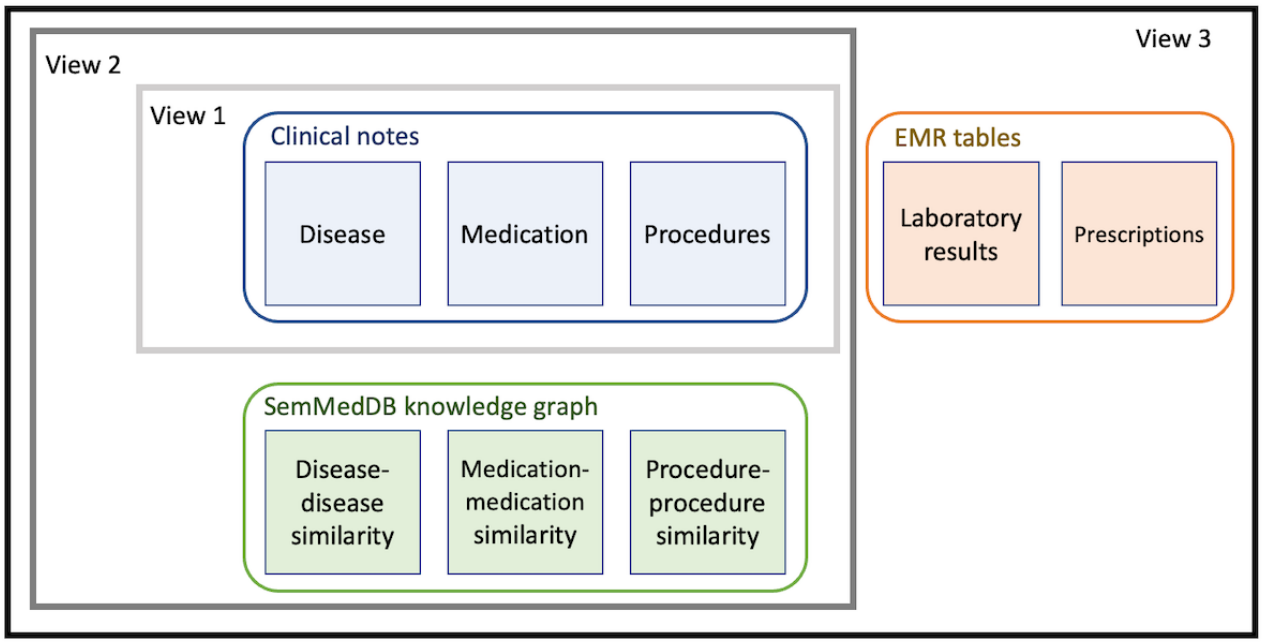
By exponentially decaying the weights, shorter paths are assigned higher weights. Thus, both the number and length of the paths are accounted for in the similarity measure.

Liu et al [29] showed that this technique generalizes and outperforms many previous graph embedding methods. To our knowledge, ours is the first study to use this technique to obtain clinical KG representations and use it within a collective matrix factorization setting to obtain patient representations.

Experiment Settings

Figure 3 shows a schematic of the experimental settings. We considered 3 views: 1, 2, and 3. View 1 comprises data extracted from clinical notes that have been used for patient representation learning in several previous studies. In view 2, data from SemMedDB KGs were extracted as described above and added to the data from view 1. In view 3, structured data from the EMR were also added to obtain patient representations. In the following section, we evaluate the performance of representations learned from these 3 views in 2 clinical decision support tasks.

Figure 3. Views 1, 2, and 3 used to obtain patient representations. EMR: electronic medical record; SemMedDB: Semantic Medline Database.



Data

Overview

We used the Medical Information Mart for Intensive Care (MIMIC) III database [41], which contains clinical data of >40,000 patients admitted to the ICUs in the Beth Israel Deaconess Medical Center in Boston, Massachusetts, between 2001 and 2012. The data were extracted and deidentified in compliance with the Health Insurance Portability and Accountability Act standards [41]. We excluded patients with >1 hospital stay at MIMIC-III. Patients aged <18 years were also excluded. A total of 28,563 patient episodes were used.

Clinical Notes Preprocessing

The NOTEEVENTS table in MIMIC-III contains all clinical notes for patients. It contains a column called IS_ERROR. A value of 1 in this column for a note indicates that a physician has identified the note as an error. Using this value, we first excluded notes that were considered erroneous. The CATEGORY column in the table indicates the type of note recorded. Discharge summaries often contain detailed information about the patient's stay, including diagnoses that are used for billing. As we wanted to predict the diagnosis category automatically from the clinical notes, we excluded all the notes that had been categorized as discharge summaries. The remaining notes were used in our analysis.

The timestamp of a clinical note is obtained from the CHARTTIME and CHARTDATE columns in the NOTEEVENTS table. They recorded the time and date, respectively, at which the notes were charted. Notes are contained in the TEXT column of the NOTEEVENTS table. To efficiently process the notes, they were aggregated over time intervals of 6 hours, starting from the time of ICU admission, and stored as text files. These text files were provided as input to the cTakes software (Apache) [42], which identifies clinical concepts in the input text and provides their CUI values. The software identifies several concept types, such as anatomical site, disease disorder, medication, procedure, and sign–symptoms. We considered only 3 concept types—medication, procedure, and disease–disorder—for our analysis.

For each of the 3 concept types, we constructed a separate matrix, where each row corresponded to a patient episode and the columns corresponded to CUI for the clinical entity. Note that concepts identified from all the notes of a patient episode were considered together to construct the row in the matrix. The disease matrix is binary, indicating the presence or absence of the CUI in the text. Thus, a 1 in the ij-th cell of the matrix indicates the presence of the j-th CUI in a note of the i-th patient episode. The medication and procedure matrices are count matrices, where each cell indicates the number of times the corresponding CUI is mentioned in the text. The total number of CUIs (ie, columns) in the disease, medication, and procedure matrices was 6604. The matrices were transformed to obtain term frequency-inverse document frequency vectors, where each identified CUI was considered a term, and all the considered notes for each patient episode were considered a document.

SemMedDB Preprocessing

SemMedDB contains 30 different kinds of relations that are organized into subject-predicate-object triplets (eg, drugA–TREATS–diseaseB), where both the subject and object are clinical concepts, and the predicate is a relation. The PREDICATION table in SemMedDB contains all the triplets, 1 in each row. The columns SUBJECT_CUI, PREDICATE, and OBJECT_CUI were used to identify the CUI of the subject, predicate, and object, respectively, for each triple. As described earlier, our aim was to obtain a set of triplets to inform us of pairwise relationships across diseases, medications, and procedures for the patient data obtained from MIMIC-III.

As the database is very large, we excluded some relations that were not directly related to clinical concepts in the patient data. These predicates included (1) PART_OF, indicating that a physical unit is a part of a larger unit; (2) LOCATION_OF, indicating the site or region of an entity; and (3) PROCESS_OF, indicating the organism in which a process occurs. In addition, all negations of the predicates in SemMedDB, which begin with NEG, were not considered. More details of these ontological predicates can be found in the study by Kilicoglu et al [34]. The rows containing these predicates were removed from the table. From the remaining rows, only those rows where both the subject and object CUIs were present in the 6604 CUIs used in the patient data were considered; the other rows were excluded.

The final set of triplets was used to construct an undirected graph in the following steps. All clinical concepts present as subjects or objects in the triplets were used as nodes. An edge was added to the graph between nodes u and v if there was a predicate with subject u and object v in the considered triplets. Note that there may be multiple triples between the same subject and object if there are different types of relations. The edges in our graph only indicated the existence of a relation and did not describe the type. Thus, our constructed KG had 6604, 4653, and 3406 nodes of 3 types—disease, medication, and procedure, respectively—and 51,326,066 edges among them. This graph was used to construct GRA similarity matrices, as described earlier for diseases, medications, and procedures.

Structured EMR Data

The prescriptions and laboratory events tables from MIMIC for the selected episodes were used directly. UMLS CUIs for medications were fetched by invoking the representational state transfer application programming interface from RxNorm [43]. The UMLS CUIs for laboratories were obtained using the MRCONSO file from UMLS [30]. Thus, we obtained 1841 and 242 CUIs for medications and laboratories, respectively.

Evaluation

Overview

We evaluated the performance of the models by constructing randomly selected held-out test sets. We split the patient episodes into 90% as training set and 10% as test set. A total of 3 different 90 to 10 splits were randomly generated, and all results shown were averaged over these 3 test sets.

Clinical Decision Support Tasks

Predictive performance was evaluated on 2 clinical decision support tasks.

The first task was that of the primary diagnosis category prediction. When a patient is discharged from the hospital, clinical coders use clinical and demographic data in EMR to assign codes in a standard format, such as ICD, for billing and reimbursement purposes. Several factors such as disease etiology, anatomical site, and severity are used in coding algorithms [44]. This is a time-consuming and error-prone process, and mistakes can lead to claim denials and underpayment for hospitals [45]. As a result, many methods have been developed for automated ICD coding [3,46,47]. An important code, from a billing perspective, that needs to be ascertained is the primary diagnosis (the reason for hospitalization). Following the study by Sushil et al [6], we predicted the category of primary diagnosis, where the categories were grouped into 18 generic categories that corresponded to diagnosis-related groups [48]. We modeled this as a multilabel classification task.

Our second task was that of mortality (risk of death) prediction. At the individual patient level, such models can be used to identify high-risk patients and prioritize their care within the ICU. It can also aid in critical decisions such as interrupting treatments or providing do-not-resuscitate orders [2,49]. MIMIC-III provides 3 different mortality labels: in-hospital, 1-month, and 1-year mortality. We used 1-year mortality, which had the least class imbalance. The label indicates whether a patient died within 1 year of discharge from the hospital. Thus, this was a binary classification task.

The label distributions for both the data sets are shown in Tables 1 and 2.

Table 1. Label distribution for 1-year mortality prediction task.

Label	Meaning	Episodes, n (%)
0	Not expired within 1 year after discharge	25,071 (87.79)
1	Expired within 1 year after discharge	3487 (12.21)

Table 2. Label distribution for diagnosis category prediction task.

Label	Meaning	Episodes, n (%)
0	Infection and parasitic diseases	2067 (7.24)
1	Neoplasms	2202 (7.71)
2	Endocrine, nutritional, and metabolic diseases and immunity disorders	616 (2.16)
3	Diseases of blood and blood-forming organs	96 (0.34)
4	Mental disorders	273 (0.96)
5	Diseases of nervous system and sense organs	487 (1.71)
6	Diseases of the circulatory system	11,249 (39.39)
7	Diseases of the respiratory system	2031 (7.11)
8	Diseases of the digestive system	2614 (9.15)
9	Diseases of the genitourinary system	505 (1.77)
10	Complications of pregnancy, childbirth, and the puerperium	119 (0.42)
11	Diseases of the skin and subcutaneous tissue	75 (0.26)
12	Diseases of the musculoskeletal system and connective tissue	372 (1.3)
13	Congenital anomalies	217 (0.76)
14	Certain conditions originating in the perinatal period	0 (0)
15	Symptoms, signs, and ill-defined conditions	333 (1.17)
16	Injury and poisoning	5210 (18.24)
17	Supplementary factors influencing health status and contact with health services	85 (0.3)
18	Supplementary classification of external causes of injury and poisoning	7 (0.02)

Models Compared

We compared 3 models to obtain patient representations. The first was the SDAE that has been used in several previous studies. It was also found to have good performance in representation learning from clinical notes for our selected tasks [6]. Note that the SDAE cannot be used when KG matrices are used.

The other 2 models are the nonneural versions of CMF and DCMF, which can be used in all 3 views. All 3 models were unsupervised learning methods. The representations learned from these methods can be used to train any off-the-shelf classifier. We evaluated the performance using 2 classifiers: random forest [50] and logistic regression. We also evaluated DCMF in the extended supervised mode, where no additional classifier was required.

The SDAE was trained following the implementation of Vincent et al [20]. A single hidden layer was used with an embedding dimension of 300, with sigmoid encoding activation and linear decoding activation. The network was trained using the RMSprop optimizer with a batch size of 32, 0.4 dropout [51], mean square error loss function, and for 20 epochs. DCMF, both supervised and unsupervised, was trained using a single hidden layer in each entity's autoencoder, with tanh activation functions. The weight decay of 1e-6 was used with a learning rate of 1e-5. The network was trained using the Adam [52]. The R package for CMF [53] was used with default parameters.

Evaluation Metrics

Diagnosis category prediction was a multilabel classification task, and we used the standard metrics of accuracy, macro F1, and weighted F1 scores. The F1 score is the harmonic mean of precision and recall. Macro F1 is the unweighted mean of the F1 score for each label. Weighted F1 determines the mean weighted by the number of true instances for each label.

Mortality prediction is a binary classification task, and we use the F1 score and area under the receiver operating characteristic (AUC) curve as evaluation metrics. The AUC shows the overall classifier performance at different thresholds that trade-off sensitivity for specificity.

Results

Overview

We first present the results of the diagnosis category prediction and then mortality prediction. For each task, we visually present the results in 2 ways: one organized by view and another organized by method. The former allowed us to compare methods within each view, and the latter allowed us to compare views within each method.

Diagnosis Category Prediction

Table 3 shows the results of the diagnosis category prediction. In view 1, predictions using supervised DCMF yielded >30% improvement in macro-F1 scores compared with classifiers with SDAE-based representations. In views 2 and 3, considerable improvement, ranging from 82% to 1955% in macro-F1 scores,

was observed over other methods that separately learned representations and classifiers. In view 1, the accuracy and weighted F1-score of supervised DCMF were comparable with those obtained from classifiers trained on SDAE-based representations. However, with the addition of knowledge matrices in view 3, which can be performed seamlessly, supervised DCMF surpassed their performance.

Table 3. Results of diagnosis category prediction.

Model and view	Accuracy (%)	F1 score-macro (%)	F1 score-weighted (%)
View 1			
SDAE[a] LR[b]	68.25	29.99	64.99
SDAE RF[c]	63.03	22.74	57.79
CMF[d] LR	6.66	0.99	2.40
CMF RF	43.96	9.08	34.57
DCMF[e] LR	62.44	22.59	58.01
DCMF RF	58.44	17.66	52.34
DCMF supervised	66.86[f]	39.22[f]	65.7[f]
View 2			
CMF LR	39.95	3.38	22.87
CMF RF	41.05	4.99	26.83
DCMF LR	63.71	25.34	59.87
DCMF RF	62.48	22.95	58.31
DCMF supervised	67.96[f]	39.58[f]	66.69[f]
View 3			
CMF LR	9.39	2.00	5.21
CMF RF	44.51	10.90	37.44
DCMF LR	60.94	22.56	56.94
DCMF RF	56.17	17.26	49.88
DCMF supervised	70.87[f]	41.10[f]	69.39[f]

[a]SDAE: stacked denoising autoencoder.

[b]LR: logistic regression.

[c]RF: random forest.

[d]CMF: collective matrix factorization.

[e]DCMF: deep collective matrix factorization.

[f]Best score for the corresponding view.

Figure 4 shows the results of the diagnosis category prediction across the 3 views. In view 1, we observed that neural representations from SDAE and DCMF outperformed nonneural representations from CMF. The supervised DCMF outperformed all other methods. The addition of information from KGs in view 2 improved the performance of DCMF, both unsupervised and supervised, in all 3 metrics. The addition of structured EMR data in view 3 further improved the performance.

Figure 4. Diagnosis category prediction across Views. Top row: accuracy; middle row: macro F1 score; bottom row: weighted F1 score. CMF: collective matrix factorization; DCMF: deep collective matrix factorization; LR: logistic regression; RF: random forest; SDAE: stacked denoising autoencoder.

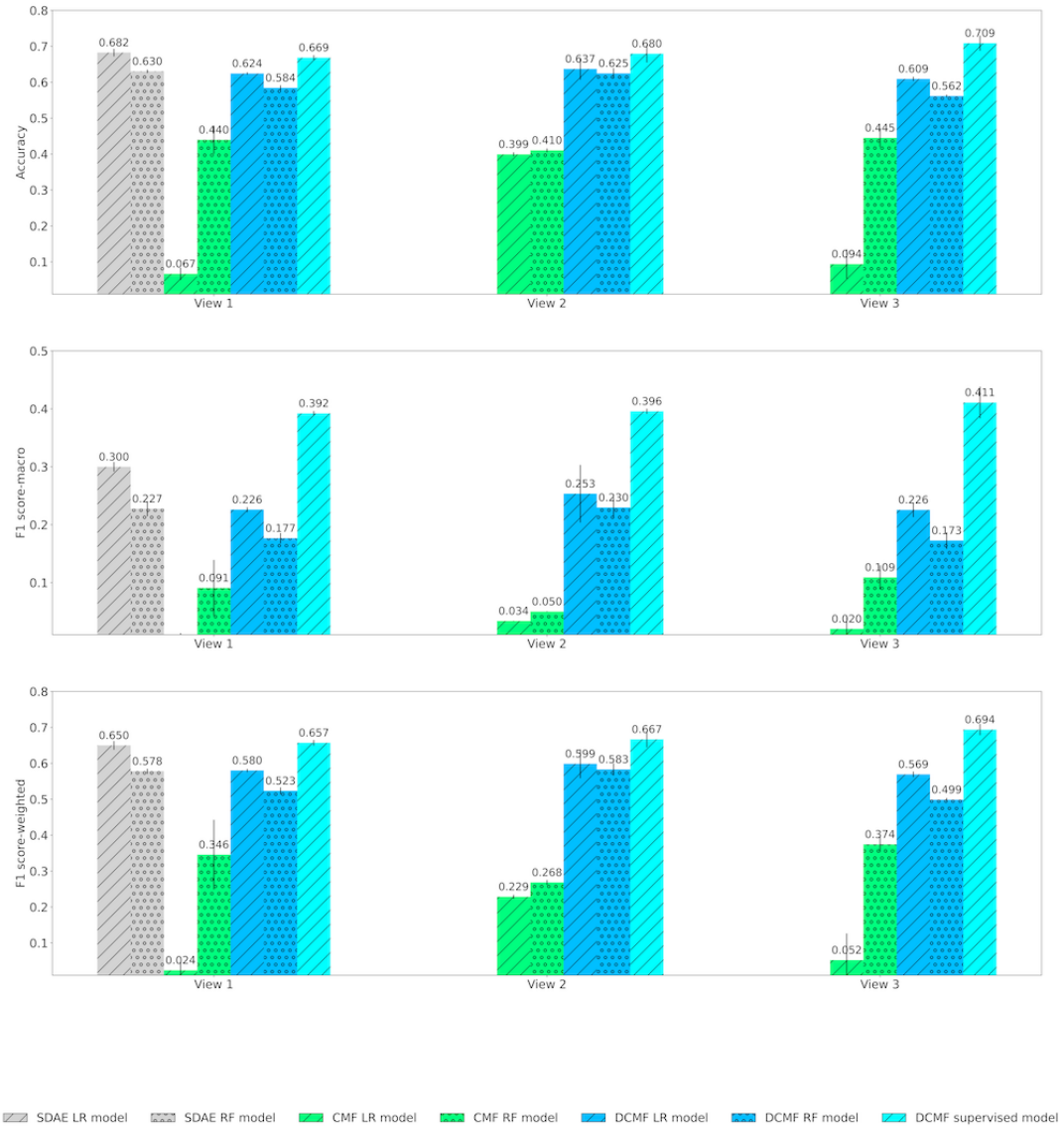
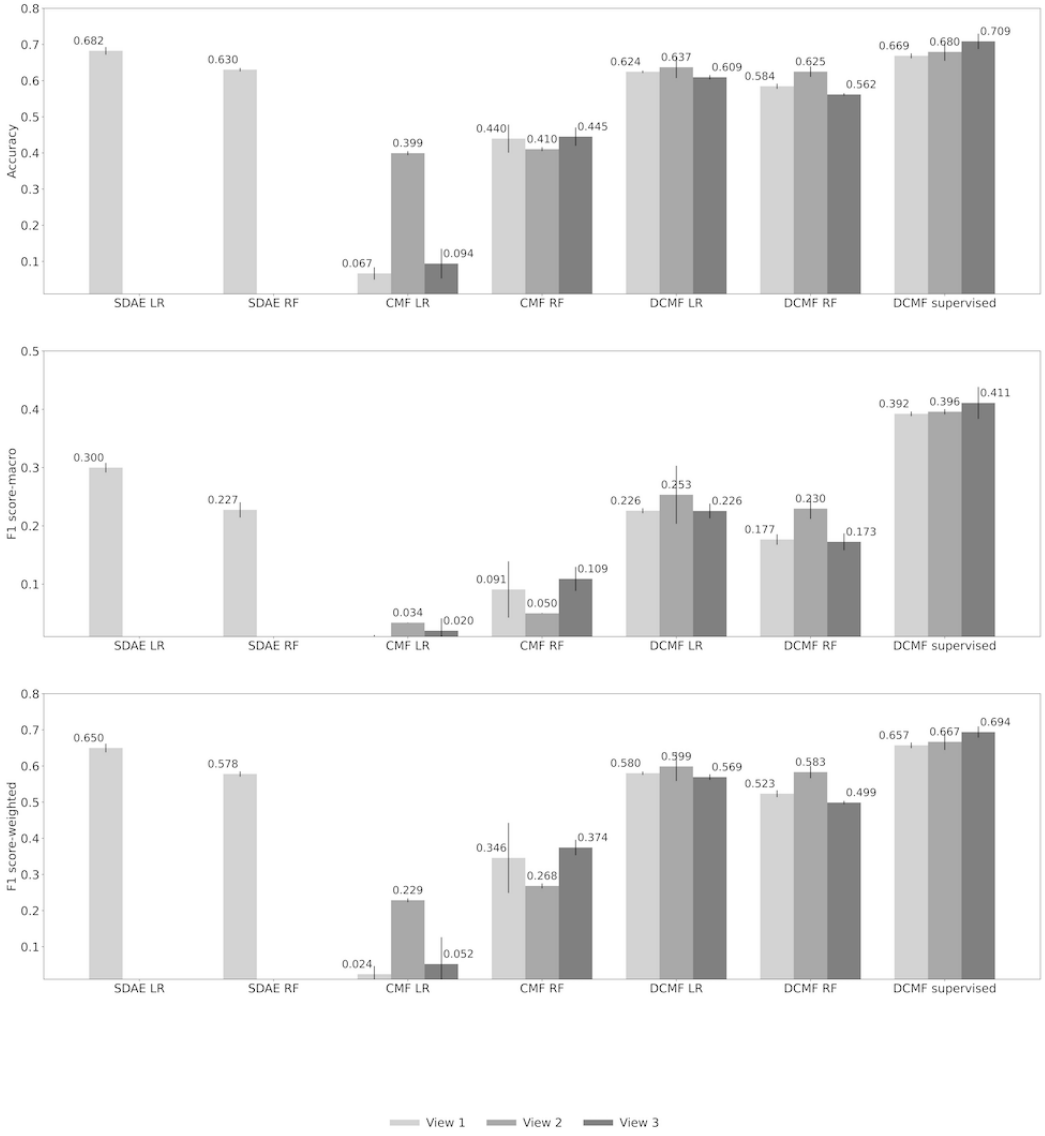


Figure 5 shows the same results of diagnosis category prediction as seen in Figure 4 but is organized based on the method. SDAE representations cannot be used in augmented multi-view settings but outperform CMF-based representations even when the CMF uses more data in views 2 and 3. This is likely because of the better representation learning capability of the neural networks. We also see that the DCMF learned better representations from all 3 views. However, although the addition of KG matrices in view 2 improved performance over view 1, further addition of data in view 3 deteriorated performance. However, with the addition of supervision from the labels, supervised DCMF was able to learn better with increasing performance across the 3 views.

Figure 5. Diagnosis category prediction across Models. Top row: accuracy; middle row: macro F1 score; bottom row: weighted F1 score. CMF: collective matrix factorization; DCMF: deep collective matrix factorization; LR: logistic regression; RF: random forest; SDAE: stacked denoising autoencoder.



Mortality Prediction

Table 4 shows the results of mortality prediction. We observed that supervised DCMF outperformed SDAE-based models by >16% in AUC and >13% in macro-F1 in view 1, where data were obtained from clinical notes. In views 2 and 3, where data from KGs and EMRs were cumulatively added to clinical notes, supervised DCMF outperformed all the baselines by similar margins. These results demonstrate the advantage of end-to-end learning using supervised DCMF over other methods that separately learn representations and classifiers.

Table 4. Results of mortality prediction.

Model and view	AUC[a] (%)	F1 score-macro (%)	F1 score-weighted (%)
View 1			
SDAE[b] LR[c]	52.06	53.15	83.95
SDAE RF[d]	51.55	47.77	82.65
CMF[e] LR	50.37	48.59	81.90
CMF RF	50.21	47.55	82.44
DCMF[f] LR	51.96	50.88	83.41
DCMF RF	50.31	47.48	82.58
DCMF supervised	60.44[g]	60.41[g]	83.99[g]
View 2			
CMF LR	50.00	46.81	82.40
CMF RF	50.04	46.91	82.43
DCMF LR	53.48	53.71	84.04
DCMF RF	51.38	49.76	83.12
DCMF supervised	60.41[g]	60.25[g]	82.97[g]
View 3			
CMF LR	49.99	46.81	82.39
CMF RF	50.00	46.95	82.37
DCMF LR	51.76	50.57	83.28
DCMF RF	50.08	47.00	82.44
DCMF supervised	61.22[g]	62.05[g]	84.43[g]

[a]AUC: area under receiver operating characteristic curve.

[b]SDAE: stacked denoising autoencoders.

[c]LR: logistic regression.

[d]RF: random forest.

[e]CMF LR: collective matrix factorization.

[f]DCMF: deep collective matrix factorization.

[g]Best score for the corresponding view.

Figure 6 shows the AUC and F1 scores obtained by the methods across the 3 views. In view 1, the SDAE representations outperform those from CMF. Results with the logistic regression classifier were marginally better than those from the random forest, with SDAE, CMF, and DCMF representations. In view 1, DCMF representations have performance comparable with that of SDAE. Supervised DCMF outperformed all other methods by a large margin. The addition of KG matrices in view 2 improved the performance of the unsupervised DCMF-based classifier. The addition of structured EMR data in view 3 improved the performance of the supervised DCMF.

Figure 6. Mortality prediction across Views. Top row: area under receiver operating characteristic curve; bottom row: F1 score. AUC: area under receiver operating characteristic curve; CMF: collective matrix factorization; DCMF: deep collective matrix factorization; LR: logistic regression; RF: random forest; SDAE: stacked denoising autoencoder.

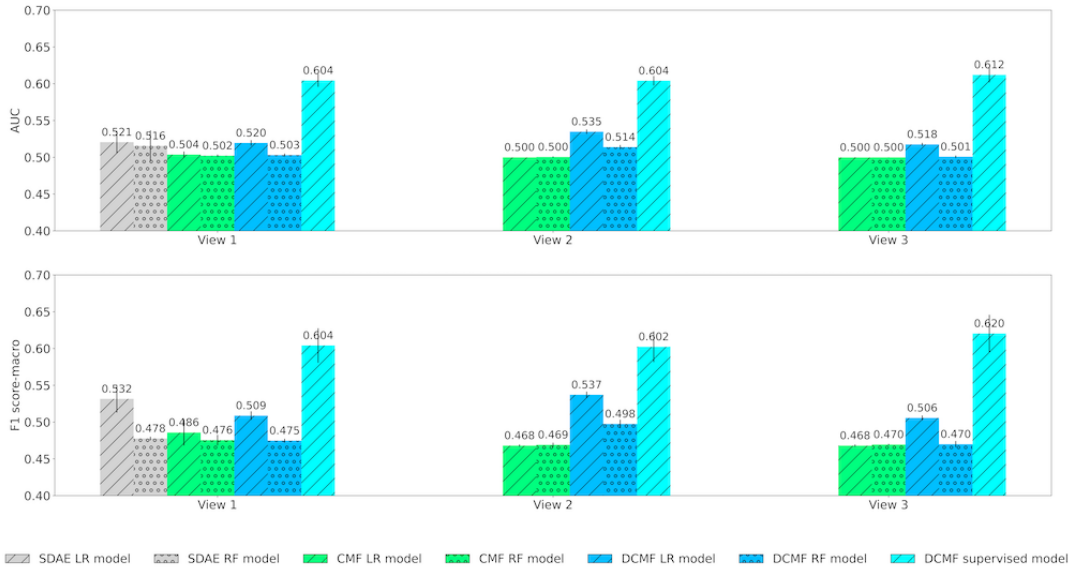
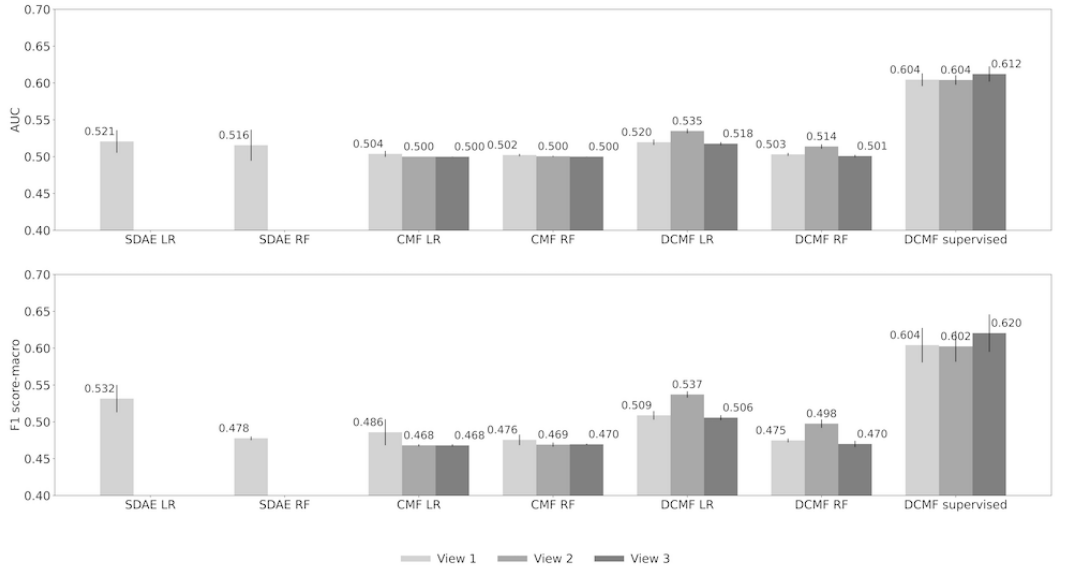


Figure 7 shows the same results from Figure 6, but is organized based on each method. The performances of the unsupervised neural methods SDAE and DCMF are comparable. DCMF can use information from KG matrices to boost its performance. However, the addition of structured EMR data did not increase its performance. However, supervised DCMF is able to use additional data well and achieves the best performance overall with view 3.

Figure 7. Mortality prediction across Models. Top row: area under receiver operating characteristic curve; bottom row: F1 score. AUC: area under receiver operating characteristic curve; CMF: collective matrix factorization; DCMF: deep collective matrix factorization; LR: logistic regression; RF: random forest; SDAE: stacked denoising autoencoder.



Discussion

Principal Findings

Our experiments strongly suggest that end-to-end models that are trained in a supervised manner outperform models comprising 2 stages of unsupervised representation learning and an independently learned classifier. An end-to-end neural model also learns patient representations internally; however, these representations are influenced by task-specific labels used for supervision. How these supervised representations perform on tasks other than what they are trained for, that is, whether they are beneficial in transfer learning, remains to be examined. Thus, for a given clinical decision support task, if labels are

available, our experiments indicate that an end-to-end model should be preferred.

DCMF provides a seamless way of integrating multiple sources of data for obtaining patient representations in both unsupervised and supervised settings. As a versatile learning method, it can be used with inputs from a single source (eg, clinical notes) as well as when inputs are from multiple sources (eg, clinical notes and structured EMR tables). Its performance in these settings is comparable with that of previous autoencoder-based representation learning methods. DCMF can also be used to obtain representations in augmented multi-view settings containing arbitrary collections of matrices, where most previous representation learning methods cannot be used directly. In such settings, its performance is considerably superior to that of the previous nonneural methods for CMF. Thus, it provides a framework for infusing valuable information from auxiliary information sources, such as KG, into patient representations.

Graph embeddings allow us to obtain vectorial representations of nodes in a graph in a way that incorporates the global structural and semantic properties of the graph. Such embeddings can be obtained for KGs as well. The technique for obtaining the embedding can be formulated as a factorization of a similarity matrix where the similarities between nodes are defined based on the number and structural characteristics of the paths between them. With this formulation, the factorization can become part of CMF, which enables us to learn patient representations from multiple clinical data sources as well as KGs. Such patient representations were found to improve downstream predictive performance, especially in supervised settings. Other ways of using KGs within DCMF were not found to be as effective; the 2 alternatives tested were directly using the adjacency matrices of the graphs and first obtaining graph embeddings and then using the embedding matrices within CMF.

Limitations

Our experimental evaluation was conducted on 2 clinical decision support tasks: a binary classification task (mortality prediction) and a multilabel classification task (primary diagnosis category prediction). Furthermore, the evaluation was performed on a subset of data sources (clinical notes, laboratory investigations, and medications) from a single hospital. The trends in performance are expected to remain the same for other tasks (eg, regression tasks) and the addition of other data sources (eg, radiology images) but must be empirically verified.

The KG used is derived automatically from biomedical literature using NLP techniques. Inaccuracies because of NLP algorithms may lead to false positives (erroneous nodes and edges) and false negatives (incompleteness) in KG. Further investigation into the effects of these inaccuracies in the representations is required. Evaluation of KGs derived from other sources can also be performed. It is possible that the results may improve with decreasing inaccuracies in the KG.

Very little hyperparameter tuning was performed for the neural models. The results of all neural models are expected to improve with more tuning. The autoencoders used within the DCMF are simple feedforward networks. Other types of autoencoders, such as SDAE or variational autoencoders, may also be used, which may improve the performance of the DCMF.

Conclusions

In this study, we investigated the use of DCMF to obtain patient representations for 2 clinical decision support tasks. The key advantage of DCMF is its versatility: it can be used to obtain representations from a single view (eg, clinical notes), from multiple views (eg, notes and structured tables in EMR data), and in *augmented* multi-view settings where it can seamlessly integrate information from diverse sources such as EMR data and KGs. Most previous representation learning methods cannot be used with such augmented multi-view data. Furthermore, DCMF can be easily used to learn representations in both supervised and unsupervised settings. In our experiments, we found that DCMF-based representations lead to predictive accuracy that is comparable with or better than previous techniques. Thus, DCMF offers an effective way of integrating heterogeneous data sources and infusing auxiliary knowledge into patient representations.

Acknowledgments

This work was supported by the Singapore Ministry of Education Academic Research Fund (R-253-000-159-114). The principal investigator was VR.

Authors' Contributions

SK implemented supervised deep collective matrix factorization (DCMF) and scripts to use baseline algorithms. AR and AN implemented the global resource allocation (GRA) similarity. SK, AN, and RM conducted the experiments. VR, SK, and RM wrote the manuscript. VR conceived and supervised the project.

Conflicts of Interest

None declared.

References

1. Huddar V, Desiraju BK, Rajan V, Bhattacharya S, Roy S, Reddy CK. Predicting complications in critical care using heterogeneous clinical data. IEEE Access 2016;4:7988-8001. [doi: 10.1109/access.2016.2618775]

2. Ghanvatkar S, Rajan V. Deep recurrent neural networks for mortality prediction in intensive care using clinical time series at multiple resolutions. In: Proceedings of the ICIS Conference 2019. 2019 Presented at: ICIS conference 2019; Dec 15-18, 2019; Munich, Germany URL: https://aisel.aisnet.org/icis2019/data_science/data_science/12/
3. Gartner D, Kolisch R, Neill D, Padman R. Machine learning approaches for early DRG classification and resource allocation. INFORMS J Comput 2015 Nov;27(4):718-734 [FREE Full text] [doi: 10.1287/ijoc.2015.0655]
4. Arnaud E, Elbattah M, Gignon M, Dequen G. Deep learning to predict hospitalization at triage: integration of structured data and unstructured text. In: Proceedings of the IEEE International Conference on Big Data (Big Data). 2020 Presented at: IEEE International Conference on Big Data (Big Data); Dec 10-13, 2020; Atlanta, GA, USA. [doi: 10.1109/bigdata50022.2020.9378073]
5. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 May 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
6. Sushil M, Šuster S, Luyckx K, Daelemans W. Patient representation learning and interpretable evaluation using clinical notes. J Biomed Inform 2018 Aug;84:103-113 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.016] [Medline: 29966746]
7. Almeida J, Silva J, Matos SJ, Oliveira JL. A two-stage workflow to extract and harmonize drug mentions from clinical notes into observational databases. J Biomed Inform 2021 Aug;120:103849. [doi: 10.1016/j.jbi.2021.103849] [Medline: 34214696]
8. Singh A, Gordon G. Relational learning via collective matrix factorizationProceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2008 Presented at: 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 24 - 27, 2008; Las Vegas Nevada USA. [doi: 10.1145/1401890.1401969]
9. Mariappan R, Rajan V. Deep collective matrix factorization for augmented multi-view learning. Mach Learn 2019 May 17;108(8-9):1395-1420 [FREE Full text] [doi: 10.1007/s10994-019-05801-6]
10. Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning. 2014 Presented at: 31st International Conference on Machine Learning; June 22-24, 2014; Beijing, China p. 1188-1196 URL: https://proceedings.mlr.press/v32/le14.html
11. Zhao J, Zhan Z, Yang Q, Zhang Y, Hu C, Li Z, et al. Adaptive learning of local semantic and global structure representations for text classification. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018 Presented at: 27th International Conference on Computational Linguistics; Aug 20-26, 2018; Santa Fe, New Mexico, USA URL: https://aclanthology.org/C18-1173/
12. Tian H, Tao Y, Pouyanfar S, Chen S, Shyu M. Multimodal deep representation learning for video classification. World Wide Web 2019 May;22(3):1325-1341 [FREE Full text] [doi: 10.1007/s11280-018-0548-3]
13. Wang W, Yin H, Du X, Hua W, Li Y, Nguyen Q. Online user representation learning across heterogeneous social networks. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019 Presented at: 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval; Jul 21 - 25, 2019; Paris France. [doi: 10.1145/3331184.3331258]
14. Liany H, Jeyasekharan A, Rajan V. Predicting synthetic lethal interactions using heterogeneous data sources. Bioinformatics 2020 Apr 01;36(7):2209-2216. [doi: 10.1093/bioinformatics/btz893] [Medline: 31782759]
15. Ruan T, Lei L, Zhou Y, Zhai J, Zhang L, He P, et al. Representation learning for clinical time series prediction tasks in electronic health records. BMC Med Inform Decis Mak 2019 Dec 17;19(Suppl 8):259 [FREE Full text] [doi: 10.1186/s12911-019-0985-7] [Medline: 31842854]
16. Cohen T, Widdows D. Embedding of semantic predications. J Biomed Inform 2017 Apr;68:150-166 [FREE Full text] [doi: 10.1016/j.jbi.2017.03.003] [Medline: 28284761]
17. Dasgupta S, Jayagopal A, Jun Hong AL, Mariappan R, Rajan V. Adverse drug event prediction using noisy literature-derived knowledge graphs: algorithm development and validation. JMIR Med Inform 2021 Oct 25;9(10):e32730 [FREE Full text] [doi: 10.2196/32730] [Medline: 34694230]
18. Dubois S, Nathanael R, David K, Nigam S, Kenneth J. Effective representations from clinical notes. arXiv. 2018 Aug 16. URL: https://arxiv.org/abs/1705.07025 [accessed 2022-01-03]
19. Suresh H, Peter S, Marzyeh G. The use of autoencoders for discovering patient phenotypes. arXiv. 2017 Mar 20. URL: https://arxiv.org/abs/1703.07004 [accessed 2022-01-02]
20. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 2010 Jan 3;11:3371-3408. [doi: 10.5555/1756006.1953039]
21. Perozzi B, Al-Rfou R, Skiena S. DeepWalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014 Presented at: 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 24 - 27, 2014; New York USA. [doi: 10.1145/2623330.2623732]
22. Grover A, Leskovec J. node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: 22nd ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining; Aug 13 - 17, 2016; San Francisco California USA. [doi: 10.1145/2939672.2939754]
23. Dong Y, Chawla N, Swami A. metapath2vec: scalable representation learning for heterogeneous networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Presented at: 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2017; Halifax NS Canada. [doi: 10.1145/3097983.3098036]
24. Zhang C, Song D, Huang C, Swami A, Chawla N. Heterogeneous graph neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019 Presented at: 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; Aug 4-8, 2019; Anchorage AK USA. [doi: 10.1145/3292500.3330961]
25. Cui P, Wang X, Pei J, Zhu W. A survey on network embedding. IEEE Trans Knowl Data Eng 2019 May 1;31(5):833-852. [doi: 10.1109/tkde.2018.2849727]
26. Cai H, Zheng VW, Chang KC. A comprehensive survey of graph embedding: problems, techniques, and applications. IEEE Trans Knowl Data Eng 2018 Sep 1;30(9):1616-1637. [doi: 10.1109/tkde.2018.2807452]
27. Yang C, Xiao Y, Zhang Y, Sun Y, Han J. Heterogeneous network representation learning: a unified framework with survey and benchmark. IEEE Trans Knowl Data Eng 2020 Dec:1. [doi: 10.1109/tkde.2020.3045924]
28. Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: a survey of approaches and applications. IEEE Trans Knowl Data Eng 2017 Dec 1;29(12):2724-2743. [doi: 10.1109/tkde.2017.2754499]
29. Liu X, Murata T, Kim K, Kotarasu C, Zhuang C. A general view for network embedding as matrix factorization. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. 2019 Presented at: Twelfth ACM International Conference on Web Search and Data Mining; Feb 11-15, 2019; Melbourne VIC Australia. [doi: 10.1145/3289600.3291029]
30. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 01;32(Database issue):267-270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
31. Fact sheet MEDLINE®. U.S. National Library of Medicine. URL: httpshttp://wayback.archive-it.org/org-350/20180312141554/https://www.nlm.nih.gov/pubs/factsheets/medline.html://www.nlm.nih.gov/pubs/factsheets/medline.html [accessed 2021-12-21]
32. Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. Elife 2017 Sep 22;6:e26726 [FREE Full text] [doi: 10.7554/eLife.26726] [Medline: 28936969]
33. Ernst P, Siu A, Weikum G. KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences. BMC Bioinformatics 2015 May 14;16:157 [FREE Full text] [doi: 10.1186/s12859-015-0549-5] [Medline: 25971816]
34. Kilicoglu H, Shin D, Fiszman M, Rosemblat G, Rindflesch TC. SemMedDB: a PubMed-scale repository of biomedical semantic predications. Bioinformatics 2012 Dec 01;28(23):3158-3160. [doi: 10.1093/bioinformatics/bts591] [Medline: 23044550]
35. Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform 2003 Dec;36(6):462-477 [FREE Full text] [doi: 10.1016/j.jbi.2003.11.003] [Medline: 14759819]
36. Che Z, Kale D, Li W, Bahadori M, Liu Y. Deep computational phenotyping. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Presented at: 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 10-13, 2015; Sydney NSW Australia. [doi: 10.1145/2783258.2783365]
37. Choi E, Bahadori M, Song L, Stewart W, Sun J. GRAM: Graph-based attention model for healthcare representation learning. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Presented at: 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13 - 17, 2017; Halifax NS Canada. [doi: 10.1145/3097983.3098126]
38. Ma F, You Q, Xiao H, Chitta R, Zhou J, Gao J. KAME: Knowledge-based attention model for diagnosis prediction in healthcare. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018 Presented at: 27th ACM International Conference on Information and Knowledge Management; Oct 22 - 26, 2018; Torino Italy. [doi: 10.1145/3269206.3271701]
39. Yin C, Zhao R, Qian B, Lv X, Zhang P. Domain knowledge guided deep learning with electronic health records. In: Proceedings of the IEEE International Conference on Data Mining (ICDM). 2019 Presented at: IEEE International Conference on Data Mining (ICDM); Nov 8-11, 2019; Beijing, China. [doi: 10.1109/icdm.2019.00084]
40. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Proceedings of the 26th International Conference on Neural Information Processing Systems. 2013 Presented at: 26th International Conference on Neural Information Processing Systems; Dec 5 - 10, 2013; Lake Tahoe Nevada URL: https://hal.archives-ouvertes.fr/hal-00920777/document

41. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
42. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
43. RxNav: browser and application programming interfaces for RxNorm. The National Center for Biomedical Ontology. URL: https://ncbo.bioontology.org/RxNav [accessed 2021-12-21]
44. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005 Nov;43(11):1130-1139. [doi: 10.1097/01.mlr.0000182534.19832.83] [Medline: 16224307]
45. Adams DL, Norman H, Burroughs VJ. Addressing medical coding and billing part II: a strategy for achieving compliance. A risk management approach for reducing coding and billing errors. J Natl Med Assoc 2002 Jun;94(6):430-447. [Medline: 12078924]
46. Xu K, Lam M, Pang J, Gao X, Band C, Mathur P, et al. Multimodal machine learning for automated ICD coding. In: Proceedings of the 4th Machine Learning for Healthcare Conference. 2019 Presented at: 4th Machine Learning for Healthcare Conference; Aug 9-10, 2019; Ann Arbor, Michigan URL: https://proceedings.mlr.press/v106/xu19a.html
47. Li F, Yu H. ICD coding from clinical text using multi-filter residual convolutional neural network. Proc Conf AAAI Artif Intell 2020 Feb;34(5):8180-8187 [FREE Full text] [doi: 10.1609/aaai.v34i05.6331] [Medline: 34322282]
48. Roth A, Dierdonck V. Hospital resource planning: concepts, feasibility, and framework. Prod Oper Manag 2009 Mar;4(1):2-29 [FREE Full text] [doi: 10.1111/j.1937-5956.1995.tb00038.x]
49. Bhattacharya S, Rajan V, Shrivastava H. ICU mortality prediction: a classification algorithm for imbalanced datasets. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2017 Presented at: AAAI Conference on Artificial Intelligence; February 4-9, 2017; San Francisco, California USA URL: https://ojs.aaai.org/index.php/AAAI/article/view/10721
50. Breiman L. Random forests. Machine Learning 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
51. Srivastava N, Hinton G, Krizhevsky A, Sutskever IR, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(56):1929-1958 [FREE Full text]
52. Kingma D, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations (ICLR). 2015 Presented at: International Conference on Learning Representations (ICLR); May 7-9, 2015; San Diego URL: https://arxiv.org/abs/1412.6980
53. Klami A, Vare L, Held F. CMF: Collective Matrix Factorization. Cran.R-Project. 2020. URL: https://cran.r-project.org/web/packages/CMF/ [accessed 2021-12-21]

Abbreviations

AUC: area under the receiver operating characteristic curve
CMF: collective matrix factorization
CUI: controlled unique identifier
DCMF: deep collective matrix factorization
EMR: electronic medical record
GRA: global resource allocation
HIN: heterogeneous information network
ICD: International Classification of Diseases
ICU: intensive care unit
KG: knowledge graph
MIMIC: Medical Information Mart for Intensive Care
NLP: natural language processing
SDAE: stacked denoising autoencoder
SemMedDB: Semantic Medline Database
UMLS: Unified Medical Language System

Edited by C Lovis; submitted 15.03.21; peer-reviewed by M Elbattah; comments to author 03.10.21; revised version received 07.11.21; accepted 14.11.21; published 20.01.22.

Please cite as:
Kumar S, Nanelia A, Mariappan R, Rajagopal A, Rajan V
Patient Representation Learning From Heterogeneous Data Sources and Knowledge Graphs Using Deep Collective Matrix Factorization: Evaluation Study
JMIR Med Inform 2022;10(1):e28842
URL: https://medinform.jmir.org/2022/1/e28842
doi:10.2196/28842
PMID:35049514

Original Paper

Evaluation of the Need for Intensive Care in Children With Pneumonia: Machine Learning Approach

Yun-Chung Liu[1,2*], MSc; Hao-Yuan Cheng[1,3*], MD, MSc; Tu-Hsuan Chang[4], MD; Te-Wei Ho[5], PhD; Ting-Chi Liu[6,7], MSc; Ting-Yu Yen[1], MD; Chia-Ching Chou[6*], PhD; Luan-Yin Chang[1*], MD, PhD; Feipei Lai[2,8,9], PhD

[1]Department of Pediatrics, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei City, Taiwan

[2]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei City, Taiwan

[3]Taiwan Centers for Disease Control, Taipei City, Taiwan

[4]Department of Pediatrics, Chi Mei Medical Center, Tainan City, Taiwan

[5]Department of Surgery, College of Medicine, National Taiwan University, Taipei City, Taiwan

[6]Institute of Applied Mechanics, National Taiwan University, Taipei City, Taiwan

[7]Department of Civil Engineering, National Taiwan University, Taipei City, Taiwan

[8]Department of Computer Science and Information Engineering, National Taiwan University, Taipei City, Taiwan

[9]Department of Electrical Engineering, National Taiwan University, Taipei City, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chia-Ching Chou, PhD
Institute of Applied Mechanics
National Taiwan University
No.1, Sec. 4, Roosevelt Rd., Da'an Dist.
Taipei City, 10617
Taiwan
Phone: 886 233665639
Email: ccchou@iam.ntu.edu.tw

Abstract

Background: Timely decision-making regarding intensive care unit (ICU) admission for children with pneumonia is crucial for a better prognosis. Despite attempts to establish a guideline or triage system for evaluating ICU care needs, no clinically applicable paradigm is available.

Objective: The aim of this study was to develop machine learning (ML) algorithms to predict ICU care needs for pediatric pneumonia patients within 24 hours of admission, evaluate their performance, and identify clinical indices for making decisions for pediatric pneumonia patients.

Methods: Pneumonia patients admitted to National Taiwan University Hospital from January 2010 to December 2019 aged under 18 years were enrolled. Their underlying diseases, clinical manifestations, and laboratory data at admission were collected. The outcome of interest was ICU transfer within 24 hours of hospitalization. We compared clinically relevant features between early ICU transfer patients and patients without ICU care. ML algorithms were developed to predict ICU admission. The performance of the algorithms was evaluated using sensitivity, specificity, area under the receiver operating characteristic curve (AUC), and average precision. The relative feature importance of the best-performing algorithm was compared with physician-rated feature importance for explainability.

Results: A total of 8464 pediatric hospitalizations due to pneumonia were recorded, and 1166 (1166/8464, 13.8%) hospitalized patients were transferred to the ICU within 24 hours. Early ICU transfer patients were younger (P<.001), had higher rates of underlying diseases (eg, cardiovascular, neuropsychological, and congenital anomaly/genetic disorders; P<.001), had abnormal laboratory data, had higher pulse rates (P<.001), had higher breath rates (P<.001), had lower oxygen saturation (P<.001), and had lower peak body temperature (P<.001) at admission than patients without ICU transfer. The random forest (RF) algorithm achieved the best performance (sensitivity 0.94, 95% CI 0.92-0.95; specificity 0.94, 95% CI 0.92-0.95; AUC 0.99, 95% CI 0.98-0.99; and average precision 0.93, 95% CI 0.90-0.96). The lowest systolic blood pressure and presence of cardiovascular and neuropsychological diseases ranked in the top 10 in both RF relative feature importance and clinician judgment.

Conclusions: The ML approach could provide a clinically applicable triage algorithm and identify important clinical indices, such as age, underlying diseases, abnormal vital signs, and laboratory data for evaluating the need for intensive care in children with pneumonia.

(JMIR Med Inform 2022;10(1):e28934) doi:10.2196/28934

KEYWORDS

child pneumonia; intensive care; machine learning; decision making; clinical index

Introduction

Despite recent advances in vaccine development, pneumonia remains a major cause of hospitalization and mortality in children in Taiwan and worldwide [1,2]. New pathogens, such as the recent coronavirus causing COVID-19, continue to cause outbreaks of pneumonia and other severe respiratory infections [3,4]. For hospitalized patients with critical conditions, the timely decision to admit them to the intensive care unit (ICU) is crucial for better prognosis and overall medical care quality [5]. The decision is usually made by doctors based on clinical criteria (eg, chief complaint, symptoms/signs, vital signs) and laboratory criteria (eg, microbiology tests, complete blood count, biochemical examinations). However, no well-structured nor quantitative approach exists.

The community-acquired pneumonia management guidelines from the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America [6] recommend that pediatric patients who need ventilation, have low blood pressure, or have low oxygen saturation be admitted to the ICU for pneumonia. Other risk factors, including white blood cell count and hemoglobin, have been associated with exacerbation among pneumonia patients during hospitalization [7]. Some studies have tried to develop clinical scoring systems to standardize prognosis and disease exacerbation evaluations. For example, a modified version of the Sequential Organ Failure Assessment score for children used vital signs (blood pressure, oxygen saturation), laboratory data (creatinine, platelet count), and medications to evaluate the risk of in-hospital mortality [8]. Other scoring systems, such as the Pediatric Early Warning Score (PEWS) and Pediatric Advanced Warning Score, have been proposed to assist the evaluation of deterioration of pediatric inpatients [9-11]. Gold et al [12] used a modified version of PEWS calculated at admission to predict ICU admission and reported an area under the receiver operating characteristic curve (AUC) of 0.86. Nevertheless, the varying sensitivity, specificity, and degrees of human effort limited their clinical application.

In the era of health data science, using large amounts of patient data to develop algorithms to solve clinical problems has become an important approach [13-18]. For example, Makino et al [19] applied a logistic regression model to predict aggravation of diabetic kidney disease 180 days after discharge using patient demographic data, lab tests, diagnosis codes, and medical history. Their model reached an AUC of 0.74 [19]. Studies conducted in the emergency service setting showed promising results in triaging patients with asthma and chronic obstructive pulmonary disease [20]. In the critical care setting, Zhang et al [16] developed an ensemble model for the prediction of agitation in invasive mechanical ventilation patients under light sedation; an automated electronic health records model to identify patients at high risk of acute respiratory failure or death was validated retrospectively and prospectively and was determined to be feasible for real-time risk identification [17]. Artificial intelligence technology is assisting us with interpreting complex data from critical patients such as patients with acute respiratory distress syndrome (ARDS) and enables us to further improve the management of critically ill patients with individual treatment plans [18]. In these studies, machine learning (ML) algorithms were usually implemented because of the strength of incorporating large data sets and exploring the hidden relationships among features [13,14]. The most common type of clinical task (eg, determining whether the patient has a specific diagnosis, the clinical severity, and the prognosis, such as survival after a specific period) was classification. Decision tree–based models usually yield the most promising results in these clinical scenarios because of their strength in classification tasks [14,20,21].

A computer-aided prognosis prediction framework has also been applied to evaluate deterioration of pediatric inpatients. Zhai et al [22] used electronic health records in a single medical center to predict the need for pediatric intensive care within the first 24 hours of admission. Their logistic regression model reached an AUC of 0.91. Mayampurath et al [23] used 6 common vital signs (eg, temperature, pulse, blood pressure) to predict an ICU transfer event up to 36 hours in advance, reaching AUCs of 0.7-0.8. Rubin et al [24] applied a boosted trees model to electronic health records to predict pediatric ICU transfer at most 2 hours to 8 hours in advance with an AUC of 0.85. These deterioration evaluation models showed promising results with general pediatric patients.

Most ML studies for pneumonia patients have focused on using clinical imaging data for diagnosis or mortality [25-27]. Few studies have explored the possibility of developing an ML-based prediction framework to evaluate the need for intensive care among pediatric pneumonia patients and to yield clinically applicable performance. Therefore, we aimed to use clinical data from children with pneumonia to develop ML algorithms to predict the need for ICU transfer within 24 hours of admission, which could support physician decision-making.

Methods

Data Source

We enrolled pneumonia patients aged under 18 years admitted to the National Taiwan University Hospital from January 2010 to December 2019. The clinical data for enrolled patients were retrieved from the National Taiwan University

Hospital-integrated Medical Database, and all data were de-identified before being analyzed. The institutional review board of the National Taiwan University Hospital approved this study and the use of de-identified electronic health records (201912131RINB).

The diagnosis of pneumonia was determined from the hospital records if both of the following criteria were met: (1) clinical manifestation of respiratory tract infection at admission, including symptoms (eg, dyspnea, rhinorrhea, cough, sputum), abnormal breath sounds (eg, rales, crackles, rhonchi), or a preliminary diagnosis recorded within 24 hours of admission (see Table S1 in Multimedia Appendix 1), and (2) the International Classification of Disease, ninth revision (ICD-9) and tenth revision (ICD-10) diagnostic codes related to pneumonia at discharge (see Table S2 in Multimedia Appendix 1).

Collection of Clinically Relevant Features

Data including demographics, underlying diseases, vital signs, pathogens, and laboratory data, which were available within 24 hours of hospitalization and prior to ICU transfer, were collected and included in the statistical analysis, model training, and performance evaluation, as seen in Table S3 in Multimedia Appendix 1. Underlying diseases were identified using ICD-9 and ICD-10 codes. The aforementioned clinically relevant features associated with pneumonia prognosis were also selected and ranked by 3 pediatricians specializing in pediatric infectious diseases, with 5, 10, and over 20 years of experience. If missing rates of cohort data were greater than 30%, features were excluded.

Outcome of Interest

The outcome of interest was ICU admission within 24 hours of hospitalization, including those directly admitted to the ICU from emergency departments or death within 24 hours of hospitalization. Therefore, patients transferred to the ICU after 24 hours of admission were excluded. Readmissions due to pneumonia within 14 days or due to other conditions within 3 days were also excluded because they might be related to previous admission. The cohort was thus categorized into 2 groups: early ICU transfer (ie, patients transferred to the ICU or who died within 24 hours of admission) and no ICU admission (patients who were not admitted to the ICU through discharge).

Statistical Analysis

In addition to descriptive analyses, we used chi-square tests for categorical variables to compare differences between the early ICU transfer group and the no ICU admission group. For numerical variables, the Shapiro-Wilk test was used to test normality, the Mann-Whitney *U* test was used for between-group comparisons if the data were not normally distributed, and the *t* test was used if data were normally distributed. The Benjamini-Hochberg procedure was applied to adjust for multiple comparisons. Adjusted *P* values <.05 were considered significant.

Model Training and Performance Evaluation

Based on previous research, we developed a logistic regression model as a baseline reference. Then, we trained random forest (RF) and eXtreme Gradient Boosting (XGB) models because of their promising performance on clinical classification tasks [14,16,17,20,28-31]. For model training, the data set was separated into development and validation sets at a 4:1 ratio via random selection. The ML models were trained using the development set with 5-fold cross-validation. The performance was then evaluated using the independent validation set. The accuracy, sensitivity (recall), specificity, positive predictive value (precision), negative predictive value, AUC, and average precision were calculated to compare different algorithms and thresholds.

We chose 3 points to operationalize the best performing model: the points with the highest Youden index [32], high specificity (0.99), and high sensitivity (0.99), which could be applied in different clinical scenarios. The CI was estimated using bootstrapping methods with 1000 samples.

Comparison of Feature Importance Between the ML Model and Physicians

With the best-performing model selected using the aforementioned performance evaluation, we further generated the relative feature importance list using Tree Explainer based on Shapley Additive Explanations (SHAP) values [21]. The relative feature importance was also ranked by 3 physicians using a 5-point scale, and the list was generated by sorting clinical features according to the average of importance scores assessed by the physicians. Then, the relative feature importance list from the ML model was compared with the relative importance ranked by the physicians.

Software

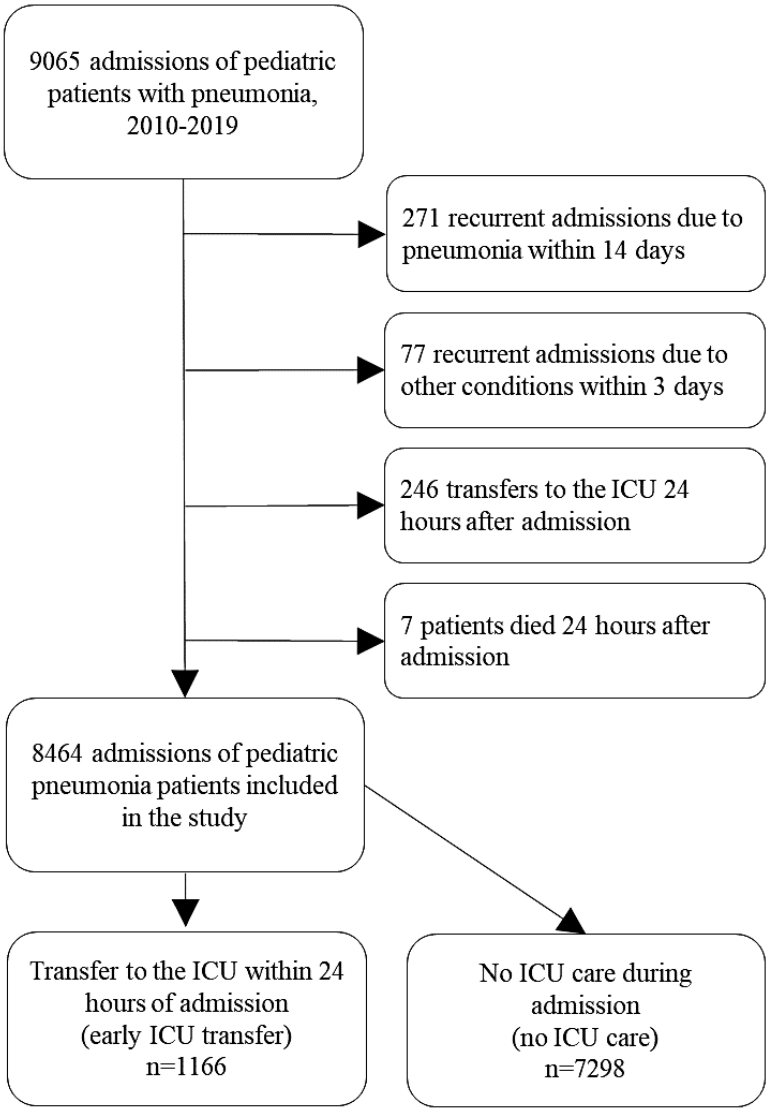
All data were managed using the NumPy (version 1.16.5) and Pandas (version 0.25.1) libraries of the Python programming language version 3.7.4 (Python Software Foundation, Fredericksburg, VA). Statistical analyses were conducted using the SciPy package version 1.3.1 [33]. To train the algorithm, we used Scikit-learn (The Scikit-learn Contributors, version 0.21.3) [34] for logistic regressions and the RF model. The XGBoost package (Version 0.90) was used for the XGB algorithm [35]. The performance evaluation was conducted using the Scikit-learn package. The Tree Explainer was built based on SHAP values [21].

Results

Cohort Description and Between-Group Comparison

A total of 6947 patients from 9065 hospitalizations due to pneumonia were included in the study based on their discharge diagnosis code and status at admission. The text mining algorithm correctly labeled 99.8% of admissions with clinical manifestations of a tentative diagnosis using admission notes as examined by the authors using 1000 randomly sampled admissions. Since 601 admissions were excluded based on the aforementioned exclusion criteria, it resulted in a final cohort of 8464 admissions (Figure 1).

Figure 1. Flowchart of patient enrollment. ICU: intensive care unit.



The male-to-female ratio was 1.16:1. The median age was 3.1 (IQR 1.7-5.1) years. Among the 8464 admissions included, 1166 admissions (13.8%) were transferred to the ICU or died in the hospital within 24 hours of admission, and they were classified as the early ICU transfer group. The most common underlying disease in the early ICU transfer group was cardiovascular disease (459/1166, 39.4%), followed by neuropsychological disease (416/1166, 35.7%) and congenital anomaly/genetic disorder (310/1166, 26.6%). Common reasons for ICU admission included respiratory failure (566/1166, 48.5%, among which 19.3% [109/566] met the criteria of ARDS), sepsis (392/1166, 33.6%), and chest tube insertion (102/1166, 8.7%). There were 1003 (1003/8464, 11.9%) admissions with a positive microbiological test (as listed in Table S3 in Multimedia Appendix 1) result within 24 hours of admission and prior to ICU transfer. The most commonly identified pathogen at admission was influenza virus type A (14/1166 admissions, 1.2%), followed by influenza virus type B (9/1166 admissions, 0.8%) and *Streptococcus pneumoniae* (5/1166 admissions, 0.4%). Younger age, higher rate of underlying diseases, higher pulse rate, higher breath rate, lower oxygen saturation, lower peak body temperature, and abnormal laboratory data were significantly associated with early ICU transfer (Table 1 and a complete list in Table S4 in Multimedia Appendix 1). However, patients with positive results for influenza A, influenza B, and *S. pneumoniae* at admission were less likely to be transferred to the ICU within 24 hours (P=.02, P<.001, and P<.001, respectively).

Table 1. Selective results of clinical feature indices based on early intensive care unit (ICU) transfer.

Features	Early ICU transfer (n= 1166)	No ICU admission (n=7298)	*P* value[a]
Demographic characteristics			
Male, n (%)	623 (53.4)	3916 (53.7)	.89
Age (years), median (IQR)	2.1 (0.5-5.3)	3.2 (1.8-5.0)	<.001
Underlying disease[b]			
Cardiovascular diseases, n (%)	459 (39.4)	599 (8.2)	<.001
Neuropsychological diseases, n (%)	416 (35.7)	836 (11.5)	<.001
CA/GD[c], n (%)	310 (26.6)	537 (7.4)	<.001
Respiratory disease, n (%)	228 (19.6)	279 (3.8)	<.001
Genital-urinary tract disease, n (%)	144 (12.3)	240 (3.3)	<.001
Vital signs[b]			
Lowest pulse (bpm), median (IQR)	136.0 (116.0-152.0)	104.0 (92.0-114.0)	<.001
Peak body temperature (°C), median (IQR)	37.6 (37.0-38.5)	38.4 (37.6-39.1)	<.001
Lowest DBP[d] (mm Hg), median (IQR)	60.0 (51.0-71.0)	66.0 (57.0-75.0)	<.001
Lowest SBP[e] (mm Hg), median (IQR)	102.0 (91.0-116.0)	107.0 (97.0-119.0)	<.001
Initial SBP (mm Hg), median (IQR)	110.0 (98.0-123.0)	112.0 (101.0-124.0)	.001
Pathogen			
Influenza virus type A, n (%)	14 (1.2)	169 (2.3)	.02
Influenza virus type B, n (%)	9 (0.8)	172 (2.4)	<.001
Streptococcus pneumoniae, n (%)	5 (0.4)	432 (5.9)	<.001
Lab data[b]			
Lymphocyte (%), median (IQR)	21.3 (12.6-36.5)	28.3 (17.2-42.9)	<.001
Creatinine (U/L), median (IQR)	0.5 (0.3-0.6)	0.4 (0.3-0.5)	<.001
Segment (%), median (IQR)	67.0 (49.0-79.3)	60.0 (44.4-73.0)	<.001
CRP[f] (mg/dL), median (IQR)	1.7 (0.5-5.6)	1.8 (0.6-4.4)	.43
Hemoglobin (g/dL), median (IQR)	12.7 (11.2-14.0)	12.5 (11.7-13.3)	.02

[a]Adjusted using the Benjamini-Hochberg procedure.

[b]Only the top 5 important features ranked by the Shapley Additive Explanations (SHAP) value are shown. The full table is shown in Table S4 in Multimedia Appendix 1.

[c]CA/GD: congenital anomaly/genetic disorder.

[d]DBP: diastolic blood pressure.

[e]SBP: systolic blood pressure.

[f]CRP: C-reactive protein.

Model Performance

After random selection, 6772 (6772/8464, 80.0%) records were included in the development set, and 1692 (1692/8464, 20.0%) were included in the validation set (Table 2). In the validation set, the RF model achieved the best performance in identifying patients transferred to the ICU within 24 hours after admission (AUC 0.987, 95% CI 0.981-0.992) compared with the XGB model (AUC 0.982, 95% CI 0.972-0.990) and logistic regression model (AUC 0.885, 95% CI 0.863-0.908). The average precision values were 0.932 (95% CI 0.904-0.956) for RF, 0.941 (95% CI 0.917-0.963) for the XGB algorithm, and 0.609 (95% CI 0.543-0.681) for the logistic regression model (Figure 2).

For the RF algorithm, at the point with the highest Youden index, the overall accuracy of the RF algorithm was 0.936 (95% CI 0.930–0.947), sensitivity was 0.940 (95% CI 0.919–0.954), and specificity was 0.935 (95% CI 0.924–0.952; Figure 2). At this threshold, there is approximately one false positive for every 3.1 positive predictions. At the point of highest sensitivity, which could include most patients with early ICU admission with some false alarms, the specificity was 0.868 (95% CI 0.642–0.917), and the negative predictive value was 0.998 (95% CI 0.995-1.000). At the point of highest specificity, which could avoid the most unnecessary ICU admissions, the sensitivity and positive predictive value (precision) for our RF algorithm were

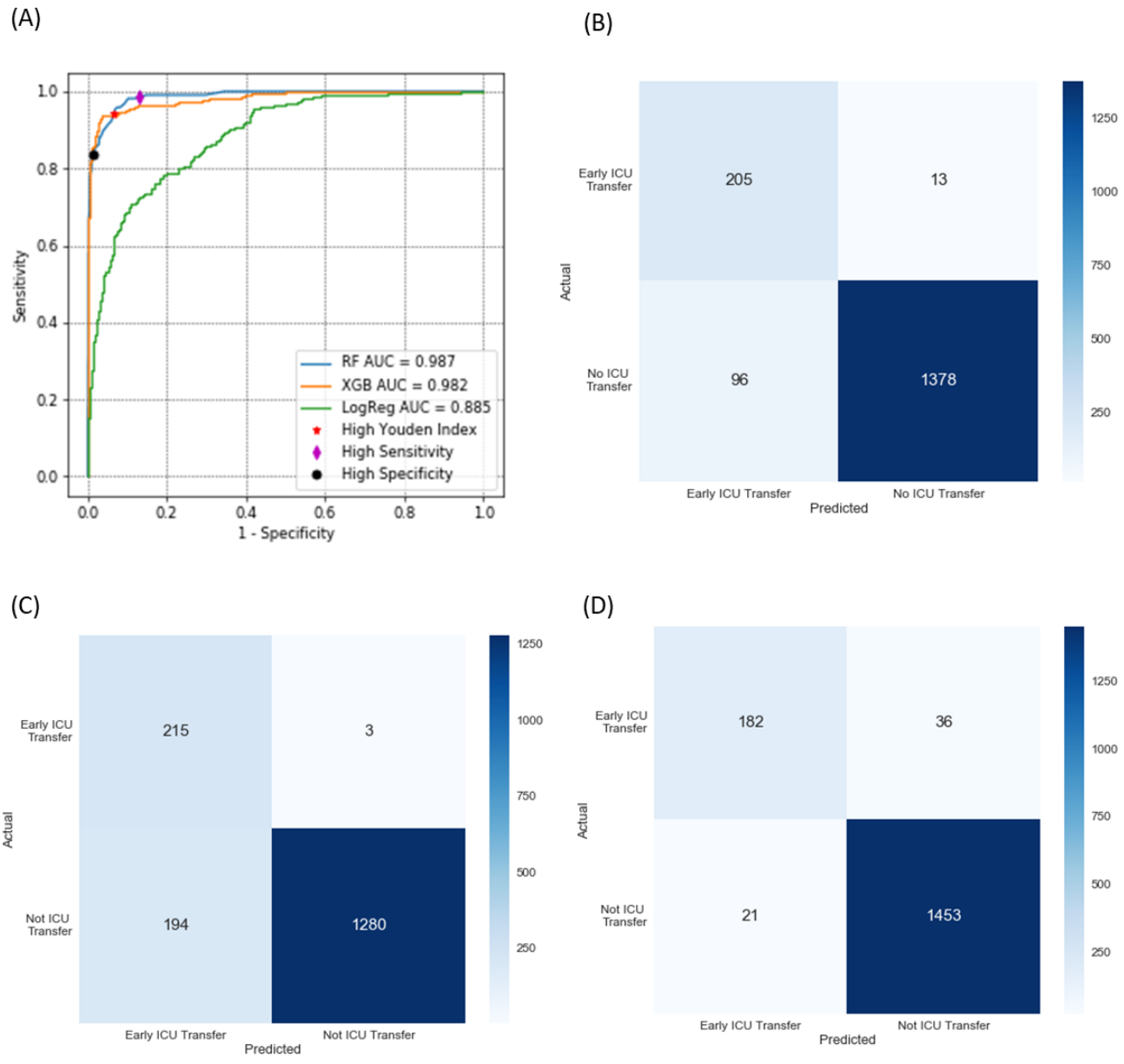
0.835 (95% CI 0.779-0.886) and 0.897 (95% CI 0.883–0.933), respectively.

Table 2. Basic characteristics of the development set and validation set.

Characteristics	Development set (n=6772)	Validation set (n=1692)
ICU[a] transfers or deaths within 24 hours after admission, n (%)	948 (14.0)	218 (12.9)
Unique individuals, n	5581	1576
Length of stay (days), median (IQR)	4.0 (3.0-7.0)	4.0 (3.0-7.0)
Age (years), mean (SD)	4.0 (3.5)	3.9 (3.3)
Male, n (%)	3625 (53.5)	914 (54.0)

[a]ICU: intensive care unit.

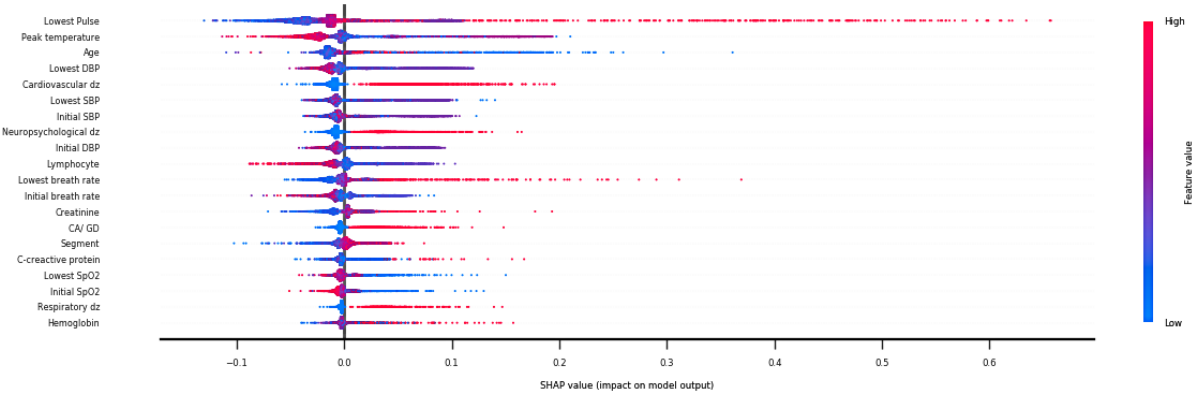
Figure 2. For the early intensive care unit (ICU) transfer and no ICU transfer groups, (A) receiver operating characteristic (ROC) curves and confusion matrices at the operational points with (B) the highest Youden index, (C) 0.99 sensitivity and the highest precision, and (D) 0.99 specificity and the highest sensitivity. AUC: area under the ROC curve; LogReg: logistic regression; RF: random forest; XGB: extreme gradient boosting.



Feature Importance From the ML Algorithm and Clinicians' Judgment

Figure 3 shows the top 20 features by relative importance from the RF algorithm based on SHAP values (see Figure S1 in Multimedia Appendix 2 for a complete list). The 5 most important features were lowest pulse rate, peak body temperature, age, lowest diastolic blood pressure, and presence of cardiovascular disease. For physician-rated relative feature importance, the presence of immunodeficiency; lowest oxygen saturation; and presence of solid neoplastic diseases, respiratory diseases, and cardiovascular diseases were considered the most important features (Figure S2 in Multimedia Appendix 3). The presence of cardiovascular diseases, the lowest systolic blood pressure, and the presence of neuropsychological diseases were ranked in the top 10 features with the highest importance measured by both SHAP values in the XGB model and physicians' judgment.

Figure 3. Top 20 important features of the random forest model based on Shapley Additive Explanations (SHAP) values. Every admission data point has one dot on each row for individual features. The color of the dot indicates the value of each feature from the admission data. The pile of dots on the same row to illustrate the density at different SHAP values. CA/GD congenital anomalies/genetic disorder; DBP: diastolic blood pressure; dz: disease; SBP: systolic blood pressure; SpO2: blood oxygen saturation.



Discussion

Principal Findings

Using the clinical data from 8464 admissions of children with pneumonia, we trained 2 ML algorithms to predict the need for ICU care within 24 hours of admission. Our study showed that ML algorithms could be applied to accurately triage hospitalized pediatric patients with pneumonia and effectively identify those who may need early ICU transfer. The high specificity and sensitivity of our algorithms supported their potential application in real-world clinical scenarios, which could provide a disease-specific alarm for severe conditions with the need for ICU care in a timely manner based on individual patient conditions. Because we only included the available features at admission, this design was considered more practical in clinical use. In addition, the list of feature importance could be explained by the clinical reasoning of human physicians. The explainability further validates the use of the ML approach for the clinical classification task. To our knowledge, our study is the first to explore the possibility of applying ML methods to large clinical data sets for triaging pediatric patients with pneumonia for ICU care.

The identification of a patient with the need for ICU care in the emergency room or in the early stage of the disease might influence medical care quality and clinical outcomes [5,36]. Previous work has revealed the ability to use decision tree–based algorithms to perform classification tasks in clinical scenarios or triage, with some promising preliminary results [13,14,16,17,20,24]. However, applications in clinical classification usually focus on triaging patients with different clinical severities and more general clinical diagnoses, such as respiratory failure, other organ failures, or sepsis [13,14,20,37]. Our work is one of the few studies to focus on a large data set for a specific diagnosis, pediatric pneumonia. Our algorithm's performance has better performance than previous studies that had AUCs ranging from approximately 0.7 to 0.9 [22-24,29], suggesting the advantage of an ML approach dedicated to children with pneumonia. With satisfactory performance, the application of the ML algorithms we proposed can be applied to support physicians' decisions for ICU care based on individual patient conditions and further improve health care quality during hospitalization. It can also help reduce clinicians' burden during outbreaks of community-acquired pneumonia, such as the recent COVID-19 outbreak, or in hospitals with insufficient human resources.

Because we could set up different operational points for the algorithm, our algorithm could be applied in various clinical settings. For example, at the high sensitivity operational point, the specificity could be kept at 0.868 (95% CI 0.642-0.917) with a negative predictive value of 0.998 (95% CI 0.995-1.000), which could be used to rule out those who did not need ICU care. Medical centers accommodating single-digit inpatients with pneumonia per day can operate on this threshold. Using the high sensitivity point, we could help clinicians identify patients who might need ICU admission earlier and reduce the number of undertriaged patients. Although there were one-quarter of the results as false positives, the burden is acceptable when the number of inpatients per day remains low, and false negatives are more harmful. When we further examined the medical records of those false negative cases in the current data set, we found that older age might be related with false negative results. Therefore, clinicians should be aware of false negative results in older children when applying the algorithm for their decision support. In contrast, at the high

specificity point (0.99), our algorithm maintained a sensitivity of 0.835 (95% CI 0.779-0.886) and a positive predictive value of 0.897 (95% CI 0.883-0.933). The high specificity with a high positive predictive value suggest that the algorithm could prevent unnecessary ICU admissions, so it may be applied when health care resources are limited or an outbreak happens. Therefore, the algorithm output could be customized according to the clinician's needs. In this way, the improved discriminability from ML algorithms could contribute to more accurate clinical decision-making and resource allocation. The ML model can not only provide automated estimation in clinical settings but also serve as a tool for training less experienced physicians or setting an alarm in hospitals with fewer human resources. Although the model does not reflect 100% of human physician decisions, it could be considered as a second opinion in the clinical setting and serve as a reference instead of being the only guideline for the final medical decision.

Our study also revealed important clinical feature indices (such as younger age, underlying diseases, higher pulse rate, and lower blood pressure) for the need for early ICU transfer, but patients with positive results for influenza A, influenza B, and *S.pneumoniae* at admission were less likely to be transferred to the ICU within 24 hours. These important clinical red flags could help physicians manage critically ill patients. In addition, early detection of the pathogens causing pneumonia in children makes early optimal treatment possible and improves the patient's clinical condition.

Limitations

There are some limitations in our study. First, we did not include imaging data, such as chest X-ray images, in our data set. However, diagnosis using the ICD codes relied on the physicians' clinical judgments, and clinicians might have already considered other clinical clues. Although most pneumonia patients are diagnosed clinically without specific radiological findings, including imaging data might still improve the judgment of clinical severity and thus influence the risk stratification for ICU care. Second, some clinically relevant parameters, such as blood gas values and procalcitonin measures, were not included in our algorithm training because of the high proportion of missing data. Third, the reasons for ICU admission usually varied (eg, ARDS, sepsis, respiratory failure, or other organ failures). Our algorithm could only evaluate the possible needs for ICU admission instead of the clinical diagnosis. With more data collected, an individual algorithm for a specific diagnosis might be developed in the future. Lastly, the algorithms were trained using a data set from a single medical center. Generalizability might be an issue if we would like to apply the findings to other hospital settings. Clinical validation in real-world settings might be required at the next stage to examine the application of ML algorithms in daily clinical work.

Comparisons With Prior Work

Compared with prior work that evaluated the need for ICU admission for pediatric patients, our disease-specific model for children with pneumonia demonstrated better performance. Our study incorporated up to 41 features from different domains (eg, demographics, vital signs, microbiological tests, and laboratory examinations) with no human-rated components (eg, behavior rating, respiratory difficulty). The strength of our tree-based ML approach is the ability to simultaneously process high-dimensional data linearly or nonlinearly [21]. With ML algorithms, we could integrate data with varying characteristics and solve complicated clinical questions (ie, predict the need for ICU care for hospitalized children with pneumonia). These characteristics enable the ML algorithm to include more clinical data and explore interactions among individual features, which was almost impossible to conduct with human intelligence or traditional statistical approaches, such as logistic regression. To further validate the algorithm's explainability, we invited 3 experienced physicians to grade the importance of ICU transfer evaluations from a clinical perspective. The results showed that features that were considered to be of higher importance by ML algorithms, such as the lowest systolic blood pressure and the presence of cardiovascular and neuropsychological diseases, were also considered essential features in the physicians' clinical judgment. The results helped us explain the findings of ML algorithms without being accused of using a "black box" for clinical decision-making. However, some discrepancies were still found. For example, human doctors tend to consider immunodeficiency and solid tumor diseases to be high-risk factors for early ICU transfer, but the importance of these 2 features in the ML algorithms is very low. This discrepancy between machine and human intelligence might be the consequence of proactive management for immunocompromised patients in clinical settings and thus inversely lowers the probability of early ICU admission. When applying the ML algorithm, we still have to consider this limitation in immunocompromised patients and combine the prediction of ML algorithms with clinical judgment. In this way, we could maximize support from machines without neglecting human intelligence.

Conclusions

In summary, we developed ML algorithms that could accurately classify the risk of early ICU transfer within 24 hours of admission for children with pneumonia. The clinical use of these algorithms might detect high-risk patients earlier and improve the quality of health care for pediatric pneumonia.

Acknowledgments

The authors would like to thank the staff of the Department of Medical Research who provided clinical data from the National Taiwan University Hospital-integrated Medical Database for their generous support.

This research was funded by grants from the Ministry of Science and Technology, Taiwan (grant numbers MOST 109-2634-F-002-029, MOST 109-2314-B-002-238, MOST 109-2321-B-002 -045, and 110-2314-B-002 -249 -). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

YL, HC, TC, and LC defined the research cohort and boundaries and formalized the design they studied. HC, TC, and LC selected the clinically relevant features for pediatric pneumonia patient intensive care unit (ICU) admission decisions and rated the relative importance. YL and LC had access to all data and were responsible for data integration. YL and TL performed data extraction and cleaning. YL and HC conducted statistical analysis, trained and validated the model, and created the figures. YL and HC contributed to the drafts of the manuscript. TC, TH, YY, and FL contributed to discussions on manuscript development. LC and CC reviewed and revised the manuscript critically. LC and FL obtained funding. All authors read and consented to the final submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOC File , 325 KB - medinform_v10i1e28934_app1.doc]

Multimedia Appendix 2
Relative feature importance of the random forest model based on Shapley values.
[PNG File , 116 KB - medinform_v10i1e28934_app2.png]

Multimedia Appendix 3
List of feature importance ranked by physicians.
[PNG File , 70 KB - medinform_v10i1e28934_app3.png]

References

1. Chou C, Shen C, Chen S, Chen H, Wang Y, Chang W, Infectious Diseases Society of Taiwan; Taiwan Society of Pulmonary and Critical Care Medicine; Medical Foundation in Memory of Dr. Deh-Lin Cheng; Foundation of Professor Wei-Chuan Hsieh for Infectious Diseases Research and Education; CY Lee's Research Foundation for Pediatric Infectious Diseases and Vaccines; 4th Guidelines Recommendations for Evidence-based Antimicrobial agents use in Taiwan (GREAT) working group. Recommendations and guidelines for the treatment of pneumonia in Taiwan. J Microbiol Immunol Infect 2019 Feb;52(1):172-199 [FREE Full text] [doi: 10.1016/j.jmii.2018.11.004] [Medline: 30612923]
2. Chen C, Lin P, Tsai M, Huang C, Tsao K, Wong K, et al. Etiology of community-acquired pneumonia in hospitalized children in northern Taiwan. Pediatr Infect Dis J 2012 Nov;31(11):e196-e201. [doi: 10.1097/INF.0b013e31826eb5a7] [Medline: 22935863]
3. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
5. Leroy O, Santré C, Beuscart C, Georges H, Guery B, Jacquier JM, et al. A five-year study of severe community-acquired pneumonia with emphasis on prognosis in patients admitted to an intensive care unit. Intensive Care Med 1995 Jan;21(1):24-31. [doi: 10.1007/BF02425150] [Medline: 7560469]
6. Bradley JS, Byington CL, Shah SS, Alverson B, Carter ER, Harrison C, Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. Clin Infect Dis 2011 Oct;53(7):e25-e76. [doi: 10.1093/cid/cir531] [Medline: 21880587]
7. Huang C, Chang L, Liu C, Huang Y, Chang L, Huang Y, Taiwan Pediatric Infectious Disease Alliance. Risk factors of progressive community-acquired pneumonia in hospitalized children: a prospective study. J Microbiol Immunol Infect 2015 Feb;48(1):36-42 [FREE Full text] [doi: 10.1016/j.jmii.2013.06.009] [Medline: 23993455]
8. Matics TJ, Sanchez-Pinto LN. Adaptation and validation of a pediatric Sequential Organ Failure Assessment score and evaluation of the sepsis-3 definitions in critically ill children. JAMA Pediatr 2017 Oct 02;171(10):e172352 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.2352] [Medline: 28783810]

9. Egdell P, Finlay L, Pedley DK. The PAWS score: validation of an early warning scoring system for the initial assessment of children in the emergency department. Emerg Med J 2008 Nov;25(11):745-749. [doi: 10.1136/emj.2007.054965] [Medline: 18955610]
10. Parshuram CS, Hutchison J, Middaugh K. Development and initial validation of the Bedside Paediatric Early Warning System score. Crit Care 2009;13(4):R135 [FREE Full text] [doi: 10.1186/cc7998] [Medline: 19678924]
11. Seiger N, Maconochie I, Oostenbrink R, Moll HA. Validity of different pediatric early warning scores in the emergency department. Pediatrics 2013 Oct;132(4):e841-e850. [doi: 10.1542/peds.2012-3594] [Medline: 24019413]
12. Gold DL, Mihalov LK, Cohen DM. Evaluating the Pediatric Early Warning Score (PEWS) system for admitted patients in the pediatric emergency department. Acad Emerg Med 2014 Nov;21(11):1249-1256 [FREE Full text] [doi: 10.1111/acem.12514] [Medline: 25377402]
13. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019 Aug;572(7767):116-119 [FREE Full text] [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
14. Goto T, Camargo CA, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. JAMA Netw Open 2019 Jan 04;2(1):e186937 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.6937] [Medline: 30646206]
15. Raita Y, Camargo CA, Macias CG, Mansbach JM, Piedra PA, Porter SC, et al. Machine learning-based prediction of acute severity in infants hospitalized for bronchiolitis: a multicenter prospective study. Sci Rep 2020 Jul 03;10(1):10979 [FREE Full text] [doi: 10.1038/s41598-020-67629-8] [Medline: 32620819]
16. Zhang Z, Liu J, Xi J, Gong Y, Zeng L, Ma P. Derivation and validation of an ensemble model for the prediction of agitation in mechanically ventilated patients maintained under light sedation. Crit Care Med 2021 Mar 01;49(3):e279-e290. [doi: 10.1097/CCM.0000000000004821] [Medline: 33470778]
17. Dziadzko MA, Novotny PJ, Sloan J, Gajic O, Herasevich V, Mirhaji P, et al. Multicenter derivation and validation of an early warning score for acute respiratory failure or death in the hospital. Crit Care 2018 Oct 30;22(1):286 [FREE Full text] [doi: 10.1186/s13054-018-2194-7] [Medline: 30373653]
18. Zhang Z, Navarese EP, Zheng B, Meng Q, Liu N, Ge H, et al. Analytics with artificial intelligence to advance the treatment of acute respiratory distress syndrome. J Evid Based Med 2020 Nov;13(4):301-312. [doi: 10.1111/jebm.12418] [Medline: 33185950]
19. Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. Sci Rep 2019 Aug 14;9(1):11862 [FREE Full text] [doi: 10.1038/s41598-019-48263-5] [Medline: 31413285]
20. Goto T, Camargo CA, Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. Am J Emerg Med 2018 Sep;36(9):1650-1654. [doi: 10.1016/j.ajem.2018.06.062] [Medline: 29970272]
21. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
22. Zhai H, Brady P, Li Q, Lingren T, Ni Y, Wheeler DS, et al. Developing and evaluating a machine learning based algorithm to predict the need of pediatric intensive care unit transfer for newly hospitalized children. Resuscitation 2014 Aug;85(8):1065-1071 [FREE Full text] [doi: 10.1016/j.resuscitation.2014.04.009] [Medline: 24813568]
23. Mayampurath A, Jani P, Dai Y, Gibbons R, Edelson D, Churpek MM. A vital sign-based model to predict clinical deterioration in hospitalized children. Pediatr Crit Care Med 2020 Sep;21(9):820-826 [FREE Full text] [doi: 10.1097/PCC.0000000000002414] [Medline: 32511200]
24. Rubin J, Potes C, Xu-Wilson M, Dong J, Rahman A, Nguyen H, et al. An ensemble boosting model for predicting transfer to the pediatric intensive care unit. Int J Med Inform 2018 Apr;112:15-20. [doi: 10.1016/j.ijmedinf.2018.01.001] [Medline: 29500014]
25. Chumbita M, Cillóniz C, Puerta-Alcalde P, Moreno-García E, Sanjuan G, Garcia-Pouton N, et al. Can artificial intelligence improve the management of pneumonia. J Clin Med 2020 Jan 17;9(1):248 [FREE Full text] [doi: 10.3390/jcm9010248] [Medline: 31963480]
26. Visweswaran S, Cooper GF. Patient-specific models for predicting the outcomes of patients with community acquired pneumonia. AMIA Annu Symp Proc 2005:759-763 [FREE Full text] [Medline: 16779142]
27. Hu C, Chen C, Fang Y, Liang S, Wang H, Fang W, TSIRC (Taiwan Severe Influenza Research Consortium). Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan. BMJ Open 2020 Feb 25;10(2):e033898 [FREE Full text] [doi: 10.1136/bmjopen-2019-033898] [Medline: 32102816]
28. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. 2015 Presented at: 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 10-13, 2015; Sydney, NSW. [doi: 10.1145/2783258.2788613]

29. Renaud B, Labarère J, Coma E, Santin A, Hayon J, Gurgui M, et al. Risk stratification of early admission to the intensive care unit of patients with no major criteria of severe community-acquired pneumonia: development of an international prediction rule. Crit Care 2009;13(2):R54 [FREE Full text] [doi: 10.1186/cc7781] [Medline: 19358736]
30. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Crit Care 2019 Apr 08;23(1):112 [FREE Full text] [doi: 10.1186/s13054-019-2411-z] [Medline: 30961662]
31. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. Crit Care Med 2018 Jul;46(7):1070-1077. [doi: 10.1097/CCM.0000000000003123] [Medline: 29596073]
32. Youden WJ. Index for rating diagnostic tests. Cancer 1950 Jan;3(1):32-35. [doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3] [Medline: 15405679]
33. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Polat, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011;12:2825-2830 [FREE Full text]
35. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939785]
36. Nadkarni VM, Larkin GL, Peberdy MA, Carey SM, Kaye W, Mancini ME, National Registry of Cardiopulmonary Resuscitation Investigators. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. JAMA 2006 Jan 04;295(1):50-57. [doi: 10.1001/jama.295.1.50] [Medline: 16391216]
37. Masino AJ, Harris MC, Forsyth D, Ostapenko S, Srinivasan L, Bonafide CP, et al. Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. PLoS One 2019;14(2):e0212665 [FREE Full text] [doi: 10.1371/journal.pone.0212665] [Medline: 30794638]

Abbreviations

ARDS: acute respiratory distress syndrome
AUC: area under the receiver operating characteristic curve
ICD: International Classification of Diseases
ICU: intensive care unit
ML: machine learning
PEWS: Pediatric Early Warning Score
RF: random forest
SHAP: Shapley Additive Explanations
XGB: extreme-gradient boosting

Edited by C Lovis; submitted 30.03.21; peer-reviewed by G Rees, Z Zhang, J Zhang; comments to author 11.05.21; revised version received 01.07.21; accepted 02.01.22; published 27.01.22.

Please cite as:
Liu YC, Cheng HY, Chang TH, Ho TW, Liu TC, Yen TY, Chou CC, Chang LY, Lai F
Evaluation of the Need for Intensive Care in Children With Pneumonia: Machine Learning Approach
JMIR Med Inform 2022;10(1):e28934
URL: https://medinform.jmir.org/2022/1/e28934
doi:10.2196/28934
PMID:35084358

Original Paper

Incidence of Diagnostic Errors Among Unexpectedly Hospitalized Patients Using an Automated Medical History–Taking System With a Differential Diagnosis Generator: Retrospective Observational Study

Ren Kawamura[1], MD; Yukinori Harada[1,2], MD, PhD; Shu Sugimoto[2], MD; Yuichiro Nagase[2], MD; Shinichi Katsukura[1], MD; Taro Shimizu[1], MPH, MBA, MD, PhD

[1]Department of Diagnostic and Generalist Medicine, Dokkyo Medical University, Mibu, Japan

[2]Department of Internal Medicine, Nagano Chuo Hospital, Nagano, Japan

Corresponding Author:
Taro Shimizu, MPH, MBA, MD, PhD
Department of Diagnostic and Generalist Medicine
Dokkyo Medical University
880 Kitakobayashi
Mibu, 321-0293
Japan
Phone: 81 282861111
Email: shimizutaro7@gmail.com

Abstract

Background: Automated medical history–taking systems that generate differential diagnosis lists have been suggested to contribute to improved diagnostic accuracy. However, the effect of these systems on diagnostic errors in clinical practice remains unknown.

Objective: This study aimed to assess the incidence of diagnostic errors in an outpatient department, where an artificial intelligence (AI)–driven automated medical history–taking system that generates differential diagnosis lists was implemented in clinical practice.

Methods: We conducted a retrospective observational study using data from a community hospital in Japan. We included patients aged 20 years and older who used an AI-driven, automated medical history–taking system that generates differential diagnosis lists in the outpatient department of internal medicine for whom the index visit was between July 1, 2019, and June 30, 2020, followed by unplanned hospitalization within 14 days. The primary endpoint was the incidence of diagnostic errors, which were detected using the Revised Safer Dx Instrument by at least two independent reviewers. To evaluate the effect of differential diagnosis lists from the AI system on the incidence of diagnostic errors, we compared the incidence of these errors between a group where the AI system generated the final diagnosis in the differential diagnosis list and a group where the AI system did not generate the final diagnosis in the list; the Fisher exact test was used for comparison between these groups. For cases with confirmed diagnostic errors, further review was conducted to identify the contributing factors of these errors via discussion among three reviewers, using the Safer Dx Process Breakdown Supplement as a reference.

Results: A total of 146 patients were analyzed. A final diagnosis was confirmed for 138 patients and was observed in the differential diagnosis list from the AI system for 69 patients. Diagnostic errors occurred in 16 out of 146 patients (11.0%, 95% CI 6.4%-17.2%). Although statistically insignificant, the incidence of diagnostic errors was lower in cases where the final diagnosis was included in the differential diagnosis list from the AI system than in cases where the final diagnosis was not included in the list (7.2% vs 15.9%, *P*=.18).

Conclusions: The incidence of diagnostic errors among patients in the outpatient department of internal medicine who used an automated medical history–taking system that generates differential diagnosis lists seemed to be lower than the previously reported incidence of diagnostic errors. This result suggests that the implementation of an automated medical history–taking system that generates differential diagnosis lists could be beneficial for diagnostic safety in the outpatient department of internal medicine.

(JMIR Med Inform 2022;10(1):e35225) doi:10.2196/35225

KEYWORDS

artificial intelligence; automated medical history–taking; diagnostic errors; outpatient; Safer Dx

Introduction

Diagnostic error, defined as the failure to establish an accurate and timely explanation of the patient's health problem or to communicate that explanation to the patient [1], is one of the most important patient safety issues that should be addressed [2,3]. The impact of diagnostic errors on patient safety is quite large [4]. First, diagnostic errors comprise around 50% of preventable harm in primary health care settings and emergency departments [5]. Second, the risk of death, significant permanent injury, and prolonged hospitalization is higher for diagnostic error cases than for other medical errors [6-12]. Third, diagnostic errors frequently occur in several settings of clinical practice; approximately 5% of patients can experience diagnostic errors in primary health care and hospital practice in the United States [13]. Therefore, effective interventions to reduce diagnostic errors are warranted.

Diagnostic error–related paid malpractice claims occur more often among outpatients than among inpatients [9], suggesting that the primary health care outpatient setting is vulnerable to diagnostic errors. The prevalence of diagnostic errors in outpatient settings has been reported to be between 3.6% and 5.1%. However, when focusing on a population of patients with a high risk for diagnostic errors who were unexpectedly hospitalized within 14 days after the index outpatient visit, the prevalence of diagnostic errors increased to as much as 21% [14]. The common contributing factors for diagnostic errors in primary care outpatient settings were reported to include problems with history-taking, overreliance on pattern recognition, and failure to consider sufficient differential diagnoses [4,15]. Therefore, strategies or systems to improve the quality of history-taking and support differential diagnosis generation are required to reduce diagnostic errors in outpatient settings.

From this perspective, newly developed technology, such as computerized automated history-taking systems and diagnostic decision support systems, can be leveraged to address this issue; these systems have a long history, since they were introduced in the 1960s and 1970s [16-18]. Computerized automated history-taking systems perform better in clinical documentation tasks for taking patient histories than do physicians [19,20]. The use of a diagnostic support system (ie, differential diagnosis generator) before collecting information by physicians showed a significant impact on the improvement of diagnostic accuracy in terms of clinical reasoning and differential diagnosis [21-23]. Moreover, a new system that combines automated medical history–taking functions with differential diagnosis generation—specialized for musculoskeletal diseases only—showed improved diagnostic accuracy among physicians in a pilot randomized controlled trial [24]. Subsequently, another system, covering broad symptoms of internal diseases, was developed and implemented in clinical practice [25]. Yet another study showed high reliability of documentation regarding clinical history to assist the diagnostic accuracy of physicians [26]; however, this was not conducted in a clinical practice setting.

These automated systems have generated concerns about their negative effects on the diagnostic accuracy of physicians. For instance, physicians may not accept correct diagnoses or may accept incorrect diagnoses generated by the systems [24,26], partly because physicians tend to be more confident with their own diagnosis than that of artificial intelligence (AI) systems when there is a discrepancy between them [27]. Therefore, the effects of the implementation of these systems on diagnostic errors in clinical practice remain unknown. This study aimed to assess the incidence of diagnostic errors in an outpatient department, where an AI-driven automated medical history–taking system that generates differential diagnosis lists was implemented in clinical practice.

Methods

Study Design

We conducted a retrospective observational study using data from Nagano Chuo Hospital in Japan. The Research Ethics Committee of Nagano Chuo Hospital approved this study (serial number: NCR202104). The requirement to obtain written informed consent from patients was waived by the Research Ethics Committee under the condition that we used an opt-out method. We informed patients by showing the detailed information of the study on the official website of Nagano Chuo Hospital.

Patient Population

We included patients aged 20 years and older who used AI Monshin—an AI-based automated medical history–taking system—in the outpatient department of internal medicine for whom the index visit was between July 1, 2019, and June 30, 2020, followed by unplanned hospitalization within 14 days. A follow-up duration of 14 days was selected to improve the sensitivity to detect diagnostic errors [14,28]. For assessing the effects of using AI Monshin on diagnostic errors, we excluded patients for whom AI Monshin did not list 10 differential diagnoses. In those cases, the AI system could not complete history-taking because patients gave up entering information or because they presented to the hospital for further investigation of abnormal test results following their annual health checkup, which was out of scope for the system during the study period. Usually, even one differential diagnosis was not generated in such cases.

Presentation of the AI Monshin Tool

The details of AI Monshin were presented in a previous report [25]. In brief, AI Monshin converts data entered by patients on tablet terminals into medical terms. Patients enter their background information, such as age and sex, and chief complaint as free text on a tablet in the waiting room. AI Monshin asks approximately 20 questions, one by one, which are tailored to the patient. The questions are optimized, based on previous answers, to generate the most relevant list of

potential differential diagnoses. Physicians can see the entered data as a summarized medical history with the top 10 possible differential diagnoses, along with their rank.

Identification of Diagnostic Errors

To identify whether diagnostic errors occurred in this study, we used the Revised Safer Dx Instrument [29]. The Safer Dx Instrument is an externally validated, structured data collection tool to improve the accuracy of assessment of diagnostic errors [30,31]; the tool has been widely used in several studies on diagnostic errors [32-36]. Recently, the tool was updated as the Revised Safer Dx Instrument [29]. The Revised Safer Dx Instrument consists of 13 items. Items 1 to 12 are used for assessing the diagnostic process, and item 13 is used to determine the possibility of diagnostic error. All items are rated by answering questions on a scale ranging from 1 (strongly disagree) to 7 (strongly agree). The Revised Safer Dx Instrument can be used to assess the entire diagnostic process of one event; however, because we focused on diagnostic errors related to the implementation of AI Monshin, which seems to mainly influence the diagnostic decision at the index visit, the evaluation of diagnostic errors in this study was based on the medical records taken during the index visit.

The identification of diagnostic errors in this study was conducted through the algorithm as discussed in this section. In the first step, two reviewers (YH and SS) independently evaluated the diagnostic process of included cases using the Revised Safer Dx Instrument by reviewing the medical records. The presence or absence of diagnostic errors in each case was judged based on the score of item 13 [29]. According to the recommendation for using the Revised Safer Dx Instrument, diagnostic error was confirmed in cases where both reviewers scored 5 or higher on item 13, and diagnostic error was denied in cases where both reviewers scored 3 or lower on item 13 [29]. The remaining cases were progressed to the second step. In the second step, the third reviewer (YN) independently evaluated the cases using the Revised Safer Dx Instrument. Diagnostic error was confirmed in cases where two out of three reviewers scored 5 or higher on item 13, and diagnostic error was denied in cases where two out of three reviewers scored 3 or lower on item 13. For the remaining cases in which diagnostic error was neither confirmed nor denied, the three reviewers (YH, SS, and YN) discussed and mutually agreed on whether diagnostic error occurred or not on a case-by-case basis.

The final diagnoses of all cases were confirmed by two reviewers (YH and SS) based on the discharge summary. Disagreements were resolved by discussion among the three reviewers (YH, SS, and YN). Based on the confirmed final diagnoses, the other two reviewers (RK and SK), who were blinded to the evaluation of diagnostic errors, independently judged whether the final diagnosis of each case was included in the list of 10 differential diagnoses generated by AI Monshin. Disagreements were resolved by discussion between the two reviewers (RK and SK).

Analysis of the Causes of Diagnostic Errors

For cases with confirmed diagnostic errors, further review was conducted to identify the contributing factors of these errors via discussion among the three reviewers (YH, SS, and YN). The Safer Dx Process Breakdown Supplement was used as a reference to classify the contributing factors of diagnostic errors and outcomes in this study [29]. To evaluate the effects of AI Monshin implementation on the diagnostic errors, other than the items in the Safer Dx Process Breakdown Supplement, the following were discussed: the frequency of the final diagnosis (ie, whether the disease was common or uncommon), typicality of the presentation for the final diagnosis (ie, typical or atypical), and initial diagnosis at the index visit.

Baseline Data Collection and Outcome

From the medical records, we extracted data on the age and sex of patients, chief complaints, and the experience of physicians who saw patients at the index visits (ie, resident: up to 5 years of experience after graduation; staff: more than 5 years of experience after graduation). The primary outcome was the incidence of diagnostic errors.

Sample Size Calculation

We calculated the required sample size to be 139 cases, with an incidence of diagnostic errors of 10.0% and a margin of 5.0%. It was estimated that there were approximately 150 patients who were eligible for this study between July 1, 2019, and June 30, 2020. Even with the expectation that approximately 5 to 10 cases could be excluded, 150 cases were a reasonable target number of cases for this study.

Statistical Analysis

Continuous data are presented as medians with the 25th and 75th percentiles. Categorical data are presented as counts and proportions (%). For the primary outcome, we calculated the incidence of diagnostic errors with 95% CI. To evaluate the baseline factors and the differential diagnosis list of AI Monshin with regard to the incidence of diagnostic errors, we compared the incidence of diagnostic errors between the groups of older adults (aged ≥65 years) and non–older adults (aged <65 years) [37-40], the groups of males and females [33], the groups seen by staff and seen by residents [26], and the groups in which AI Monshin generated or did not generate the final diagnosis in the differential diagnosis list [26]; these comparisons were made using the Fisher exact test. We also calculated the odds ratio (OR) with 95% CI for the incidence of diagnostic errors in these groups. *P* values were based on 2-tailed statistical tests, and *P* values less than .05 were considered statistically significant. All statistical analyses were conducted using R (version 4.1.0; The R Foundation).

Results

Baseline Patient Characteristics

A total of 150 cases were unexpectedly hospitalized within 14 days after the index visit that took place at the outpatient department of internal medicine; AI Monshin was used at the index visit. Only 2 (1.3%) patients did not complete history-taking by AI Monshin: a woman in her 70s complained of an uncomfortable feeling on her tongue, abdominal pain with distention, and appetite loss, and a man in his 70s complained that his cold was not getting better. After excluding 4 (2.7%)

cases in which AI Monshin did not develop 10 differential diagnoses (2 cases: incomplete history-taking; 2 cases: patients presented for further investigation for abnormal test results), the data from 146 cases were analyzed for this study. The median age of the patients was 71 (IQR 59-82) years, 72 (49.3%) were male, 71 (48.6%) were seen by residents at the index visit, and 103 (70.5%) were admitted to the hospital on the same day as the index visit.

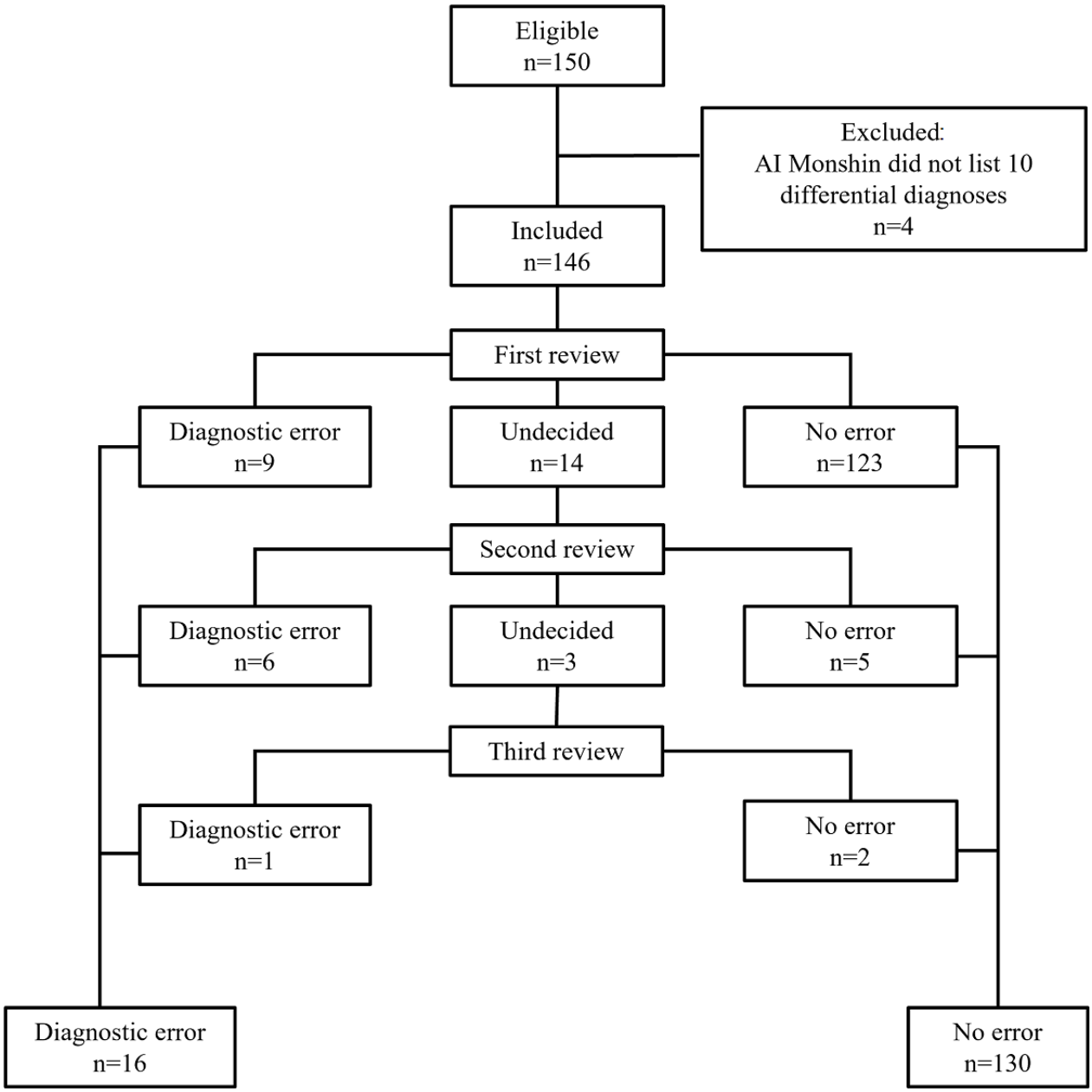
Chief Complaints and the Final Diagnosis

The top three most common chief complaints were abdominal pain (37/146, 25.3%), fever (20/146, 13.7%), and melena or hematochezia (15/146, 10.3%). During follow-up outpatient visits or admission, the final diagnosis was confirmed for 138 patients (94.5%). The most common diagnosis was lower respiratory tract infection (15/138, 10.9%), followed by ischemic colitis (8/138, 5.8%), diverticular bleeding (8/138, 5.8%), and congestive heart failure (8/138, 5.8%). The final diagnosis was based on the differential diagnosis list from AI Monshin for 69 out of 138 patients (50.0%).

Primary Outcome

Figure 1 shows the steps of the review for confirming the diagnostic errors in this study. In the first step of the review, diagnostic errors were confirmed in 9 cases and denied in 123 cases. Among the remaining 14 cases, diagnostic errors were confirmed in 6 cases and denied in 5 cases in the second step of the review. Among the remaining 3 cases, diagnostic errors were confirmed in 1 case and denied in 2 cases in the third step of the review. In total, diagnostic errors were confirmed in 16 out of 146 cases (11.0%, 95% CI 6.4%-17.2%).

Figure 1. Flow of reviews for confirming diagnostic errors. AI: artificial intelligence.



The incidence of diagnostic errors was significantly higher in patients aged 65 years and older compared to those under 65 years of age (15/96, 16% vs 1/50, 2%; OR 9.1, 95% CI 1.2-70.8; *P*=.01). There were no significant differences in the incidence of diagnostic errors between male and female patients (11/72, 15% vs 5/74, 7%; OR 2.5, 95% CI 0.8-7.6; *P*=.12), between patients who were seen by a resident and those who were seen by a physician at the index visit (9/71, 13% vs 7/75, 9%; OR 1.4, 95% CI 0.5-4.0; *P*=.60), and between cases in which the final diagnosis was not included in the differential diagnosis list from AI Monshin and those in which the final diagnosis was included in the same list (11/69, 16% vs 5/69, 7%; OR 2.4, 95% CI 0.8-7.4; *P*=.18).

Details Regarding Cases With Diagnostic Errors

Table 1 and Multimedia Appendix 1 show the details of the 16 cases where there were diagnostic errors. All cases had common final diagnoses (ie, cholangitis, cholecystitis, diverticular bleeding, pneumonia, interstitial pneumonia, intestinal obstruction, pyelonephritis, infectious enteritis, heart failure, and pulmonary artery embolism), and the final diagnosis presentation was typical for 15 out of 16 cases (94%). The most common chief complaint in the 16 cases with diagnostic errors was abdominal pain (n=5, 31%), followed by cough (n=4, 25%) and fever (n=3, 19%).

According to the Safer Dx Process Breakdown Supplement, the most common contributing factors for diagnostic errors in 16 cases were “problems ordering diagnostic tests for further workup” (n=13, 81%), followed by “problems with data integration and interpretation” (n=10, 63%), “problems with physical exam” (n=9, 56%), and “performed tests not interpreted correctly” (n=8, 50%; Table 2).

From the aspect of the differential diagnosis list for cases with diagnostic errors, AI Monshin listed the final diagnosis in the list in 5 out of 16 cases (31%) and the initial diagnosis in 4 out of 16 cases (25%). On the other hand, in cases without diagnostic errors, AI Monshin listed the final diagnosis in the differential list in 64 out of 122 cases (52.5%, excluding 8 cases where the final diagnosis was unknown). In summary, despite using AI Monshin, physicians could not make the correct diagnoses as were suggested in the differential diagnosis list in 5 of 69 cases (7% omission errors). On the other hand, the incorrect initial diagnoses made by physicians were listed in the differential diagnosis list in 4 of 69 cases (6% commission errors). Regarding the outcome, no cases of diagnostic errors resulted in death or permanent harm. A total of 2 cases out of 16 (13%) were classified as Category C: “An error occurred that reached the patient but did not cause the patient harm.” Diagnostic errors resulted in some harm in 14 out of 16 cases (88%; 2 cases were classified as Category E: “An error occurred that may have contributed to or resulted in temporary harm to the patient and required intervention”; 12 cases were classified as Category F: “An error occurred that may have contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization”). The median time between the index visit and the time that the final diagnosis was made was 3 (IQR 2-6) days.

Table 1. The details of 16 diagnostic error cases.

Case No.[a]	Age (y)	Sex[b]	Physician of first contact	Chief complaint	Initial diagnosis	Final diagnosis	Index visit to final diagnosis (days), n	Outcome category[c]	Initial diagnosis was on list[d]	Final diagnosis was on list[d]
1	95	F	Resident	Fever	URI[e]	Cholangitis	4	F	No	No
2	76	M	Resident	Abdominal pain	GERD[f]	Cholecystitis	2	F	Yes; rank 4	No
3	83	M	Resident	Abdominal pain	Costo-chondritis	Pneumonia	3	F	No	No
4	55	M	Resident	Hematochezia	Infectious enteritis	Diverticular bleeding	2	F	Yes; rank 3	Yes; rank 1
5	89	F	Staff	Nausea	Unknown	Acute pyelonephritis	3	F	No	No
6	75	M	Staff	Cough	URI	Interstitial pneumonia	3	F	No	Yes; rank 10
7	66	M	Resident	Abdominal pain	Constipation	Intestinal obstruction	6	F	Yes; rank 4	No
8	70	F	Staff	Cough	Unknown	Heart failure	3	F	No	Yes; rank 8
9	77	F	Resident	Palpitation	Heart failure	Pulmonary embolism	2	E	Yes; rank 10	No
10	82	M	Staff	Fever	URI	Cholecystitis	3	F	No	No
11	81	F	Resident	Anorexia	Choledocholithiasis	Acute pyelonephritis	2	C	No	No
12	72	M	Staff	Headache, lightheadedness	Fatigue	Vestibular neuritis	8	E	No	No
13	86	M	Resident	Abdominal pain	Enteritis	Intestinal obstruction	0[g]	F	No	Yes; rank 9
14	78	M	Staff	Abdominal pain	Hemorrhoid	Infectious enteritis	9	C	No	No
15	91	M	Staff	Fever, cough, back pain	URI	Acute pyelonephritis	7	F	No	Yes; rank 3
16	72	M	Resident	Dyspnea, cough, malaise	URI	Interstitial pneumonia	11	F	No	No

[a]All diagnoses were common. All cases had typical presentations except for case 2.

[b]Female (F) or male (M).

[c]Outcome was classified, along with the Safer Dx Process Breakdown Supplement, as follows: Category C, "An error occurred that reached the patient but did not cause the patient harm"; Category E, "An error occurred that may have contributed to or resulted in temporary harm to the patient and required intervention"; Category F, "An error occurred that may have contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization" [29].

[d]AI Monshin's differential list; where a diagnosis was on the list, its rank on the list is indicated.

[e]URI: upper respiratory infection.

[f]GERD: gastroesophageal reflux disease.

[g]The final diagnosis was made at the second visit, which was on the same day as the index visit.

Table 2. Breakdown analysis of the contributing factors for diagnostic errors.

Contributing factors and details	Cases (N=16), n (%)
Patient-related factors	
Delay in seeking care	0 (0)
Lack of adherence to appointments	0 (0)
Other	0 (0)
Patient-provider encounter	
Problems with history	4 (25)
Problems with physical exam	9 (56)
Problems ordering diagnostic tests for further workup	13 (81)
Failure to review previous documentation	4 (25)
Problems with data integration and interpretation	10 (63)
Other	0 (0)
Diagnostic tests	
Ordered test was not performed at all	0 (0)
Ordered test was not performed correctly	0 (0)
Performed test was not interpreted correctly	8 (50)
Misidentification	1 (6)
Other	0 (0)
Follow-Up and tracking	
Problems with timely follow-up of abnormal diagnostic test results	1 (6)
Problems with scheduling of appropriate and timely follow-up visits	2 (13)
Problems with diagnostic specialties returning test results to clinicians	2 (13)
Problems with clinicians reviewing test results	0 (0)
Problems with clinicians documenting action or response to test results	0 (0)
Problems with notifying patients of test results	0 (0)
Problems with monitoring patients through follow-up	0 (0)
Other	0 (0)
Referrals	
Problems initiating referral	1 (6)
Lack of appropriate actions on requested consultation	0 (0)
Communication breakdown from consultant to referring provider	0 (0)
Other	0 (0)

Discussion

Principal Findings

Among 146 patients who used the AI-driven, automated history-taking system, which developed a list of the top 10 differential diagnoses, diagnostic errors occurred in 11.0% of cases. These patient histories were collected at the index visit to the outpatient department of internal medicine, followed by unplanned hospitalization of the patient within 14 days. The incidence of diagnostic errors was statistically higher among older adult patients; however, the sex of the patients, the experience of the physicians, and the accuracy of the differential diagnosis list of the AI system were not statistically associated with the incidence of diagnostic errors. In all cases where diagnostic errors occurred, the final diagnoses were common diseases, as reported in a previous study that was conducted in primary care settings in the United States between 2006 and 2007 [4], and the clinical presentation was typical, except in one case.

Limitations

To the best of our knowledge, this is the first observational study that evaluated the effects of implementation of an automated medical history–taking system with a differential diagnosis generator in routine clinical practice using the validated Revised Safer Dx Instrument to detect diagnostic errors. However, this study also had some limitations. First, this

study did not include patients who did not use an automated history-taking system with a differential diagnosis generator or those who were not admitted; therefore, the incidence of diagnostic errors should be interpreted with caution. Second, exclusion of the cases in which AI Monshin did not develop 10 differential diagnoses may have reduced the incidence of diagnostic errors in this study. Since inadequate and inappropriate history could be a contributing factor for diagnostic errors, excluding such a case may merit the optimistic assumption of AI Monshin's performance. Third, because the judgment of diagnostic errors was conducted by a retrospective review of the charts, some bias could not be avoided. However, as the review process was predefined and at least two reviewers independently assessed each case, we are sure that these biases were avoided as much as possible. Fourth, we are unsure of the effects of COVID-19 on diagnostic errors in the outpatient department. Future studies may focus on the incidence of diagnostic errors between hospitals with and without implementation of an automated medical history–taking system with a diagnostic decision support function in a prospective design.

Comparison With Prior Work

The incidence of diagnostic errors in this study was 11.0%, which was lower than that reported in previous studies (13.7% and 20.9%) that included cases similar to this study (ie, patients who were unexpectedly hospitalized within 14 days after their index visit) [14,28]. In addition, the incidence of diagnostic errors in this study was lower than that reported in retrospective studies with chart review (13.3% to 21.8%) [11,41-43] or in prospective studies (12.3% to 20.0%) [12,44] that investigated the rate of discrepancy in the diagnosis between admission and discharge. Therefore, it is possible that the implementation of an automated history-taking system with a differential diagnosis generator reduced the incidence of diagnostic errors in the outpatient department of internal medicine.

The quality of clinical history documented by AI Monshin may be a key component of the results. There may be high discrepancies in clinical history between patient reports and physician documentation [45]; in addition, the automated medical history–taking system, as compared to physicians, may have the potential to take clinical histories that are more diagnostically useful and of higher quality [19,20]. Therefore, routine use of automated history-taking systems may improve diagnostic accuracy by establishing a high-quality base of clinical history for the correct diagnosis. Indeed, in a previous study that used the documentation made by an automated medical history–taking system from real patients, the correct diagnosis appeared in 56.3% of the top three differential diagnoses made by physicians without using a differential diagnosis list from an AI-driven system; this increased to 72.7% in cases where the correct diagnosis was included in the AI-driven differential diagnosis list [26]. Furthermore, a previous study of another automated medical history–taking system with a differential diagnosis generator—DIAANA, specializing in injury or disease of the musculoskeletal system—showed that the diagnostic accuracy was superior in the group in which physicians used the system compared to the group in which physicians did not use the system; this was a pilot randomized controlled trial conducted in a real clinical practice setting [24]. In contrast to the previous study that identified history-taking as the most common contributing factor of diagnostic errors [4], the breakdown analysis of the diagnostic errors in this study did not identify history-taking as the main contributing factor of these errors, indicating that the implementation of an automated history-taking system with diagnostic decision support could reduce the diagnostic errors associated with poor clinical history–taking.

In addition to making a high-quality document of medical history, an automated medical history–taking system with a differential diagnosis generator seems to have some advantages. First, this system can be integrated into routine diagnostic processes in clinical practice. Currently, one of the most important concerns in the diagnostic decision support system is its low usage rate. For example, in the case of Isabel, which is one of the most famous AI-driven diagnostic decision support systems that generates a differential diagnosis list based on entered information by physicians, a previous study showed that only 7.9% of participants who were given open access to Isabel reported using Isabel at least once a week, whereas the others never used it [46]. According to the other two studies, on average, Isabel was used for only 3 out of 4840 patients (0.06%) for 3 months [47], and the usage rate did not increase despite frequent reminders for clinicians to use Isabel on a regular basis [48]. Such low use of a diagnostic decision support system appeared to be caused by physicians who did not recognize the need for diagnostic support, relying on their own acumen to deliver the correct diagnosis [49]. However, diagnostic decision support systems should operate seamlessly in the background in the diagnostic process in clinical practice, regardless of whether the physicians need it or not [49]. An automated medical history–taking system with a differential diagnosis generator can address such an unmet need and may reduce diagnostic errors through routine support. Second, the use of a diagnostic decision support system at the early stage of the diagnostic process was reported to be more useful than its use at a later stage. To date, several studies have been conducted to evaluate the impact of the timing of using a diagnostic decision support system. According to their studies, physician diagnosis was associated with their first impression [50], and early use of diagnostic support systems before collecting information by physicians significantly improved the diagnostic accuracy [21-23]. These findings may support the positive effects of the implementation of an automated medical history–taking system with a differential diagnosis generator, which can provide diagnostic decision support before physicians collect information. Third, an automated medical history–taking system with a differential diagnosis generator can be used without additional time consumption. Another barrier for clinicians to use diagnostic decision support systems in routine clinical practice is time constraint, as previous studies have shown that using Isabel usually requires an additional 4 to 7 minutes per case [47,48]. On the other hand, an automated history-taking system with a differential diagnosis generator increased only 0.3 minutes of examination time per case in an internal medicine outpatient department [25]. Therefore, clinicians can use automated history-taking systems with

differential diagnosis generators without wasting additional time.

Furthermore, several limitations exist regarding the implementation of automated history-taking systems with differential diagnosis generators. First, at present, the accuracy of differential diagnosis lists of AI systems is not sufficiently high to believe the lists every time. A previous study reported that the prevalence of the correct diagnosis in the top 10 list of differential diagnoses from diagnostic decision support systems in clinical practice settings was around 50% [51]; similar to that study, the correct diagnosis appeared in only 50% of the top 10 lists of differential diagnoses from AI Monshin in this study. As an a priori incorrect diagnosis before a patient encounter can lead physicians to an incorrect final diagnosis [52], the relatively low accuracy of the differential diagnosis list from AI Monshin may prevent the positive effect of the implementation of an automated history-taking system with a differential diagnosis generator on the reduction of diagnostic errors. Although statistically insignificant, the incidence of diagnostic errors in cases where the correct diagnosis was included in the differential diagnosis list from the AI system was twice as high as that in cases where the correct diagnosis was not included in the list. However, among the 69 cases in which the final diagnosis was not included in the differential diagnosis list from the AI system, an incorrect diagnosis by a physician was observed in the differential diagnosis list from AI Monshin in only 4 cases (6%). In addition, a previous study showed that only 15% of physicians' diagnoses seemed to be associated with the differential diagnosis list from the AI system [53]. This indicates that the majority of diagnostic errors in this study were not related to the incorrect differential diagnosis list from the AI system. Second, the correct diagnosis in the automated differential diagnosis list cannot always be accepted as the most likely diagnosis by a physician. In 5 out of 69 cases (7%) where the correct diagnosis was included in the AI-generated differential diagnosis list, the correct diagnosis was not accepted as the initial diagnosis by the physician in this study. However, this type of error was also lower than that reported in previous studies (10.0% and 15.9%) [24,53]. Third, automated medical history–taking systems have had difficulty in precise history-taking for specific patients, such as older adult patients [54]. Indeed, in cases with diagnostic errors in this study, important past medical history was not imputed for 3 patients. However, such missed information seemed to be easily covered by physicians by checking the past medical history directly from the patient or reviewing the previous documentation.

Conclusions

The incidence of diagnostic errors seems to be reduced by the implementation of an automated medical history–taking system with a diagnostic decision support function in the outpatient department. Although the accuracy of the differential diagnosis list from AI Monshin remains low, the negative effects of incorrect differential diagnosis lists from AI systems on the diagnostic accuracy of physicians could be counteracted by the high-quality clinical history taken by AI systems. Therefore, in total, the implementation of an automated history-taking system with diagnostic decision support may have more beneficial impacts than negative effects on diagnostic safety in the outpatient department.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (KAKENHI) program (grant JP21K10355).

Authors' Contributions

RK and YH were responsible for conceptualization of the study and for developing the study methodology. YH conducted the formal analysis, was responsible for securing resources, performed data curation, and was responsible for project administration and funding acquisition. RK, YH, SS, YN, and SK conducted the study investigation. RK was responsible for writing and preparing the original draft of the manuscript. YH and TS were responsible for reviewing and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the histories written by AI Monshin in 16 diagnostic error cases. AI: artificial intelligence.
[DOCX File , 32 KB - medinform_v10i1e35225_app1.docx]

References

1. Balogh EP, Miller BT, Ball JR, editors. Overview of diagnostic error in health care. In: Improving Diagnosis in Health Care. Washington, DC: The National Academies Press; 2015:81-144.
2. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. BMJ Qual Saf 2017 Jun;26(6):484-494 [FREE Full text] [doi: 10.1136/bmjqs-2016-005401] [Medline: 27530239]
3. Cresswell KM, Panesar SS, Salvilla SA, Carson-Stevens A, Larizgoitia I, Donaldson LJ, World Health Organization's (WHO) Safer Primary Care Expert Working Group. Global research priorities to better understand the burden of iatrogenic

harm in primary care: An international Delphi exercise. PLoS Med 2013 Nov;10(11):e1001554 [FREE Full text] [doi: 10.1371/journal.pmed.1001554] [Medline: 24260028]
4. Singh H, Giardina TD, Meyer AND, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. JAMA Intern Med 2013 Mar 25;173(6):418-425 [FREE Full text] [doi: 10.1001/jamainternmed.2013.2777] [Medline: 23440149]
5. Fernholm R, Pukk Härenstam K, Wachtler C, Nilsson GH, Holzmann MJ, Carlsson AC. Diagnostic errors reported in primary healthcare and emergency departments: A retrospective and descriptive cohort study of 4830 reported cases of preventable harm in Sweden. Eur J Gen Pract 2019 Jul;25(3):128-135 [FREE Full text] [doi: 10.1080/13814788.2019.1625886] [Medline: 31257959]
6. Gupta A, Snyder A, Kachalia A, Flanders S, Saint S, Chopra V. Malpractice claims related to diagnostic errors in the hospital. BMJ Qual Saf 2017 Aug 09;27(1):53-60. [doi: 10.1136/bmjqs-2017-006774] [Medline: 28794243]
7. Watari T. Malpractice claims of internal medicine involving diagnostic and system errors in Japan. Intern Med 2021 Sep 15;60(18):2919-2925 [FREE Full text] [doi: 10.2169/internalmedicine.6652-20] [Medline: 33776001]
8. Watari T, Tokuda Y, Mitsuhashi S, Otuki K, Kono K, Nagai N, et al. Factors and impact of physicians' diagnostic errors in malpractice claims in Japan. PLoS One 2020;15(8):e0237145 [FREE Full text] [doi: 10.1371/journal.pone.0237145] [Medline: 32745150]
9. Saber Tehrani AS, Lee H, Mathews SC, Shore A, Makary MA, Pronovost PJ, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: An analysis from the National Practitioner Data Bank. BMJ Qual Saf 2013 Aug;22(8):672-680. [doi: 10.1136/bmjqs-2012-001550] [Medline: 23610443]
10. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, et al. Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims. Ann Intern Med 2006 Oct 03;145(7):488-496. [Medline: 17015866]
11. Bastakoti M, Muhailan M, Nassar A, Sallam T, Desale S, Fouda R, et al. Discrepancy between emergency department admission diagnosis and hospital discharge diagnosis and its impact on length of stay, up-triage to the intensive care unit, and mortality. Diagnosis (Berl) 2021 Jul 05. [doi: 10.1515/dx-2021-0001] [Medline: 34225399]
12. Hautz WE, Kämmer JE, Hautz SC, Sauter TC, Zwaan L, Exadaktylos AK, et al. Diagnostic error increases mortality and length of hospital stay in patients presenting through the emergency room. Scand J Trauma Resusc Emerg Med 2019 May 08;27(1):54 [FREE Full text] [doi: 10.1186/s13049-019-0629-z] [Medline: 31068188]
13. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014 Sep;23(9):727-731 [FREE Full text] [doi: 10.1136/bmjqs-2013-002627] [Medline: 24742777]
14. Singh H, Giardina TD, Forjuoh SN, Reis MD, Kosmach S, Khan MM, et al. Electronic health record-based surveillance of diagnostic errors in primary care. BMJ Qual Saf 2012 Feb;21(2):93-100 [FREE Full text] [doi: 10.1136/bmjqs-2011-000304] [Medline: 21997348]
15. Goyder CR, Jones CHD, Heneghan CJ, Thompson MJ. Missed opportunities for diagnosis: Lessons learned from diagnostic errors in primary care. Br J Gen Pract 2015 Dec;65(641):e838-e844 [FREE Full text] [doi: 10.3399/bjgp15X687889] [Medline: 26622037]
16. Slack WV, Hicks GP, Reed CE, Van Cura LJ. A computer-based medical-history system. N Engl J Med 1966 Jan 27;274(4):194-198. [doi: 10.1056/NEJM196601272740406] [Medline: 5902618]
17. Mayne JG, Weksel W, Sholtz PN. Toward automating the medical history. Mayo Clin Proc 1968 Jan;43(1):1-25. [Medline: 5635452]
18. De Dombal FT, Leaper DJ, Horrocks JC, Staniland JR, McCann AP. Human and computer-aided diagnosis of abdominal pain: Further report with emphasis on performance of clinicians. Br Med J 1974 Mar 02;1(5904):376-380 [FREE Full text] [doi: 10.1136/bmj.1.5904.376] [Medline: 4594585]
19. Zakim D, Brandberg H, El Amrani S, Hultgren A, Stathakarou N, Nifakos S, et al. Computerized history-taking improves data quality for clinical decision-making: Comparison of EHR and computer-acquired history data in patients with chest pain. PLoS One 2021;16(9):e0257677 [FREE Full text] [doi: 10.1371/journal.pone.0257677] [Medline: 34570811]
20. Almario CV, Chey W, Kaung A, Whitman C, Fuller G, Reid M, et al. Computer-generated vs physician-documented history of present illness (HPI): Results of a blinded comparison. Am J Gastroenterol 2015 Jan;110(1):170-179 [FREE Full text] [doi: 10.1038/ajg.2014.356] [Medline: 25461620]
21. Kostopoulou O, Lionis C, Angelaki A, Ayis S, Durbaba S, Delaney BC. Early diagnostic suggestions improve accuracy of family physicians: A randomized controlled trial in Greece. Fam Pract 2015 Jun;32(3):323-328. [doi: 10.1093/fampra/cmv012] [Medline: 25800247]
22. Kostopoulou O, Rosen A, Round T, Wright E, Douiri A, Delaney B. Early diagnostic suggestions improve accuracy of GPs: A randomised controlled trial using computer-simulated patients. Br J Gen Pract 2015 Jan;65(630):e49-e54 [FREE Full text] [doi: 10.3399/bjgp15X683161] [Medline: 25548316]
23. Sibbald M, Monteiro S, Sherbino J, LoGiudice A, Friedman C, Norman G. Should electronic differential diagnosis support be used early or late in the diagnostic process? A multicentre experimental study of Isabel. BMJ Qual Saf 2021 Oct 05 [FREE Full text] [doi: 10.1136/bmjqs-2021-013493] [Medline: 34611040]

24. Schwitzguebel AJ, Jeckelmann C, Gavinio R, Levallois C, Benaïm C, Spechbach H. Differential diagnosis assessment in ambulatory care with an automated medical history-taking device: Pilot randomized controlled trial. JMIR Med Inform 2019 Nov 04;7(4):e14044 [FREE Full text] [doi: 10.2196/14044] [Medline: 31682590]
25. Harada Y, Shimizu T. Impact of a commercial artificial intelligence-driven patient self-assessment solution on waiting times at general internal medicine outpatient departments: Retrospective study. JMIR Med Inform 2020 Aug 31;8(8):e21056 [FREE Full text] [doi: 10.2196/21056] [Medline: 32865504]
26. Harada Y, Katsukura S, Kawamura R, Shimizu T. Efficacy of artificial-intelligence-driven differential-diagnosis list on the diagnostic accuracy of physicians: An open-label randomized controlled study. Int J Environ Res Public Health 2021 Feb 21;18(4):2086 [FREE Full text] [doi: 10.3390/ijerph18042086] [Medline: 33669930]
27. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: An online mobile survey. J Med Internet Res 2019 Mar 25;21(3):e12422 [FREE Full text] [doi: 10.2196/12422] [Medline: 30907742]
28. Aaronson E, Jansson P, Wittbold K, Flavin S, Borczuk P. Unscheduled return visits to the emergency department with ICU admission: A trigger tool for diagnostic error. Am J Emerg Med 2020 Aug;38(8):1584-1587. [doi: 10.1016/j.ajem.2019.158430] [Medline: 31699427]
29. Singh H, Khanna A, Spitzmueller C, Meyer AND. Recommendations for using the Revised Safer Dx Instrument to help measure and improve diagnostic safety. Diagnosis (Berl) 2019 Nov 26;6(4):315-323. [doi: 10.1515/dx-2019-0012] [Medline: 31287795]
30. Al-Mutairi A, Meyer AND, Thomas EJ, Etchegaray JM, Roy KM, Davalos MC, et al. Accuracy of the Safer Dx Instrument to identify diagnostic errors in primary care. J Gen Intern Med 2016 Jun;31(6):602-608 [FREE Full text] [doi: 10.1007/s11606-016-3601-x] [Medline: 26902245]
31. Davalos MC, Samuels K, Meyer AND, Thammasitboon S, Sur M, Roy K, et al. Finding diagnostic errors in children admitted to the PICU. Pediatr Crit Care Med 2017 Mar;18(3):265-271. [doi: 10.1097/PCC.0000000000001059] [Medline: 28125548]
32. Searns JB, Williams MC, MacBrayne CE, Wirtz AL, Leonard JE, Boguniewicz J, et al. Handshake antimicrobial stewardship as a model to recognize and prevent diagnostic errors. Diagnosis (Berl) 2021 Aug 26;8(3):347-352. [doi: 10.1515/dx-2020-0032] [Medline: 33112779]
33. Bergl PA, Taneja A, El-Kareh R, Singh H, Nanchal RS. Frequency, risk factors, causes, and consequences of diagnostic errors in critically ill medical patients: A retrospective cohort study. Crit Care Med 2019 Nov;47(11):e902-e910. [doi: 10.1097/CCM.0000000000003976] [Medline: 31524644]
34. Cifra CL, Ten Eyck P, Dawson JD, Reisinger HS, Singh H, Herwaldt LA. Factors associated with diagnostic error on admission to a PICU: A pilot study. Pediatr Crit Care Med 2020 May;21(5):e311-e315 [FREE Full text] [doi: 10.1097/PCC.0000000000002257] [Medline: 32097247]
35. Liberman AL, Bakradze E, Mchugh DC, Esenwa CC, Lipton RB. Assessing diagnostic error in cerebral venous thrombosis via detailed chart review. Diagnosis (Berl) 2019 Nov 26;6(4):361-367. [doi: 10.1515/dx-2019-0003] [Medline: 31271550]
36. Fletcher TL, Helm A, Vaghani V, Kunik ME, Stanley MA, Singh H. Identifying psychiatric diagnostic errors with the Safer Dx Instrument. Int J Qual Health Care 2020 Jul 20;32(6):405-411. [doi: 10.1093/intqhc/mzaa066] [Medline: 32671387]
37. Avelino-Silva TJ, Steinman MA. Diagnostic discrepancies between emergency department admissions and hospital discharges among older adults: Secondary analysis on a population-based survey. Sao Paulo Med J 2020;138(5):359-367 [FREE Full text] [doi: 10.1590/1516-3180.0471.R1.05032020] [Medline: 32935740]
38. Skinner TR, Scott IA, Martin JH. Diagnostic errors in older patients: A systematic review of incidence and potential causes in seven prevalent diseases. Int J Gen Med 2016;9:137-146 [FREE Full text] [doi: 10.2147/IJGM.S96741] [Medline: 27284262]
39. Horberg MA, Nassery N, Rubenstein KB, Certa JM, Shamim EA, Rothman R, et al. Rate of sepsis hospitalizations after misdiagnosis in adult emergency department patients: A look-forward analysis with administrative claims data using Symptom-Disease Pair Analysis of Diagnostic Error (SPADE) methodology in an integrated health system. Diagnosis (Berl) 2021 Nov 25;8(4):479-488. [doi: 10.1515/dx-2020-0145] [Medline: 33894108]
40. Cheraghi-Sohi S, Holland F, Singh H, Danczak A, Esmail A, Morris RL, et al. Incidence, origins and avoidable harm of missed opportunities in diagnosis: Longitudinal patient record review in 21 English general practices. BMJ Qual Saf 2021 Dec;30(12):977-985 [FREE Full text] [doi: 10.1136/bmjqs-2020-012594] [Medline: 34127547]
41. Lim G, Seow E, Koh G, Tan D, Wong H. Study on the discrepancies between the admitting diagnoses from the emergency department and the discharge diagnoses. Hong Kong J Emerg Med 2017 Dec 11;9(2):78-82. [doi: 10.1177/102490790200900203]
42. Adnan M, Baharuddin K, Mohammad J, Yazid B, Ahmad M. A study on the diagnostic discrepancy between admission and discharge in Hospital Universiti Sains Malaysia. Malays J Med Health Sci 2021 Jan;17(1):105-110 [FREE Full text]
43. Fatima S, Shamim S, Butt AS, Awan S, Riffat S, Tariq M. The discrepancy between admission and discharge diagnoses: Underlying factors and potential clinical outcomes in a low socioeconomic country. PLoS One 2021;16(6):e0253316. [doi: 10.1371/journal.pone.0253316] [Medline: 34129648]

44. Peng A, Rohacek M, Ackermann S, Ilsemann-Karakoumis J, Ghanim L, Messmer AS, et al. The proportion of correct diagnoses is low in emergency patients with nonspecific complaints presenting to the emergency department. Swiss Med Wkly 2015;145:w14121 [FREE Full text] [doi: 10.4414/smw.2015.14121] [Medline: 25741894]
45. Eze-Nliam C, Cain K, Bond K, Forlenza K, Jankowski R, Magyar-Russell G, et al. Discrepancies between the medical record and the reports of patients with acute coronary syndrome regarding important aspects of the medical history. BMC Health Serv Res 2012 Mar 26;12:78 [FREE Full text] [doi: 10.1186/1472-6963-12-78] [Medline: 22448755]
46. Ramnarayan P, Roberts GC, Coren M, Nanduri V, Tomlinson A, Taylor PM, et al. Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: A quasi-experimental study. BMC Med Inform Decis Mak 2006 Apr 28;6:22 [FREE Full text] [doi: 10.1186/1472-6947-6-22] [Medline: 16646956]
47. Henderson EJ, Rubin GP. The utility of an online diagnostic decision support system (Isabel) in general practice: A process evaluation. JRSM Short Rep 2013 May;4(5):31 [FREE Full text] [doi: 10.1177/2042533313476691] [Medline: 23772310]
48. Cheraghi-Sohi S, Alam R, Hann M, Esmail A, Campbell S, Riches N. Assessing the utility of a differential diagnostic generator in UK general practice: A feasibility study. Diagnosis (Berl) 2021 Feb 23;8(1):91-99. [doi: 10.1515/dx-2019-0033] [Medline: 32083441]
49. Delaney BC, Kostopoulou O. Decision support for diagnosis should become routine in 21st century primary care. Br J Gen Pract 2017 Nov;67(664):494-495 [FREE Full text] [doi: 10.3399/bjgp17X693185] [Medline: 29074677]
50. Kostopoulou O, Sirota M, Round T, Samaranayaka S, Delaney BC. The role of physicians' first impressions in the diagnosis of possible cancers without alarm symptoms. Med Decis Making 2017 Jan;37(1):9-16 [FREE Full text] [doi: 10.1177/0272989X16644563] [Medline: 27112933]
51. Graber MA, VanScoy D. How well does decision support software perform in the emergency department? Emerg Med J 2003 Sep;20(5):426-428 [FREE Full text] [doi: 10.1136/emj.20.5.426] [Medline: 12954680]
52. Meyer FML, Filipovic MG, Balestra GM, Tisljar K, Sellmann T, Marsch S. Diagnostic errors induced by a wrong a priori diagnosis: A prospective randomized simulator-based trial. J Clin Med 2021 Feb 18;10(4):826 [FREE Full text] [doi: 10.3390/jcm10040826] [Medline: 33670489]
53. Harada Y, Katsukura S, Kawamura R, Shimizu T. Effects of a differential diagnosis list of artificial intelligence on differential diagnoses by physicians: An exploratory analysis of data from a randomized controlled study. Int J Environ Res Public Health 2021 May 23;18(11):5562 [FREE Full text] [doi: 10.3390/ijerph18115562] [Medline: 34070958]
54. Brandberg H, Sundberg CJ, Spaak J, Koch S, Zakim D, Kahan T. Use of self-reported computerized medical history taking for acute chest pain in the emergency department - The Clinical Expert Operating System Chest Pain Danderyd Study (CLEOS-CPDS): Prospective cohort study. J Med Internet Res 2021 Apr 27;23(4):e25493 [FREE Full text] [doi: 10.2196/25493] [Medline: 33904821]

Abbreviations

AI: artificial intelligence
KAKENHI: Grants-in-Aid for Scientific Research
OR: odds ratio

Edited by C Lovis; submitted 26.11.21; peer-reviewed by D Zakim, J Carter; comments to author 03.12.21; revised version received 11.12.21; accepted 02.01.22; published 27.01.22.

Please cite as:
Kawamura R, Harada Y, Sugimoto S, Nagase Y, Katsukura S, Shimizu T
Incidence of Diagnostic Errors Among Unexpectedly Hospitalized Patients Using an Automated Medical History–Taking System With a Differential Diagnosis Generator: Retrospective Observational Study
JMIR Med Inform 2022;10(1):e35225
URL: https://medinform.jmir.org/2022/1/e35225
doi:10.2196/35225
PMID:35084347

Original Paper

Using Electronic Health Records for Personalized Dosing of Intravenous Vancomycin in Critically Ill Neonates: Model and Web-Based Interface Development Study

Ka Ho Matthew Hui[1], BPharm, PhD; Hugh Simon Lam[2], MD; Cheuk Hin Twinny Chow[1*], BPharm, MCP; Yuen Shun Janice Li[1*], BPharm, MCP; Pok Him Tom Leung[1*], BPharm; Long Yin Brian Chan[1*], BPharm; Chui Ping Lee[1,3], PharmD; Celeste Lom Ying Ewig[1,3], PharmD; Yin Ting Cheung[1], BSc (Hons), PhD; Tai Ning Teddy Lam[1,3], PharmD, PhD

[1]School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[2]Department of Paediatrics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[3]Department of Pharmacy, Prince of Wales Hospital, Hospital Authority, Hong Kong, Hong Kong

[*]these authors contributed equally

Corresponding Author:
Tai Ning Teddy Lam, PharmD, PhD
School of Pharmacy
Faculty of Medicine
The Chinese University of Hong Kong
8/F, Lo Kwee-Seong Integrated Biomedical Sciences Building, Area 39
Shatin, New Territories
Hong Kong
Hong Kong
Phone: 852 39436827
Fax: 852 26035295
Email: teddylam@cuhk.edu.hk

Abstract

Background: Intravenous (IV) vancomycin is used in the treatment of severe infection in neonates. However, its efficacy is compromised by elevated risks of acute kidney injury. The risk is even higher among neonates admitted to the neonatal intensive care unit (NICU), in whom the pharmacokinetics of vancomycin vary widely. Therapeutic drug monitoring is an integral part of vancomycin treatment to balance efficacy against toxicity. It involves individual dose adjustments based on the observed serum vancomycin concentration (VC_s). However, the existing trough-based approach shows poor evidence for clinical benefits. The updated clinical practice guideline recommends population pharmacokinetic (popPK) model–based approaches, targeting area under curve, preferably through the Bayesian approach. Since Bayesian methods cannot be performed manually and require specialized computer programs, there is a need to provide clinicians with a user-friendly interface to facilitate accurate personalized dosing recommendations for vancomycin in critically ill neonates.

Objective: We used medical data from electronic health records (EHRs) to develop a popPK model and subsequently build a web-based interface to perform model-based individual dose optimization of IV vancomycin for NICU patients in local medical institutions.

Methods: Medical data of subjects prescribed IV vancomycin in the NICUs of Prince of Wales Hospital and Queen Elizabeth Hospital in Hong Kong were extracted from EHRs, namely the Clinical Information System, In-Patient Medication Order Entry, and electronic Patient Record. Patient demographics, such as body weight and postmenstrual age (PMA), serum creatinine (SCr), vancomycin administration records, and VC_s were collected. The popPK model employed a 2-compartment infusion model. Various covariate models were tested against body weight, PMA, and SCr, and were evaluated for the best goodness of fit. A previously published web-based dosing interface was adapted to develop the interface in this study.

Results: The final data set included EHR data extracted from 207 subjects, with a total of 689 VC_s measurements. The final model chosen explained 82% of the variability in vancomycin clearance. All parameter estimates were within the bootstrapping CIs. Predictive plots, residual plots, and visual predictive checks demonstrated good model predictability. Model approximations

showed that the model-based Bayesian approach consistently promoted a probability of target attainment (PTA) above 75% for all subjects, while only half of the subjects could achieve a PTA over 50% with the trough-based approach. The dosing interface was developed with the capability to optimize individual doses with the model-based empirical or Bayesian approach.

Conclusions: Using EHRs, a satisfactory popPK model was verified and adopted to develop a web-based individual dose optimization interface. The interface is expected to improve treatment outcomes of IV vancomycin for severe infections among critically ill neonates. This study provides the foundation for a cohort study to demonstrate the utility of the new approach compared with previous dosing methods.

(JMIR Med Inform 2022;10(1):e29458) doi:10.2196/29458

KEYWORDS

digital health; web-based user interface; personalized medicine; dose individualization; therapeutic drug monitoring; Bayesian estimation; antibiotics; vancomycin; infectious disease; neonate

Introduction

Intravenous Vancomycin

Intravenous (IV) vancomycin has long been the first-line treatment for severe bacterial infections, especially in cases involving *Staphylococci* species [1]. Despite its well-established efficacy, vancomycin has a narrow therapeutic index and is commonly associated with acute kidney injury (AKI), especially at high levels of exposure [2]. It was shown that even small acute increases in serum creatinine (SCr) could be detrimental to long-term survival in critically ill patients [3]. Therapeutic drug monitoring (TDM) for vancomycin is a recommended practice to balance efficacy and the risk of AKI. This involves the monitoring of the systemic serum vancomycin concentration (VC_s) over time after drug administration and subsequent adjustments of vancomycin dosage as necessary.

Pharmacokinetics of Vancomycin

Vancomycin is eliminated from the systemic circulation primarily through glomerular filtration in the kidneys. Thus, glomerular filtration rate (GFR) is closely correlated with vancomycin clearance (CL), which is the main factor affecting VC_s [4]. Since GFR is clinically estimated by creatinine clearance, the major determinants of creatinine clearance, including body size and SCr, are among the major covariates of CL among patients from all age groups [5,6]. To improve the prediction of VC_s, the pharmacokinetics of vancomycin has been widely studied to understand the mathematical relationship between CL and these covariates [7-11].

Vulnerability of Critically Ill Neonates Requiring Vancomycin

In the neonatal intensive care unit (NICU), the pharmacokinetics of vancomycin among neonates is highly variable due to dynamic patient conditions and interventions [11]. Moreover, for neonates, it is necessary to account for the maturation of renal function, a process unique to the neonatal population that occurs over the first weeks to months postpartum and is associated with postmenstrual age (PMA) [12]. These conditions put NICU patients at a higher risk for suboptimal therapeutic effects of vancomycin and AKI, making accurate TDM of vancomycin indispensable in this population.

TDM of Vancomycin

A steady-state area under the curve of the VC_s-time profile (AUC) over 24 hours (AUC_{24}) to minimum inhibitory concentration (MIC) ratio (AUC_{24}/MIC) of ≥400 hours has been advocated as the primary predictor of vancomycin efficacy [13]. Nevertheless, since AUC estimation requires measuring multiple VC_s values, which is often impractical in the clinical setting, the American Society of Health-System Pharmacists, the Infectious Diseases Society of America, and the Society of Infectious Diseases Pharmacists published a consensus report in 2009 recommending the steady-state trough VC_s ($VC_{s,ss,trough}$) as a surrogate marker for the AUC target (assuming MIC at 1 mg/L) [13]. However, data on the efficacy and safety profile with this trough-based approach are lacking [14]. On the other hand, there is further evidence supporting AUC_{24}/MIC as the pharmacokinetic target. The requirement of multiple VC_s measurements could also be resolved by employing the Bayesian approach as supported by recent research [15,16].

In response, the guideline was updated in 2020 jointly by the 3 societies publishing the 2009 report, together with the Pediatric Infectious Diseases Society, giving new recommendations. First, $VC_{s,ss,trough}$ is no longer recommended as a pharmacokinetic target; dose optimization should instead target an AUC_{24}/MIC of 400 to 600 hours. Second, the preferred method to estimate individual AUC is to apply Bayesian estimation using 1 trough VC_s ($VC_{s,trough}$, presteady-state or steady-state trough) and preferably 1 peak VC_s ($VC_{s,peak}$, presteady-state or steady-state peak), based on a population pharmacokinetic (popPK) model for vancomycin. Third, a less preferred method to calculate individual AUC is to use the first-order equations on a set of measured $VC_{s,ss,trough}$ and steady-state peak VC_s ($VC_{s,ss,peak}$) values [14]. Recommendations for initial dosing were also revised. A popPK model–based estimation of individual AUC is preferred over using a universal weight-based dosing scheme [17]. As AUC becomes the basis of dose optimization, ensuring a reliable approach for AUC estimation is a prerequisite of dose optimality. Table 1 summarizes the approaches to AUC estimation and hence dose optimization used in this text.

Table 1. Summary of approaches to vancomycin dosing.

Dosing approach	Weight-based	Empirical dosing with popPK[a] parameter estimates	Steady-state trough target	Estimation of AUC[b] by steady-state peak and trough	Model-based Bayesian optimization
Which dose to guide?	Initial dose	Initial dose	Maintenance dose	Maintenance dose	Maintenance dose
When to use?	Before the first dose	Before the first dose	When VC_s[c] measurement is available	When VC_s measurement is available	When VC_s measurement is available
Required VC_s measurements	N/A[d]	N/A	$VC_{s,ss,trough}$[e]	$VC_{s,ss,trough}$ + $VC_{s,ss,peak}$[f]	$VC_{s,trough}$[g] (+$VC_{s,peak}$[h])[i]
PK[j] target	N/A	AUC	$VC_{s,ss,trough}$	AUC	AUC
popPK model-based?	N/A	Yes[k]	N/A	N/A	Yes[k]
Bayesian estimation required?	N/A	N/A	N/A	N/A	Yes
Recommended?	Yes	Yes[l]	No longer	Yes, less preferred	Yes, preferred

[a]popPK: population pharmacokinetic.

[b]AUC: steady-state area under the curve of the serum vancomycin concentration-time profile.

[c]VC_s: serum vancomycin concentration.

[d]N/A: not applicable.

[e]$VC_{s,ss,trough}$: steady-state trough serum vancomycin concentration.

[f]$VC_{s,ss,peak}$: steady-state peak serum vancomycin concentration.

[g]$VC_{s,trough}$: trough serum vancomycin concentration (presteady-state or steady-state trough).

[h]$VC_{s,peak}$: peak serum vancomycin concentration (presteady-state or steady-state peak).

[i]Preferrably with $VC_{s,peak}$.

[j]PK: pharmacokinetic.

[k]The 2 approaches are collectively called the model-based approaches.

[l]Potentially better compared with the weight-based approach.

Multifaceted Roles of Digital Health in the TDM of Vancomycin in the NICU Population

The rapid development in digital health has made this study and the proposed clinical improvements possible in multiple ways. They are elaborated in the following paragraphs.

To keep up with the current standard of treatment and given the large variability in NICU patients, separate popPK analyses for vancomycin are required for the local NICU population [18]. However, prospective data collection is often costly and burdensome in the clinical environment, while the unstructured collection of retrospective data is prone to errors. Fortunately, as digital records are becoming vital on the clinical frontline, electronic health records (EHRs) now present extractable information for data analyses [19]. It is now feasible to consolidate data retrieved from multiple EHR sources to reconcile a data set suitable for popPK analyses [20].

Establishing a popPK model is the first step to upgrade the TDM practice for IV vancomycin in the local NICU population. To maximize the utility of the popPK model, it is necessary to enable Bayesian estimation for accurate estimations of individual AUC [14]. Unlike conventional strategies to individual dose optimization by equations and nomograms, which can be carried out manually, Bayesian estimation requires numerical approximation processes that can only be performed digitally using computers.

Putting popPK model–based Bayesian estimation into clinical practice is difficult because most clinicians are not experts in this area. To tackle this, a fully automated web-based interface incorporating a popPK model, a numerical approximation solution to Bayesian estimation, and algorithms for dose optimization would be an ideal tool for clinical use. In contrast with a client-based interface, a web-based interface (1) allows remote access with various browser-enabled devices, including clinical computer workstations, tablets, and smartphones; (2) saves installation issues; and (3) is easier to maintain. Such an interface is designed to guide and validate necessary inputs from clinicians, followed by suggestions of dosing regimens, which are expected to help clinicians decide the optimal treatment plan that can enhance clinical outcomes.

Summary and Study Objectives

The use of a model-based dosing interface is in its pilot stage in Hong Kong. Neonatal vancomycin is among the first drugs being investigated. Experiences gained in this study are expected to improve IV vancomycin treatment significantly and, perhaps more importantly, lay the foundation for the extraction of popPK data from EHRs and web-based dose optimization interfaces for other drugs with narrow therapeutic indices. In support of its implementation, this study was conducted in local medical

institutions to develop the popPK model of vancomycin for NICU patients using real-world data from EHR resources. Besides, a previously reported framework of a web-based interface performing Bayesian estimation and individual dose optimization for the use of high-dose methotrexate in local institutions will be adopted to create the dosing interface for neonatal IV vancomycin.

Methods

Study Population, EHR Use, and Data Preprocessing

The study data set consists of all Chinese patients within 1 year of postnatal age (PNA) admitted to the NICUs of Prince of Wales Hospital and Queen Elizabeth Hospital in Hong Kong between January 1, 2016, and December 31, 2017. Each potential subject had to be prescribed IV vancomycin, and have at least one VC_s measurement and one SCr measurement in order to be eligible. Subjects with major congenital heart diseases were excluded. Subjects with vancomycin initiated within 7 days of birth were also excluded due to the variable effects of maternal creatinine on the estimation of neonatal renal function. Eligible subjects were identified through the Clinical Data Analysis and Reporting System (CDARS), a database developed and maintained by the Hong Kong Hospital Authority for audit and research purposes. Data of selected subjects were then collected from several in-house EHR platforms, namely the Clinical Information System (CIS), In-Patient Medication Order Entry (IPMOE), electronic Patient Record (ePR), and, whenever necessary, original copies of medical charts. Ethical approval was obtained from the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (reference number: 2018.094) and Kowloon Central Cluster/Kowloon East Cluster Research Ethics Committee (reference number: KC/KE-18-0096/ER-1) for data collection. Parental consent was not required due to the anonymized and retrospective nature of data collection.

Constant data items collected were sex, birth weight, gestation age, and date of birth. Time-dependent measurements included body weight, VC_s, and SCr. Dosing records were collected in terms of the dose administered and the infusion rate (assuming constant rate). The date and time tags of all time-dependent measurements and dosing records were also collected. Dosing records were available from IPMOE, while other data items were collected from CIS and ePR.

All VC_s records collected over 7 days after the start of the last infusion of vancomycin were removed. All subjects started on vancomycin within the first 7 days of birth were also removed. At each unique time tag of each subject, PMA was calculated as the time difference between the tag and the estimated first day of the last menstrual period of the subject's mother. SCr was imputed as the previous or next available value, whichever was closer in time. Body weight was imputed by linearly interpolating and extrapolating available values. All VC_s values measured during infusion were removed. SCr values below the lower limit of quantification (LLOQ) (ie, 15 µmol/L) were replaced by 7.5 µmol/L. VC_s values measured had an LLOQ of 1 mg/L, and records below the LLOQ (below the limit of quantification [BLQ]) were flagged.

Model Structure and Parameterization

A popPK model adopts the structure of a nonlinear mixed-effect model [21,22]. A 2-compartment infusion model with first-order elimination was applied, for which the pharmacokinetic parameters CL, central volume (V_c), intercompartmental clearance (Q), and peripheral volume (V_p) of vancomycin were defined [9]. The between-subject variability (BSV) in CL and the between-occasion variability (BOV, variability in CL in the same subject between episodes) were expressed in terms of the coefficient of variance (CV) (ie, CVCL and $CVCL_{BOV}$, respectively). Both CVCL and $CVCL_{BOV}$ were assumed to follow the log-normal distribution [23]. Residual unexplained variability was described by a combined proportional-additive error model [24].

In building the pharmacokinetic parameter model, allometric scaling was applied to describe the association of CL, V_c, Q, and V_p against body weight using the power function, with fixed exponents of 0.75 and 1, respectively [25]. This was tested against freely estimated exponents (one for CL and Q, and another for V_c and V_p) using the likelihood ratio test. The maturation of renal function was described as a function of PMA and tested against the linear, exponential, first-order, and Hill functions [26-28]. The function that returned the best goodness of fit was chosen. The renal function with respect to SCr was described using the power model [29].

Parameter Estimation and Model Evaluation

Parameter estimation was executed with NONMEM version 7.4 (Icon plc) using first-order conditional estimation with interaction [30]. BLQ data were handled using the M3 method [31]. Perl-speaks-NONMEM was used to coordinate NONMEM execution [32]. Residual plots, predictive plots, and a prediction-corrected visual predictive check (pcVPC) were generated [33]. Bootstrapping using 1000 resampled data sets was performed to assess the stability of parameter estimates [34]. R and the R package ggplot2 were used for graphics generation [35,36].

Dose Individualization

Model-based approaches to dose optimization rely on the estimation of pharmacokinetic parameters for a subject based on the verified popPK model as described above. The set of pharmacokinetic parameter estimates are then used to approximate the AUC distributions at different doses, such that the dose at which the probability of attaining an AUC_{24}/MIC of 400 to 600 hours (probability of target attainment [PTA]) is maximized (ie, the optimal dose) can be identified by numerical approximation. Practically, the empirical approach helps decide the initial dose, while the Bayesian approach informs dose adjustments afterwards (see Table 1).

Web-Based Dosing Interface

The web-based dosing interface in this study is designed to perform the model-based approach to dose optimization. The framework of the interface was replicated from that reported in

a previous study for the dose adjustments of single-dose high-dose methotrexate in the pediatric population [37]. On performing the Bayesian approach, the interface demonstrated the ability to generate individual estimates of AUC identical to and more efficiently than NONMEM. The interface was modified to adapt to the popPK model for IV vancomycin estimated and verified in this study, enable empirical dose (the first dose) suggestion, and allow dose optimization at various dosing intervals.

Results

Data Set Management

One VC_s value was measured 7 days after the start of the last infusion and thus removed. Forty-five subjects with PNA <7 days when vancomycin was first started were also removed. Data extraction from the EHRs and data exclusion resulted in a final data set consisting of 207 patients and a total of 689 VC_s measurements. The demographics are detailed in Table 2. The time profile of observed VC_s values is shown in Figure 1.

Table 2. Demographic and data characteristics of the final data set (N=207).

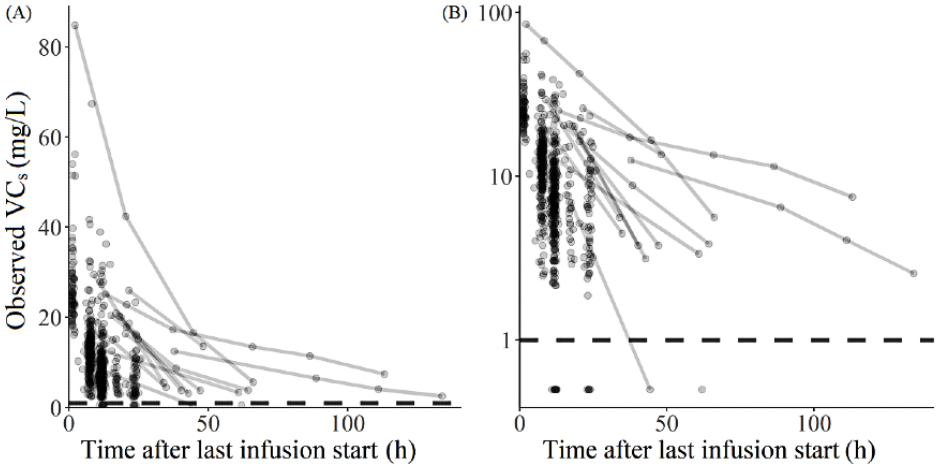
Characteristic	Value
Site, n (%)	
Prince of Wales Hospital	156 (75.4)
Queen Elizabeth Hospital	51 (24.6)
Sex, n (%)	
Male	112 (54.1)
Female	95 (45.9)
Gestation age (weeks), median±IQR (min-max)	30.1±6.9 (24.1-41.3)
Postnatal age at first dose (days), median±IQR (min-max)	17±14 (7-114)
Postmenstrual age at first dose (weeks), median±IQR (min-max)	33.7±7.3 (25.7-53.3)
Birth weight (kg), median±IQR (min-max)	1.32±0.89 (0.44-4.14)
Median body weight (kg), median±IQR (min-max)	1.68±1.13 (0.47-7.36)
Dose infused (mg/kg), median±IQR (min-max)	14±3 (5-31)
SCr[a] (µmol/L), median±IQR (min-max)	42±34 (15-252) (plus 12 BLQ[b] measures of SCr)
Number of VC_s[c] measurements by subject, n (%)	
1	63 (30.4)
2	43 (20.8)
3	26 (12.6)
4	27 (13.0)
5	21 (10.1)
6-8	16 (7.7)
10-21	11 (5.3)
Measured VC_s (mg/L), median±IQR (min-max)	9.9±9.4 (1.9-84.8) (plus 16 BLQ measures of VC_s)
Number of episodes (after combining) by subject, n (%)	
1	131 (63.3)
2	48 (23.2)
3	14 (6.8)
4	5 (2.4)
5	5 (2.4)
6	4 (1.9)

[a]SCr: serum creatinine concentration.

[b]BLQ: below limit of quantification.

[c]VC_s: serum vancomycin concentration.

Figure 1. Profile of the observed VC_s in the final data set. The graphs show the profile of the observed VC_s in linear (A) and logarithmic (B) scales. Observed VC_s values after the same last dose in the same subject are joined with a solid line. The dashed horizontal line denotes the lower limit of quantification, below which all measured VC_s values are displayed at 0.5 mg/L. VC_s: serum vancomycin concentration.



Model Comparison

In the final model, a fixed exponent is used for the power functions of the body weight effect and a Hill function is used to describe the PMA-CL relationship. The final model has a minimum objective function value (OFV) of 2329.272. Allowing freely estimated exponents for body weight functions on pharmacokinetic parameters only led to statistically insignificant improvements in goodness of fit (χ^2 value approximates change in OFV [dOFV]=−5.583, *P*=.06, at a degree of freedom of 2). Replacing the Hill function with a linear, exponential, or first-order function resulted in worsened goodness of fit (dOFV=+58.440, +83.391, and +140.645, respectively). None of these alternative models led to significantly better goodness of fit.

Final Model Parameter Estimates and Evaluation

Parameter estimates for the final model are shown in Table 3. They are very close to the bootstrap means and well within the bootstrap CI and have a condition number of 226 (which is within the usual reference limit of 1000). Accounting for both CVCL and $CVCL_{BOV}$, the effects of body weight, PMA, and SCr alone, and all combined explained 43%, 63%, 54%, and 82% of the variability in CL, respectively. Besides, the overall shrinkage of random effects in CL is estimated to be 16.9%, which is within the acceptable range. Predictive and residual plots of the final model are available in Figure 2. The prediction-corrected visual predictive check is shown in Figure 3.

Table 3. Parameter estimates of the final model.

Parameter[a]	Estimate (90% CI)[b]	Bootstrap mean (90% CI)
TVCL[c], L/h	0.140 (0.123-0.159)	0.142 (0.119-0.163)
$\theta_{PMA,CL,Hill}$[d]	7.02 (5.05-9.76)	6.76 (5.08-9.70)
$\theta_{PMA,CL,Mat50}$[e], days	197 (188-206)	199 (185-209)
$\theta_{Scr,CL}$[f]	0.541 (0.455-0.644)	0.530 (0.455-0.643)
TVV_c[g], L	0.769 (0.705-0.839)	0.782 (0.681-0.868)
TVQ[h], L/h	0.147 (0.087-0.249)	0.0887 (0.0267-0.8149)
TVV_p[i], L	0.285 (0.211-0.385)	0.287 (0.169-0.482)
$\theta_{WT,CL}$[j] and $\theta_{WT,Q}$[k]	Fixed at 0.75	Fixed at 0.75
$\theta_{WT,Vc}$[l] and $\theta_{WT,Vp}$[m]	Fixed at 1	Fixed at 1
CVCL[n], %	12.3 (9.0-14.9)	11.9 (8.8-15.0)
$CVCL_{BOV}$[n], %	13.3 (9.8-16.1)	13.3 (10.4-15.7)
σ_{prop}[o], %	16.8 (12.2-23.2)	16.1 (12.0-23.5)
σ_{add}[p], mg/L	1.76 (1.25-2.47)	1.68 (1.29-2.40)

[a]The equations for population values are as follows:  ;  ;  ;  , where CL is vancomycin clearance, PMA is postmenstrual age in days, Q is vancomycin intercompartmental clearance, S_{Cr} is serum creatinine level in μmol/L, V_c is vancomycin central volume, V_p is vancomycin peripheral volume, and WT is body weight in kg.

[b]Parameters were estimated on the logarithmic scale (except for coefficient of variance describing between-subject variability in clearance [CVCL] and coefficient of variance describing between-occasion variability in clearance [$CVCL_{BOV}$]), and the displayed CIs are calculated based on the estimated standard errors on the logarithmic scale assuming normal distribution.

[c]TVCL: typical value of vancomycin clearance.

[d]$\theta_{PMA,CL,Hill}$: Hill factor describing the association between postmenstrual age in days and vancomycin clearance.

[e]$\theta_{PMA,CL,Mat50}$: postmenstrual age in days at which maturation in vancomycin clearance is 50%.

[f]$\theta_{Scr,CL}$: exponent describing serum creatinine effect on vancomycin clearance.

[g]TVV_c: typical value of vancomycin central volume.

[h]TVQ: typical value of vancomycin intercompartmental clearance.

[i]TVV_p: typical value of vancomycin peripheral volume.

[j]$\theta_{WT,CL}$: exponent describing body weight effect on vancomycin clearance.

[k]$\theta_{WT,Q}$: exponent describing body weight effect on vancomycin intercompartmental clearance.

[l]$\theta_{WT,Vc}$: exponent describing body weight effect on vancomycin central volume.

[m]$\theta_{WT,Vp}$: exponent describing body weight effect on vancomycin peripheral volume.

[n]Coefficient of variance describing between-subject variability in vancomycin clearance [CVCL] and coefficient of variance describing between-occasion variability in vancomycin clearance [$CVCL_{BOV}$] are converted from the estimated variance of random effects (ω^2) using the formula .

[o]σ_{prop}: proportional component of residual unexplained variability.

[p]σ_{add}: additive component of residual unexplained variability.

Figure 2. Predictive and residual plots of the final model. The observed VC_s and CWRES are plotted against the population and individual predicted VC_s of the final model in the graphs, as indicated. The dashed lines in the CWRES plots indicate the range of −1.96 to +1.96, within which 95% of the data points should fall. The observed agreements between observed and predicted VC_s and the distributions of CWRES demonstrate the good predictive power of the final model. VC_s: serum vancomycin concentration; CWRES: conditional weighted residual.

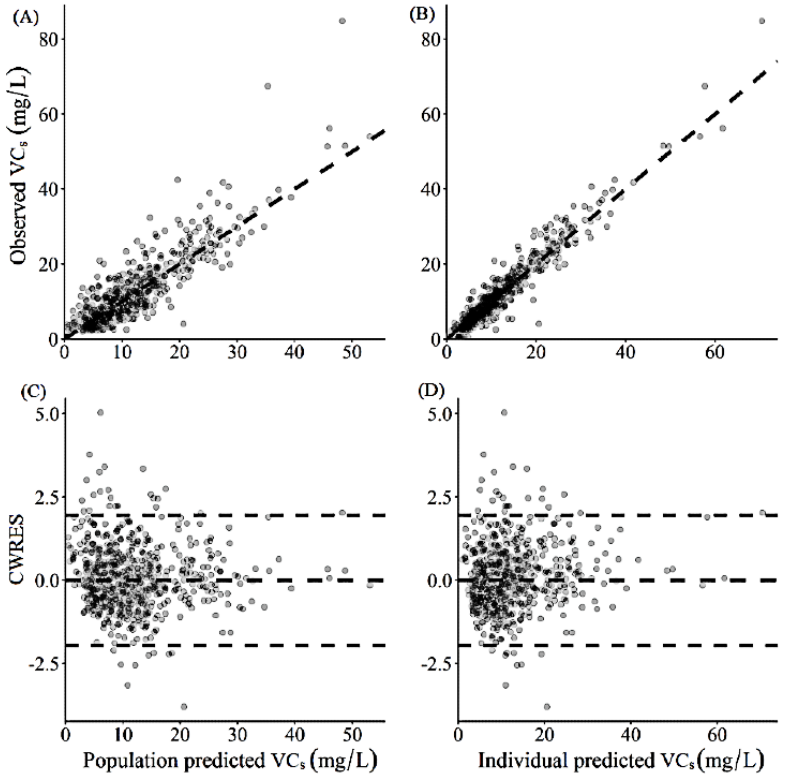
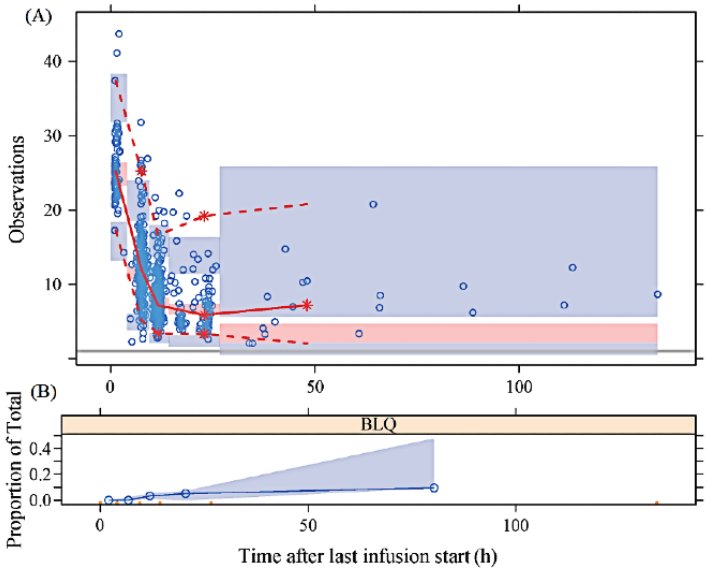


Figure 3. Prediction-corrected visual predictive check of the final model. (A) The 3 shaded areas (from bottom to top) for each time bin represent the 95% CI of the 5th percentiles, medians, and 95th percentiles of the corrected predictions; the dots represent the corrected observed VC_s; the solid line represents the binned medians of the corrected observed VC_s; the dashed lines represent the binned 5th and 95th percentiles of the corrected observed VC_s. Ideally, the percentiles of the observed VC_s should fall within the indicated CIs of predicted percentiles. (B) The shaded area and the line represent the 95% CI of predicted proportions and the observed proportions of BLQ concentrations, respectively. Most binned percentiles of the corrected observed VC_s fall within or are very close to the 95% CI of corrected predictions, demonstrating the predictive power of the final model. BLQ: below limit of quantification; VC_s: serum vancomycin concentration.

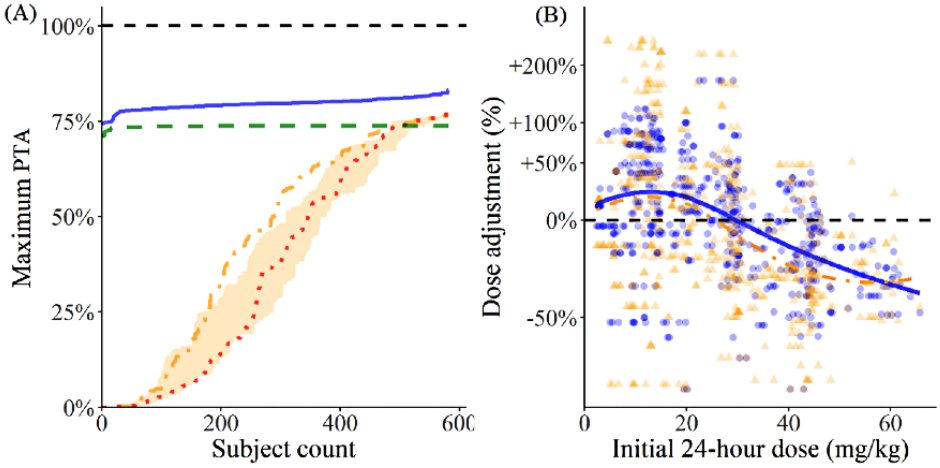


Performance of Dose Individualization

Based on the validated popPK model, the PTAs of different dosing approaches for the subjects in the data set were approximated. The graph on the left in Figure 4 shows that dose adjustments by the steady-state trough approach result in only half of the subjects achieving a PTA over 50%, which is only slightly improved when compared to maintaining the initial

doses given. It is also outperformed by the model-based approaches, namely, the empirical approach and the Bayesian approach, which reliably raise the PTA to above 75% for most subjects. Meanwhile, the graph on the right in Figure 4 shows some extreme dose adjustments with the steady-state trough approach when compared to the Bayesian approach, indicating overcorrection of doses with the trough approach without achieving a better PTA profile.

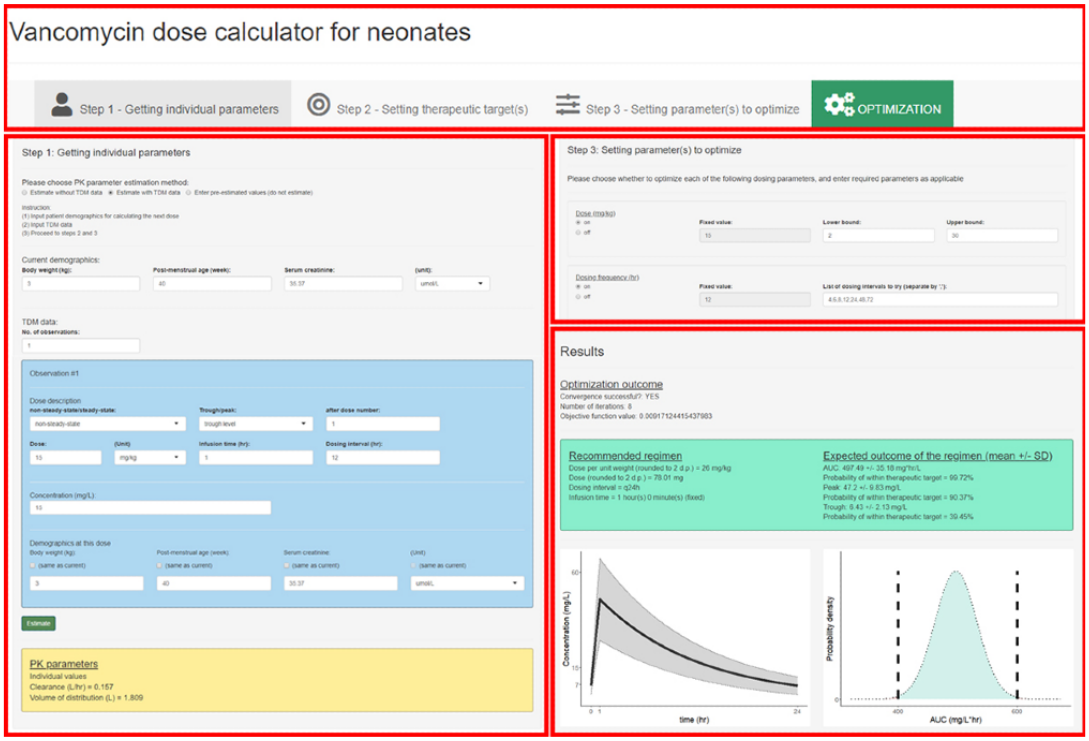
Figure 4. Probability of target attainments with different dose adjustment approaches. (A) The maximum PTA among the indicated count of subjects with the lowest PTAs under different dosing approaches. The lines represent the approximated outcomes of (1) maintaining the initial dose given (red dotted), (2) steady-state trough approach by targeting a $VC_{s,ss,trough}$ of 8.5 mg/L (orange dot-dashed, where the shaded region represents the previously recommended target range of 7-10 mg/L), (3) the model-based empirical approach (green dashed), and (4) the model-based Bayesian approach (blue solid). (B) The percentage changes from the initial 24-hour doses to the optimal doses with the steady-state trough approach (orange triangles with an orange dot-dashed fitting curve) and the model-based Bayesian approach (blue circles with a blue solid fitting curve). The downward sloping fitting curves agree with the general trend that the dose is increased (or decreased) when it is too low (or high). PTA: probability of target attainment. $VC_{s,ss,trough}$: steady-state trough serum vancomycin concentration.



Web-Based Dosing Interface

A composite screenshot of the developed interface is available in Figure 5. Detailed screenshots of the developed interface are available in Multimedia Appendix 1. The top panel is always displayed and allows the user to navigate different steps using the interface. By clicking a tab, the corresponding panel will be displayed below the top panel. Step 1 requires user inputs to estimate individual parameters. The user may choose between the model-based approaches (the empirical or Bayesian approach), depending on whether VC_s data are available. If the latter is chosen, then apart from the current body weight, PMA, and SCr, the user also needs to input previous doses administered, measured VC_s, and previous body weight, PMA, and SCr. In step 2, the user may specify the desired therapeutic targets, which defaults to an AUC_{24}/MIC of 400 to 600 hours without constraints by $VC_{s,ss,trough}$ and $VC_{s,ss,peak}$. Step 3 allows the user to specify the range of doses and dosing intervals allowed during optimization, which are by default set according to the usual practices of the hospitals using the interface. In most cases, accepting the defaults for steps 2 and 3 suffices. After inputting the required data, the results of individual dose optimization will be generated in the "optimization" tab, which suggests the dose (and dosing intervals) required to maximize PTA and the graphical illustrations of the steady-state VC_s profile and the expected probability distribution of AUC_{24}.

Figure 5. Example screenshot of the individual dose optimization interface. The top part shows the set of ordered tab buttons that are always displayed at the top of the window to guide the users through the steps of using the interface. The left part shows the panel for step 1 to get individual parameter estimates. The upper-right part shows the panel for step 3 to set the ranges of dosing parameters (range of doses and dosing intervals) to optimize. The lower-right part is the panel showing the results of individual dose optimization.



Discussion

Fulfillment of Study Objectives

Real-world data from EHRs were successfully used to develop a popPK model of IV vancomycin for the local NICU population. Based on a previously published dosing interface for high-dose methotrexate, a dosing interface for individual dose optimization of IV vancomycin for the local NICU population was created.

Reconciliation of popPK Data From EHRs

Despite the promising aspects of using EHRs, challenges were present when attempting to reconcile a popPK data set with EHR resources. First, especially when encountering a new EHR source, efforts were required to understand and validate the data structure of the source to ensure the likeliness of generating necessary tables for statistical analyses. To enable popPK analyses, it is essential to ensure that the target information can be reformatted into tables with different row representations (ie, 1 row per, for example, subject, dose, and observed VC_s). Then, since EHRs are primarily archived automatically during clinical operation, there is the issue of unstandardized or ambiguous inputs, especially for manual fields, because different clinics may have different logging practices. For instance, laboratories may run assays with different LLOQs, which could be logged onto the EHR systems using various syntaxes. Other problems encountered were suspected duplicated or missing records. For example, detectable VC_s measured before the first recorded dose or 7 days after the last dose in subjects with normal renal function may indicate missing dosing records.

A major limitation of using EHR data following the above issues is that data errors and ambiguity are often untraceable. To ensure the robustness of the final data set submitted for popPK analyses, it is crucial to remove problematic data that cannot be clarified from the EHR sources while keeping an eye on the possible risk of causing biased estimates (eg, censored data that are missing not at random).

While having to identify unsalvageable data is a downside, using EHRs is a convenient way to obtain a useful volume of data. Under the hectic environment of hospital wards, it is often difficult for clinicians to cater to the collection of study data. Making use of EHRs can ease the data collection process by minimizing the clinical workforce required. Moreover, since most EHR fields are already standardized, organized, and validated to a certain extent, typographical errors are less of a concern when extracting information from EHR sources.

popPK Model Development

The covariates can explain a significant proportion of BSV as expected. Diagnostic plots and the prediction-corrected visual predictive check show good predictive performance. The agreement between final parameter estimates against bootstrapping results and the relatively small condition number demonstrates the stability of the estimates. The choices of parameter-covariate relationships in the final model structures and the resultant parameter estimates in this study generally agree with previously reported models [26,28,29]. These positive results of evaluations help establish the validity of the model for implementation into the dose individualization interface.

Advantages of the Web-Based Dosing Interface

It is anticipated that the implementation of the developed interface can bring about several improvements to the current practice of administering IV vancomycin to treat severe infections in critically ill neonates. First, with the support of the popPK model developed, the interface can estimate individual AUC more accurately and enhance the optimality of the recommended initial dose (with the empirical approach) and maintenance dose (with the Bayesian approach). The recommended dose is also adaptive to significant changes in individual vancomycin PK due to variations in body weight, PMA, and renal functions during treatment. Moreover, with the Bayesian approach, presteady-state VC_s is also usable for estimation, such that waiting until the release of a steady-state VC_s measurement result for dose adjustment is no longer required. Together, these advantages promote the PTA profile and shorten the time to achieve the pharmacokinetic target by reducing the number of dose adjustments required. This is, in turn, expected to improve the treatment outcomes by promoting recovery while mitigating the risk of developing AKI. Apart from that, implementing the interface eliminates the need for manual calculation and thus reduces the risks of arithmetic errors in dose adjustments. The interface is also designed in a user-friendly and foolproof manner to ease its application by clinicians. Furthermore, the interface is developed using open-source software such that accessibility is guaranteed and licensing costs can be saved.

Conclusions and Future Studies

Based on a data set reconciled from real-world data extracted from multiple EHR sources, a popPK model of IV vancomycin has been developed and verified for the local NICU population. Based on the verified model and adoption of a previously published framework, a web-based dosing interface has been built to apply model-based approaches to individual AUC estimation and dose optimization of IV vancomycin. The developed interface is expected to improve clinical outcomes of the treatment of severe infections compared with previously adopted approaches, namely, the weight-based approach for initial dosing and the trough-based approach for dose adjustments. A cohort study will be performed later to show the superiority of using the interface compared with the previous approaches in terms of clinical outcomes. The experiences gained in this study will be valuable for the future use of the data collected from EHR sources for popPK analyses and the development of similar interfaces for other drug entities with narrow therapeutic indices.

Acknowledgments

The authors would like to thank Ms IU, Pui Ching Iris for her assistance with the manuscript.

Authors' Contributions

KHMH and TNTL designed the research. CHTC, YSJL, PHTL, and LYBC collected the data. KHMH analyzed the data and developed the dosing interface. KHMH wrote the manuscript. HSL, TNTL, CPL, CLYE, and YTC reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed screenshots of the web-based dosing interface.
[DOCX File , 441 KB - medinform_v10i1e29458_app1.docx]

References

1. Baumgart S, Hall SE, Campos JM, Polin RA. Sepsis with coagulase-negative staphylococci in critically ill newborns. Am J Dis Child 1983 May 01;137(5):461-463. [doi: 10.1001/archpedi.1983.02140310043012] [Medline: 6846275]
2. Lodise TP, Patel N, Lomaestro BM, Rodvold KA, Drusano GL. Relationship between initial vancomycin concentration-time profile and nephrotoxicity among hospitalized patients. Clin Infect Dis 2009 Aug 15;49(4):507-514. [doi: 10.1086/600884] [Medline: 19586413]
3. Linder A, Fjell C, Levin A, Walley KR, Russell JA, Boyd JH. Small acute increases in serum creatinine are associated with decreased long-term survival in the critically ill. Am J Respir Crit Care Med 2014 May 01;189(9):1075-1081. [doi: 10.1164/rccm.201311-2097OC] [Medline: 24601781]
4. Matzke GR, Kovarik JM, Rybak MJ, Boike SC. Evaluation of the vancomycin-clearance:creatinine-clearance relationship for predicting vancomycin dosage. Clin Pharm 1985;4(3):311-315. [Medline: 4006395]
5. Cockcroft DW, Gault H. Prediction of creatinine clearance from serum creatinine. Nephron 1976;16(1):31-41. [doi: 10.1159/000180580] [Medline: 1244564]
6. Schwartz GJ, Haycock GB, Edelmann CM, Spitzer A. A simple estimate of glomerular filtration rate in children derived from body length and plasma creatinine. Pediatrics 1976 Aug;58(2):259-263. [Medline: 951142]

7. Chu Y, Luo Y, Ji S, Jiang M, Zhou B. Population pharmacokinetics of vancomycin in Chinese patients with augmented renal clearance. Journal of Infection and Public Health 2020 Jan;13(1):68-74 [FREE Full text] [doi: 10.1016/j.jiph.2019.06.016] [Medline: 31277936]
8. Marqués-Miñana MR, Saadeddin A, Peris J. Population pharmacokinetic analysis of vancomycin in neonates. A new proposal of initial dosage guideline. Br J Clin Pharmacol 2010 Nov;70(5):713-720 [FREE Full text] [doi: 10.1111/j.1365-2125.2010.03736.x] [Medline: 21039765]
9. Moffett BS, Morris J, Munoz F, Arikan AA. Population pharmacokinetic analysis of vancomycin in pediatric continuous renal replacement therapy. Eur J Clin Pharmacol 2019 Aug;75(8):1089-1097. [doi: 10.1007/s00228-019-02664-7] [Medline: 30937470]
10. Okada A, Kariya M, Irie K, Okada Y, Hiramoto N, Hashimoto H, et al. Population Pharmacokinetics of Vancomycin in Patients Undergoing Allogeneic Hematopoietic Stem-Cell Transplantation. J Clin Pharmacol 2018 Sep;58(9):1140-1149. [doi: 10.1002/jcph.1106] [Medline: 29762865]
11. Stockmann C, Hersh AL, Roberts JK, Bhongsatiern J, Korgenski EK, Spigarelli MG, et al. Predictive Performance of a Vancomycin Population Pharmacokinetic Model in Neonates. Infect Dis Ther 2015 Jun;4(2):187-198 [FREE Full text] [doi: 10.1007/s40121-015-0067-9] [Medline: 25998107]
12. Rhodin MM, Anderson BJ, Peters AM, Coulthard MG, Wilkins B, Cole M, et al. Human renal function maturation: a quantitative description using weight and postmenstrual age. Pediatr Nephrol 2009 Jan;24(1):67-76. [doi: 10.1007/s00467-008-0997-5] [Medline: 18846389]
13. Rybak M, Lomaestro B, Rotschafer JC, Moellering R, Craig W, Billeter M, et al. Therapeutic monitoring of vancomycin in adult patients: a consensus review of the American Society of Health-System Pharmacists, the Infectious Diseases Society of America, and the Society of Infectious Diseases Pharmacists. Am J Health Syst Pharm 2009 Jan 01;66(1):82-98. [doi: 10.2146/ajhp080434] [Medline: 19106348]
14. Rybak MJ, Le J, Lodise TP, Levine DP, Bradley JS, Liu C, et al. Therapeutic monitoring of vancomycin for serious methicillin-resistant Staphylococcus aureus infections: A revised consensus guideline and review by the American Society of Health-System Pharmacists, the Infectious Diseases Society of America, the Pediatric Infectious Diseases Society, and the Society of Infectious Diseases Pharmacists. Am J Health Syst Pharm 2020 May 19;77(11):835-864. [doi: 10.1093/ajhp/zxaa036] [Medline: 32191793]
15. Neely MN, Kato L, Youn G, Kraler L, Bayard D, van Guilder M, et al. Prospective Trial on the Use of Trough Concentration versus Area under the Curve To Determine Therapeutic Vancomycin Dosing. Antimicrob Agents Chemother 2018 Feb;62(2):e02042 [FREE Full text] [doi: 10.1128/AAC.02042-17] [Medline: 29203493]
16. Neely MN, Youn G, Jones B, Jelliffe RW, Drusano GL, Rodvold KA, et al. Are vancomycin trough concentrations adequate for optimal dosing? Antimicrob Agents Chemother 2014;58(1):309-316 [FREE Full text] [doi: 10.1128/AAC.01653-13] [Medline: 24165176]
17. Frymoyer A, Stockmann C, Hersh AL, Goswami S, Keizer RJ. Individualized Empiric Vancomycin Dosing in Neonates Using a Model-Based Approach. J Pediatric Infect Dis Soc 2019 May 11;8(2):97-104. [doi: 10.1093/jpids/pix109] [Medline: 29294072]
18. de Hoog M, Schoemaker RC, Mouton JW, van den Anker JN. Vancomycin population pharmacokinetics in neonates. Clin Pharmacol Ther 2000 Apr;67(4):360-367. [doi: 10.1067/mcp.2000.105353] [Medline: 10801244]
19. Eichler H, Bloechl-Daum B, Broich K, Kyrle PA, Oderkirk J, Rasi G, et al. Data Rich, Information Poor: Can We Use Electronic Health Records to Create a Learning Healthcare System for Pharmaceuticals? Clin Pharmacol Ther 2019 Apr;105(4):912-922 [FREE Full text] [doi: 10.1002/cpt.1226] [Medline: 30178490]
20. Choi L, Beck C, McNeer E, Weeks HL, Williams ML, James NT, et al. Development of a System for Postmarketing Population Pharmacokinetic and Pharmacodynamic Studies Using Real-World Data From Electronic Health Records. Clin Pharmacol Ther 2020 Apr;107(4):934-943 [FREE Full text] [doi: 10.1002/cpt.1787] [Medline: 31957870]
21. Mould DR, Upton RN. Basic concepts in population modeling, simulation, and model-based drug development. CPT Pharmacometrics Syst Pharmacol 2012 Sep 26;1(9):1-14 [FREE Full text] [doi: 10.1038/psp.2012.4] [Medline: 23835886]
22. Mould DR, Upton RN. Basic concepts in population modeling, simulation, and model-based drug development-part 2: introduction to pharmacokinetic modeling methods. CPT Pharmacometrics Syst Pharmacol 2013 Apr 17;2(4):1-14 [FREE Full text] [doi: 10.1038/psp.2013.14] [Medline: 23887688]
23. Karlsson MO, Sheiner LB. The importance of modeling interoccasion variability in population pharmacokinetic analyses. Journal of Pharmacokinetics and Biopharmaceutics 1993 Dec 1;21(6):735-750. [doi: 10.1007/bf01113502]
24. Keizer RJ, Karlsson MO, Hooker A. Modeling and Simulation Workbench for NONMEM: Tutorial on Pirana, PsN, and Xpose. CPT Pharmacometrics Syst Pharmacol 2013 Jun 26;2(6):1-9 [FREE Full text] [doi: 10.1038/psp.2013.24] [Medline: 23836189]
25. Holford NHG, Anderson BJ. Allometric size: The scientific theory and extension to normal fat mass. Eur J Pharm Sci 2017 Nov 15;109S:S59-S64. [doi: 10.1016/j.ejps.2017.05.056] [Medline: 28552478]
26. Anderson BJ, Woollard GA, Holford NH. A model for size and age changes in the pharmacokinetics of paracetamol in neonates, infants and children. Br J Clin Pharmacol 2000 Aug;50(2):125-134 [FREE Full text] [doi: 10.1046/j.1365-2125.2000.00231.x] [Medline: 10930964]

27. Kimura T, Sunakawa K, Matsuura N, Kubo H, Shimada S, Yago K. Population Pharmacokinetics of Arbekacin, Vancomycin, and Panipenem in Neonates. Antimicrob Agents Chemother 2004 Apr;48(4):1159-1167 [FREE Full text] [doi: 10.1128/aac.48.4.1159-1167.2004] [Medline: 15047516]
28. Anderson BJ, Allegaert K, Van den Anker JN, Cossey V, Holford NHG. Vancomycin pharmacokinetics in preterm neonates and the prediction of adult clearance. Br J Clin Pharmacol 2007 Jan;63(1):75-84 [FREE Full text] [doi: 10.1111/j.1365-2125.2006.02725.x] [Medline: 16869817]
29. Grimsley C, Thomson AH. Pharmacokinetics and dose requirements of vancomycin in neonates. Arch Dis Child Fetal Neonatal Ed 1999 Nov 23;81(3):F221-F227 [FREE Full text] [doi: 10.1136/fn.81.3.f221] [Medline: 10525029]
30. Boeckmann A, Sheiner L, Beal S. NONMEM Users Guide. Ellicott City, MD: Icon Development Solutions; 2009.
31. Ahn JE, Karlsson MO, Dunne A, Ludden TM. Likelihood based approaches to handling data below the quantification limit using NONMEM VI. J Pharmacokinet Pharmacodyn 2008 Aug;35(4):401-421. [doi: 10.1007/s10928-008-9094-4] [Medline: 18686017]
32. Lindbom L, Pihlgren P, Jonsson EN. PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. Comput Methods Programs Biomed 2005 Sep;79(3):241-257. [doi: 10.1016/j.cmpb.2005.04.005] [Medline: 16023764]
33. Nguyen THT, Mouksassi M, Holford N, Al-Huniti N, Freedman I, Hooker AC, Model Evaluation Group of the International Society of Pharmacometrics (ISoP) Best Practice Committee. Model Evaluation of Continuous Data Pharmacometric Models: Metrics and Graphics. CPT Pharmacometrics Syst Pharmacol 2017 Feb;6(2):87-109 [FREE Full text] [doi: 10.1002/psp4.12161] [Medline: 27884052]
34. Thai H, Mentré F, Holford NHG, Veyrat-Follet C, Comets E. Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics. J Pharmacokinet Pharmacodyn 2014 Feb;41(1):15-33. [doi: 10.1007/s10928-013-9343-z] [Medline: 24317870]
35. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.r-project.org/ [accessed 2022-01-17]
36. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer-Verlag; 2016.
37. Hui KH, Chu HM, Fong PS, Cheng WTF, Lam TN. Population Pharmacokinetic Study and Individual Dose Adjustments of High-Dose Methotrexate in Chinese Pediatric Patients With Acute Lymphoblastic Leukemia or Osteosarcoma. J Clin Pharmacol 2019 Apr;59(4):566-577. [doi: 10.1002/jcph.1349] [Medline: 30556906]

Abbreviations

AKI: acute kidney injury
AUC: steady-state area under the curve of the serum vancomycin concentration-time profile
AUC_{24}: steady-state area under the curve of the serum vancomycin concentration-time profile over 24 hours
BLQ: below the limit of quantification
BOV: between-occasion variability
BSV: between-subject variability
CIS: Clinical Information System
CL: vancomycin clearance
CV: coefficient of variance
CVCL: coefficient of variance of between-subject variability in vancomycin clearance
$CVCL_{BOV}$: coefficient of variance of between-occasion variability in vancomycin clearance
dOFV: change in objective function value
EHR: electronic health record
ePR: electronic Patient Record
GFR: glomerular filtration rate
IPMOE: In-Patient Medication Order Entry
IV: intravenous
LLOQ: lower limit of quantification
MIC: minimum inhibitory concentration
NICU: neonatal intensive care unit
OFV: objective function value
pcVPC: prediction-corrected visual predictive check
PMA: postmenstrual age
PNA: postnatal age
popPK: population pharmacokinetic
PTA: probability of target attainment
Q: vancomycin intercompartmental clearance
SCr: serum creatinine

TDM: therapeutic drug monitoring
V_c: vancomycin central volume
VC_s: serum vancomycin concentration
$VC_{s,peak}$: peak serum vancomycin concentration
$VC_{s,ss,peak}$: steady-state peak serum vancomycin concentration
$VC_{s,ss,trough}$: steady-state trough serum vancomycin concentration
$VC_{s,trough}$: trough serum vancomycin concentration
V_p: vancomycin peripheral volume

Edited by C Lovis; submitted 27.04.21; peer-reviewed by R Mirzaei, A Allam; comments to author 30.07.21; revised version received 06.08.21; accepted 02.01.22; published 31.01.22.

Please cite as:
Hui KHM, Lam HS, Chow CHT, Li YSJ, Leung PHT, Chan LYB, Lee CP, Ewig CLY, Cheung YT, Lam TNT
Using Electronic Health Records for Personalized Dosing of Intravenous Vancomycin in Critically Ill Neonates: Model and Web-Based Interface Development Study
JMIR Med Inform 2022;10(1):e29458
URL: https://medinform.jmir.org/2022/1/e29458
doi:10.2196/29458
PMID:35099393

Corrigenda and Addenda

Correction: Candidemia Risk Prediction (CanDETEC) Model for Patients With Malignancy: Model Development and Validation in a Single-Center Retrospective Study

Junsang Yoo[1*], RN, PhD; Si-Ho Kim[2*], MD; Sujeong Hur[3,4], RN, MS; Juhyung Ha[5]; Kyungmin Huh[6], MD; Won Chul Cha[4,7,8], MD

[1]Department of Nursing, College of Nursing, Sahmyook University, Seoul, Republic of Korea
[2]Division of Infectious Disease, Samsung Changwon Hospital, Sungkyunkwan University School of Medicine, Changwon, Republic of Korea
[3]Department of Patient Experience Management, Samsung Medical Center, Seoul, Republic of Korea
[4]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea
[5]Department of Computer Science, Indiana University Bloomington, Bloomington, IN, United States
[6]Division of Infectious Disease, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[7]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[8]Digital Innovation Center, Samsung Medical Center, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Won Chul Cha, MD
Department of Emergency Medicine
Samsung Medical Center
Sungkyunkwan University School of Medicine
81, Irwon-ro, Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 2053
Email: wc.cha@samsung.com

Related Article:

Correction of: https://medinform.jmir.org/2021/7/e24651

(JMIR Med Inform 2022;10(1):e36385) doi:10.2196/36385

In “Candidemia Risk Prediction (CanDETEC) Model for Patients With Malignancy: Model Development and Validation in a Single-Center Retrospective Study” (JMIR Med Inform 2021;9(7):e24651), one error was noted.

Due to a system error, the ORCID number of author Sujeong Hur was incorrectly published as:

0000-0001-7763-8940

This has been corrected to:

0000-0003-1335-576X

The correction will appear in the online version of the paper on the JMIR Publications website on January 19, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 12.01.22; this is a non–peer-reviewed article;accepted 12.01.22; published 19.01.22.

Please cite as:
Yoo J, Kim SH, Hur S, Ha J, Huh K, Cha WC
Correction: Candidemia Risk Prediction (CanDETEC) Model for Patients With Malignancy: Model Development and Validation in a Single-Center Retrospective Study
JMIR Med Inform 2022;10(1):e36385
URL: https://medinform.jmir.org/2022/1/e36385
doi:10.2196/36385
PMID:35044928

Original Paper

Effects of Social Media Use for Health Information on COVID-19–Related Risk Perceptions and Mental Health During Pregnancy: Web-Based Survey

Qian Wang[1], PhD; Luyao Xie[2], MMed; Bo Song[1], MMed; Jiangli Di[1], PhD; Linhong Wang[3], MMed; Phoenix Kit-Han Mo[2], PhD

[1]National Center for Women and Children's Health, Chinese Center for Disease Control and Prevention, Beijing City, China

[2]Center for Health Behaviours Research, Faculty of Medicine, School of Public Health and Primary Care, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, Hong Kong

[3]National Center for Chronic and Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing City, China

Corresponding Author:
Phoenix Kit-Han Mo, PhD
Center for Health Behaviours Research, Faculty of Medicine
School of Public Health and Primary Care
The Chinese University of Hong Kong, Prince of Wales Hospital
202D
Hong Kong
Hong Kong
Phone: 852 2252 8765
Email: phoenix.mo@cuhk.edu.hk

Abstract

Background: Social media has become an important source of health information during the COVID-19 pandemic. Very little is known about the potential mental impact of social media use on pregnant women.

Objective: This study aims to examine the association between using social media for health information and risk perception for COVID-19, worry due to COVID-19, and depression among pregnant women in China.

Methods: A total of 4580 pregnant women were recruited from various provinces of China. The participants completed a cross-sectional, web-based survey in March 2020.

Results: More than one-third (1794/4580, 39.2%) of the participants reported always using social media for obtaining health information. Results of structural equation modeling showed that the frequency of social media use for health information was positively associated with perceived susceptibility (β=.05; *P*<.001) and perceived severity (β=.12; *P*<.001) of COVID-19, which, in turn, were positively associated with worry due to COVID-19 (β=.19 and β=.72, respectively; *P*<.001). Perceived susceptibility (β=.09; *P*<.001), perceived severity (β=.08; *P*<.001), and worry due to COVID-19 (β=.15; *P*<.001) all had a positive association with depression. Bootstrapping analysis showed that the indirect effects of frequency of social media use for health information on both worry due to COVID-19 (β=.09, 95% CI 0.07-0.12) and depression (β=.05, 95% CI 0.02-0.07) were statistically significant.

Conclusions: This study provides empirical evidence on how social media use for health information might have a negative impact on the mental health of pregnant women. Interventions are needed to equip this population with the skills to use social media properly and with caution.

(JMIR Med Inform 2022;10(1):e28183) doi:10.2196/28183

KEYWORDS

COVID-19; pregnant; social media use; risk perception; worry; depression

Introduction

COVID-19 is an infectious disease caused by a newly discovered coronavirus, named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus is known to have originated in Wuhan, China, in December 2019, and since then, it has spread rapidly, resulting in a global pandemic [1]. The rapid transmission of COVID-19 has caused massive disruption worldwide. As of February 14, 2021, more than 108 million people across 235 countries were infected with COVID-19, and more than 2 million associated deaths were reported [2].

Pregnant women are more susceptible to the morbidity and mortality associated with COVID-19, owing to the physiological changes that occur in the immune and cardiopulmonary systems during pregnancy [3,4]. A systematic review of 27 studies reported that 9.3% of pregnant women with COVID-19 were admitted to the intensive care unit, and 5.4% of them required mechanical ventilation [5]. As a uniquely vulnerable group, pregnant women require special attention and care during a pandemic. However, reduced access to health facilities during the COVID-19 pandemic caused significant psychological toll among people [6]. Pregnant women also experience serious stress and anxiety due to fear of infection, antenatal care suspension, boredom, frustration, and worries about the health of the fetus [7,8]. This may also lead to adverse effects for the child, such as inefficient mother-infant bonding [9] and the risk of inherited psychiatric illness [10]. Previous studies on pregnant women during the severe acute respiratory syndrome (SARS) outbreak in Hong Kong have suggested that 12.3% scored higher than the cut-off for depression, and 87.8% of pregnant women reported higher than moderate level of anxiety during the COVID-19 pandemic [11,12]. A recent study among 900 pregnant women in Canada found that, compared to the pre–COVID-19 period, pregnant women showed a significantly higher level of depression (from 15% to 41%) and anxiety (from 29% to 72%) during the COVID-19 pandemic [13]. Other studies have also shown that pregnant women experienced greater psychological distress than the general population during the pandemic [14,15].

In recent years, the widespread use of the internet has allowed individuals to access health information and receive support in their health care [16]. Women tend to be more involved in seeking health information on the internet [17], and web searches for health information have been found to be popular among pregnant women. For example, a study among 332 Chinese pregnant women showed that 88.7% of them used the internet to obtain health information, starting from the beginning of their pregnancies [18]. In general, between 28% and 95% of pregnant women use the internet for health information [19]. Common web-based search topics for obtaining health information included fetal development [18,20], stages of childbirth [20], antenatal pregnancy complications [21], and pregnancy nutrition [18]. The ease and accessibility of searching the internet during pregnancy met novice mothers' information needs [21-23] and provided them with opportunities to share similar experiences and apprehensions with other women [23,24].

During the COVID-19 pandemic, the demand for information about the pandemic soared, and all published reports elevated public concerns about the serious threats of the pandemic. Social isolation measures resulted in consideration of the internet and social media as the primary sources of information about the pandemic [25,26]. Although social media has served as a powerful tool for disseminating health information, challenges and concerns over social media use have been raised. For example, there are serious concerns about misinformation and unsubstantiated rumors that were rapidly spread through social media, causing distrust and posing additional challenges for public health efforts to combat the pandemic [27]. Second, social media tends to overemphasize risks; repeated exposure to such platforms may, therefore, increase negative emotions such as panic and fear [28]. Third, compared to traditional media, social media not only provides information but also allows personal sharing and emotional expressions. Negative emotions are more likely to be conveyed on social media during an infectious disease pandemic [29]. A recent content analysis of messages posted to social media platforms in China during the COVID-19 period showed that personal posts are likely to attribute blame to other individuals or the government and express concerns about the pandemic [30]. Some studies have shown that social media exposure was positively related to increased anxiety, fear, posttraumatic stress disorder, and forming risk perceptions during previous outbreaks, such as the Middle East respiratory syndrome (MERS) outbreak, and the current COVID-19 crisis [29,31,32]. A study among factory workers in Shenzhen, China, conducted at the beginning of the COVID-19 pandemic, also found that higher exposure via unofficial web-based media was associated with higher depressive symptoms [25].

There could be more than a direct association between exposure to health information and mental distress during the COVID-19 period. Exposure to distressing health information on social media may intensify risk perceptions, leading to poor mental outcomes. Risk perception refers to an individual's subjective evaluation of the possibility of a negative event; it consists of two key components: perceived susceptibility (ie, perception of the likelihood of contracting the disease) and perceived severity (ie, perception of the extent of harm of the disease). The cognitive model suggested that negative perceptions about a disease could increase worry or anxiety of one's health status [33]. Evidence from previous public health crises (ie, Ebola and H1N1 outbreaks) also revealed that when a community crisis is repeatedly exposed in the media, considerable information about the risk of the health crisis could unintendedly lead to heightened anxiety and stress reactions [34,35].

Furthermore, information about the pandemic might change an individual's perceived susceptibility and perceived harm of the disease [36,37]. Studies have found that during a global pandemic, mass media information would likely affect the perceived threat from the disease [38]; perceived threat, in turn, has shown to have a direct positive effect on negative mental outcomes, such as sadness, depression, anxiety, and anger [39]. Gender-based difference was also observed, with women perceiving higher levels of threat than men [38]. For pregnant women, pregnancy itself is characterized with heightened

worries. Given the stress and uncertainty brought by the COVID-19 pandemic, using social media to obtain health information can be accompanied by various stressors, such as excessive information, long-term confinement, and fear of infection, all of which might increase the risk perceptions of this population [6].

Perception of susceptibility and severity may also lead to negative emotions, which in turn, could affect an individual's mental health. The Appraisal Theory posits that emotions result from an individual's evaluation or appraisal of an event, even in the absence of physiological arousal [40]. The appraisal process involves evaluation of two aspects of a situation: motivational relevance and motivational congruence. Motivational relevance assesses the relevance of the situation to one's well-being, whereas motivational congruence evaluates the congruence of the situation with one's goal. More intense emotional responses occur when a situation is judged to be highly relevant to one's well-being and inconsistent to one's goal [41]. It is therefore contended that the perception that one is at risk for COVID-19 infection and that the disease would have severe negative consequences will elicit negative emotions, leading to an adverse mental response. The association between risk perception of a disease and negative emotional reactions has been widely demonstrated in the literature [42,43]. In the context of COVID-19, studies from some Asian countries, including the Philippines and Vietnam, also support the findings that perceived susceptibility and impact of COVID-19 are related to negative emotions and poor mental health [44,45].

Based on the Appraisal Theory, this study aims to investigate whether and how social media use for health information might be associated with mental health outcomes among Chinese pregnant women during the COVID-19 era. In particular, the relationship between the use of social media for health information, risk perception (ie, perceived susceptibility and perceived severity of COVID-19), worry due to COVID-19, and depression were examined. It was hypothesized that using social media for health information would be associated with a higher level of risk perception that, in turn, would be associated with higher levels of worry and depression. Worry due to COVID-19 would also be positively associated with depression.

Methods

Study Design

A web-based, cross-sectional survey was conducted in March 2020. Pregnant women who were availing health services from maternal health care centers in Mainland China and who intended to continue the pregnancy were included in this study. Those who planned to terminate their pregnancy were excluded from the sample.

Procedures

Participants were recruited from maternal health care centers of various provinces of China (ie, Beijing, Chongqing, Guangdong, Guangxi, Hainan, Shandong, Tianjin, and Xinjiang). Eligible women were first identified from medical records obtained from the center, and they were invited to participate in the survey through WeChat. Interested participants could access the web-based survey by scanning the quick response (QR) code or by clicking the link provided in the WeChat invitation message. Information about the purpose and procedure of the survey was provided on the first page of the web-based survey. Participants were assured about the confidentiality of the study and that refusal to participate in the survey would not affect any future services they would avail at the center. Informed consent was obtained from the participants by asking them to click on the "I agree" button on the first page of the survey. Ethical approval was obtained from the authors' institution. A total of 4580 complete responses were collected (70% response rate).

Measures

Sociodemographic Characteristics

Sociodemographic and pregnancy-related characteristics, including age, education level, parity, gestational age, and whether the participants had any pregnancy-related complications were collected.

Frequency of Social Media Use for Obtaining Health Information

Participants were asked to rate a single item about their frequency of using social media to seek health information in the past week. Responses are rated on a 4-point Likert scale ranging from 1 (never) to 4 (always).

Perceived Susceptibility to COVID-19

Participants were asked to rate 2 items on the extent to which they perceived that they and their family members would likely contract COVID-19. Responses were recorded on a 4-point Likert scale ranging from 1 (very little) to 4 (very much). A higher score indicated a higher level of perceived susceptibility. The internal reliability of the items was satisfactory (Cronbach α=.93).

Perceived Severity of COVID-19

Participants were asked to rate 3 items on their perceived consequences of COVID-19 (eg, "maternal infection with COVID-19 will affect the health of the newborn"). Items were rated on a 4-point Likert scale ranging from 1 (very little) to 4 (very much), with a higher score indicating a higher level of perceived severity. The internal reliability of the items was satisfactory (Cronbach α=.92).

Worry Due to COVID-19

Participants were asked to rate 4 items assessing their level of worry on various aspects related to COVID-19 (eg, "you will be infected with COVID-19 when you attend the prenatal check-up"). Items were rated on a 4-point Likert scale ranging from 1 (very little) to 4 (very much), with a higher score indicating a higher level of worry. The internal reliability of the items was satisfactory (Cronbach α=.91).

Depression

Depression was measured using the Patient Health Questionnaire-9 [46], which has been validated and used in the Chinese population [47,48]. Participants were asked to rate how

often they have been bothered by COVID-19–related symptoms in the past 2 weeks on a 4-point Likert scale ranging from 0 (not at all) to 3 (almost every day). Total scores ranged from 0 to 27, with a higher score indicating higher level of depression. A score of 0 to 4, 5 to 9, 10 to 14, 15 to 19, and 20 to 27 represented minimal, mild, moderate, moderately severe, and severe depression, respectively.

Data Analyses

Descriptive statistics and zero-order correlations among all variables were performed. To evaluate the association between social media use for health information, risk perception, worry due to COVID-19 and depression, confirmatory factor analysis was conducted to assess the goodness of fit of the measurement model [49]. Structural equation modeling (SEM) was then performed to assess the hypothesized associations between the variables. Bootstrapping analysis, based on 2000 samples, was used to test the indirect effect. To evaluate the overall model fit, we considered the following indices: χ^2 statistic, comparative fit index (CFI), incremental fit index (IFI), and root mean square error of approximation (RMSEA). Analyses were performed using AMOS 26 (IBM Corp) and tested using the maximum likelihood method.

Results

Descriptive Statistics of Study Participants

Of the 4580 participants, one-third (n=1538, 33.6%) were above 30 years of age; half (n=2334, 51%) had received postsecondary level of education; and a similar number (n=2300, 50.2%) were nulliparous. Slightly less than half (2143/4580, 46.8%) the participants were in their third trimester of pregnancy. A small number (n=310, 6.8%) of all participants reported having some pregnancy-related complications. Slightly less than half (n=2226, 48.6%) scored higher than the cut-off score for mild depression, and more than one-third (n=1794, 39.2%) reported always using social media for health information in the past week. More than one-third (n=2887, 63.1% to n=3104, 67.7%) of all participants showed a high level of susceptibility toward COVID-19. Furthermore, between 79.2% (n=3630) and 86.4% (n=3959) and between 68.5% (n=3136) and 75.5% (n=3462) of the 4580 participants reported a high level of severity and worry about COVID-19 (Table 1).

Table 1. Background characteristics of study participants (N=4580)

Characteristic	Value, n (%)
Age (years)	
≤19	62 (1.4)
20-25	967 (21.1)
26-30	2013 (44)
31-35	1197 (26.1)
36-40	274 (6)
≥41	67 (1.5)
Education level	
Primary or lower	117 (2.6)
Junior secondary	1130 (24.7)
Senior secondary	999 (21.8)
Matriculation	1218 (26.6)
Undergraduate	987 (21.6)
Postgraduate or higher	129 (2.8)
Parity	
Nulliparous	2300 (50.2)
Primiparous	2001 (43.7)
Multiparous	279 (6.1)
Gestational age	
First trimester (≤12 weeks)	904 (19.7)
Second trimester (13-26 weeks)	1533 (33.5)
Third trimester (≥27 weeks)	2143 (46.8)
Pregnancy-related complications	
No	4270 (93.2)
Yes	310 (6.8)
Depression (measured by PHQ-9[a])	
Minimal (0-4)	2354 (51.4)
Mild (5-9)	1302 (28.4)
Moderate (10-14)	567 (12.4)
Moderately severe (15-19)	252 (5.5)
Severe (20-27)	105 (2.3)
Frequency of social media use for health information in the past week	
Never	338 (7.4)
Seldom	845 (18.4)
Sometimes	1603 (35)
Always	1794 (39.2)
Perceived susceptibility (score ≥3)	
Likelihood of contracting COVID-19 themselves	3104 (67.7)
Likelihood of family members contracting COVID-19	2887 (63.1)
Perceived severity (score ≥3)	
“COVID-19 will be transmitted from mother to child”	3630 (79.2)
“Maternal infection of COVID-19 will be more difficult to cure than the general population”	3827 (83.5)

Characteristic	Value, n (%)
"Maternal infection of COVID-19 will affect the health of the child"	3959 (86.4)
Worry (score ≥3)	
Worry that you will be infected with COVID-19 when you attend the prenatal check-up	3462 (75.5)
Worry that your hospital delivery arrangement will be infected due to COVID-19	3136 (68.5)
Worry that accompany delivery will not be available due to COVID-19	3197 (69.8)
Worry that child health services will be affected after delivery due to COVID-19	3391 (74)

[a]PHQ-9: Patient Health Questionnaire-9.

Correlation Between Study Variables

Among the sociodemographic and pregnancy-related characteristics, age had a negative correlation with depression (r=–0.03, P<.05). The frequency of using social media for health information had a positive correlation with perceived susceptibility (r=0.05, P<.001), perceived severity (r=0.11, P<.001), and worry due to COVID-19 (r=0.09, P<.001), but it had no significant correlation with depression. Perceived susceptibility (r=0.15, P<.001), perceived severity (r=0.19, P<.001), and worry due to COVID-19 (r=0.22, P<.001) all had a significant correlation with depression (Table 2).

Table 2. Correlation between study variables.

	Age	Education level	Parity	Gestational age	Complications[a]	Frequency of use[b]	Susceptibility[c]	Severity[d]	Worry[e]	Depression
Age										
r	1	0.12	0.37	–0.02	0.06	0.04	–0.03	0.003	–0.02	–0.03
P value	—[f]	<.001	<.001	.12	<.001	.01	.04	.85	.19	.03
Education level										
r	0.12	1	–0.30	–0.10	0.05	0.25	0.01	–0.01	–0.05	–0.01
P value	<.001	—	<.001	<.001	.001	<.001	.44	.41	<.001	.34
Parity										
r	0.37	–0.30	1	0.09	–0.02	–0.08	–0.03	0.02	0.04	–0.004
P value	<.001	<.001	—	<.001	.12	<.001	.08	.20	.01	.76
Gestational age										
r	–0.02	–0.10	0.09	1	0.14	0.03	0.06	0.14	0.18	0.001
P value	.12	<.001	<.001	—	<.001	.02	<.001	<.001	<.001	.92
Complications[a]										
r	0.06	0.05	–0.02	0.14	1	0.02	0.01	0.03	0.03	0.001
P value	<.001	.001	.12	<.001	—	.16	.40	.03	.02	.996
Frequency of use[b]										
r	0.04	0.25	–0.08	0.03	0.02	1	0.05	0.11	0.09	–0.001
P value	.01	<.001	<.001	.02	.16	—	.001	<.001	<.001	.955
Susceptibility[c]										
r	–0.03	0.01	–0.03	0.06	0.01	0.05	1	0.31	0.37	0.15
P value	.04	.44	.08	<.001	.40	.001	—	<.001	<.001	<.001
Severity[d]										
r	0.003	–0.01	0.02	0.14	0.03	0.11	0.31	1	0.71	0.19
P value	.85	.41	.20	<.001	.03	<.001	<.001	—	<.001	<.001
Worry[e]										
R	–0.02	–0.05	0.04	0.18	0.03	0.09	0.37	0.71	1	0.22
P value	.19	<.001	.01	<.001	.02	<.001	<.001	<.001	—	<.001
Depression										
r	–0.03	–0.01	–0.004	0.001	0.001	–0.001	0.15	0.19	0.22	1
P value	.03	.34	.76	.92	.996	.955	<.001	<.001	<.001	—

[a]Pregnancy-related complications.

[b]Frequency of social media use for health information.

[c]Perceived susceptibility of COVID-19.

[d]Perceived severity of COVID-19.

[e]Worry due to COVID-19.

[f]Not applicable.

SEM Results

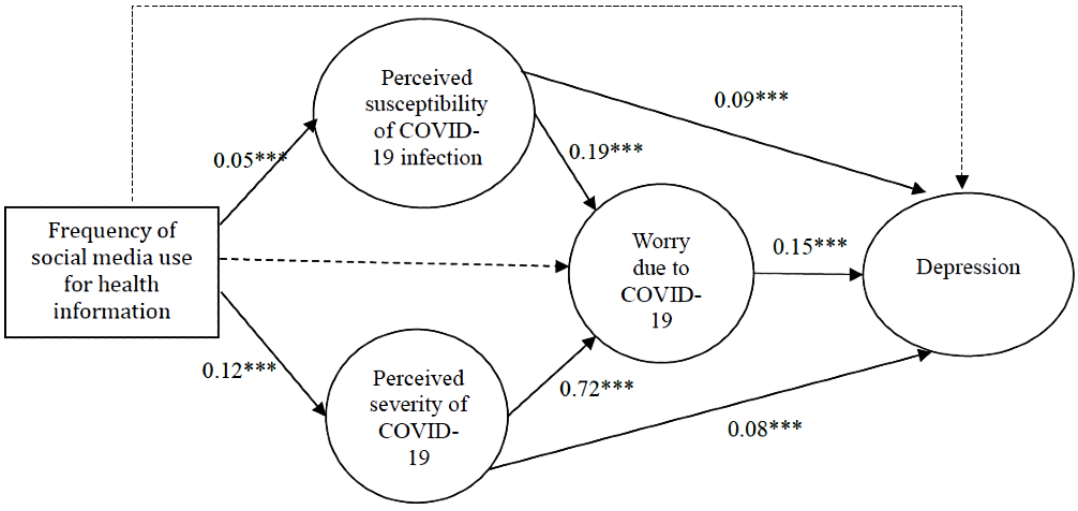
Results from the confirmatory factor analysis suggested that the measurement model showed good fit to the data (χ^2_{48}=695.76; *P*=.01; CFI=0.99; IFI=0.99; RMSEA=0.05). All factor loadings were significant at *P*<.001 (Table 3). SEM results also showed that the structural model fitted the data well (χ^2_{57}=1143.3; *P*<.001; CFI=0.98; IFI=0.97; RMSEA=0.06). Frequency of social media use for health information was positively associated with perceived susceptibility (β=.05; *P*<.001) and perceived severity of COVID-19 (β=.12; *P*<.001), which in turn were positively associated with worry due to

COVID-19 (β=.19 and β=.72, respectively; *P*<.001). Perceived susceptibility (β=.09; *P*<.001), perceived severity (β=.08; *P*<.001), and worry due to COVID-19 (β=.15; *P*<.001) all had a significant positive association with depression. In contrast, frequency of social media use for health information did not have a significant association with worry due to COVID-19 and depression (Figure 1). Bootstrapping analysis showed that the indirect effects of frequency of social media use for health information on worry due to COVID-19 (β=.09, 95% CI 0.07-0.12) and depression (β=.05, 95% CI 0.02-0.07) were both statistically significant.

Table 3. Unstandardized and standardized loadings for the measurement model.

Parameter estimates	Unstandardized loading (SE)	Standardized loading
Perceived susceptibility of COVID-19		
Item 1	1.00	0.95
Item 2	0.99 (0.03)	0.92
Perceived severity of COVID-19		
Item 1	1.00	0.86
Item 2	1.05 (0.01)	0.93
Item 3	1.02 (0.01)	0.91
Worry due to COVID-19		
Item 1	1.00	0.82
Item 2	1.16 (0.02)	0.90
Item 3	1.12 (0.02)	0.85
Item 4	1.11 (0.02)	0.88
Depression		
Parcel score 1	1.00	0.85
Parcel score 2	1.01 (0.01)	0.93
Parcel score 3	0.73 (0.01)	0.79

Figure 1. Structural equation model for social media use for health information, risk perceptions of COVID-19, worry due to COVID-19, and depression among pregnant women. The standardized coefficients of structural paths are shown after controlling significant background variables. Nonsignificant path is shown as a dotted line. Factor loadings and measurement errors have been omitted for clarity. ***P<.001.



Discussion

Principal Findings

With the proliferation and rapid development of internet technologies and social networking sites, social media has become an important source of health information. In this study, more than one-third (39.2%) of the participants reported that they always used social media for obtaining health information during the COVID-19 pandemic. These findings are consistent with previous reports documenting an extensive use of social media for health information following the COVID-19 outbreak [25,26]. With the practices of physical and social distancing, individuals have increasingly turned to social media for information related to safety precautions and news updates related to COVID-19. Understanding how social media use for

health information may change the information-seeking behaviors and health of pregnant women would be particularly valuable and meaningful.

To our knowledge, this is the first study that examined the associations between social media use for health information and depressive symptoms among pregnant women during the COVID-19 pandemic. It is important to note that, in the present study, nearly half (48.6%) of the participants were classified as having mild to severe depression—a figure that was significantly higher than that reported in the general population of pregnant women (ie, 7.4% to 12.8%) [50]. Furthermore, our findings show that the frequency of social media use for health information was indirectly associated with higher levels of worry and depression. These findings are in line with previous reports of a substantial proportion of pregnant women being confused about the complex or incorrect information available on the internet and experiencing heightened anxiety [24], as well as reports that have documented a positive relation between social media use and spread of fear and panic related to COVID-19 [51]. This study was conducted during the early phase of the COVID-19 pandemic, when China was significantly impacted; hence, it may be possible that participants in the present study did not only search for factual information about the pandemic, but they might also be exposed to the sharing of negative views, hot debate and arguments, and exaggerated worries toward COVID-19 via social media [30]. The mental impact of social media use during the COVID-19 period thus requires additional public health attention.

It would be important to understand the underlying mechanism through which social media use is associated with mental health. It is intriguing that using social media for health information was found to be indirectly associated with depression through perceived susceptibility and perceived severity about COVID-19, suggesting that social media use can affect the formation of risk perception of a pandemic. These findings are consistent with previous studies, which have documented that exposure to news media about a disease, such as H1N1, is associated with the formation of risk perceptions of the disease [52,53]. Social media has served as a useful tool for obtaining instant and up-to-date information during the COVID-19 outbreak. Nevertheless, since anyone can post on social media, it may also facilitate the sharing of inaccurate or unfiltered information, or the sharing of negative views, including uncertainty, severity, or suspicions of the disease. It is likely that exposure to the symptoms or complications related to COVID-19 may increase one's *perceived severity* of the disease, whereas exposure to the statistics about disease prevalence or mortality rates may increase one's *perceived susceptibility* to the disease. In general, the focus on negative information on social media may increase individuals' level of risk perceptions toward the pandemic.

When an individual faces a health threat, they generate not only cognitive appraisal regarding the level of disease risk but also affective and emotional responses. Our study findings show that perceived susceptibility and severity are directly and indirectly associated with depression as a result of worry due to COVID-19. These findings are supported by the Appraisal Theory, which advocated that appraising an event as highly relevant and influential to one's well-being leads to an emotional and affective response [40,41]. These findings concurred with previous studies, which showed that negative emotions during the COVID-19 outbreak could be amplified by misinformation fueled with rumors about the severity of the disease [54]. The perception about increased susceptibility and disease severity may mislead the public and increase uncontrolled panic associated with COVID-19 [44,45,55,56]. Our findings are also consistent to the extant literature that individuals who are exposed to excessive information about the harmful effects of a health issue might experience higher levels of health-related anxiety [57,58].

Study Implications

Findings from this study suggest that the mental health of pregnant women during the COVID-19 pandemic warrants special attention. Screening for mental health problems, continuous monitoring of mental health status, and provision of psychological support throughout the pregnancy during a pandemic are highly warranted. Furthermore, these findings have raised the potential impact of social media in shaping risk perceptions and negative mental response among pregnant women. As social media has become one of the most important sources of health information during the COVID-19 pandemic, there is an urgent need to formulate strategies to minimize the potential negative effect that its use may have for pregnant women. It is suggested that accurate information and effective communication can be valuable to reduce misperception of risk, fear, and negative reactions toward the pandemic. It is important that appropriate social media strategies are developed to counter misinformation or negative information, and to ensure the credibility and accuracy of information shared during this period. Interventions to detect and counter inaccurate information about the media would also be important to reduce its negative impact.

Findings of our study also call for the need for intervention to guide pregnant women regarding the proper use of social media for health information. Alarming evidence suggests that most pregnant women perceive health information available on the internet to be reliable and that they rarely discuss the information with their physicians or midwives [18,20]. Without proper guidance, using social media for health information may lead to harmful consequences, such as information overload, or consumption of unreliable or misleading information [16,59,60]. Interventions are thus needed to empower pregnant women with the skills to identify credible source for obtaining health information, to provide thoughtful consideration of the veracity and quality of health information, and to process the information in an objective manner. Previous studies have also shown that people are likely to absorb negative information and react emotionally on social media. It is important to educate them about the potential bias that may occur in social media, and how these will affect their mental health during the pandemic. They should also be guided to manage their negative emotions, which may be elicited by exposure to stressful information and how to seek social support when they encounter stress as a result of social media use.

Limitations

This study is subjected to several limitations. First, this study was cross-sectional in nature, so causality between the variables cannot be assumed. Nevertheless, it is important to note that the hypothesized association between social media use for health information, perceived susceptibility, perceived severity, worry due to COVID-19, and depression made theoretical sense. The cross-sectional nature of the study also precluded the opportunity to investigate change in the study variables. Second, only pregnant women from several provinces of China were recruited in this study; hence, the sample may not be generalizable to the whole population of pregnant women in China. Third, since no validated measures for measuring social media use for health information, perceived susceptibility, perceived severity, and worry related to COVID-19 were available, items were self-developed with reference to previous studies on other pandemics. The validity of survey items should therefore be cautioned. Fourth, as no information about those who did not participate in the study was available, no comparison between respondents and nonrespondents could be made. Finally, as the current model was based on the Appraisal Theory that highlights the important role of cognitive appraisal and resulting emotions, only cognitive and emotional factors were included in the study; other factors of depression, such as media literacy, resilience, confidence in fighting against the pandemic, and social support, have not been considered. Future studies could include a broader range of factors from different perspectives to allow a better understanding on the role of social media on mental health among pregnant women.

Conclusions

Despite the limitations, given the scarcity of studies on the role of social media use for health information and mental health among pregnant women during the COVID-19 pandemic and the limited application of theoretical frameworks in understanding the topic, we believe that the findings of this study would provide valuable insights into the potential mental impact of social media use on mental health of pregnant women. This study shows that more than one-third of pregnant women surveyed reported that they always used social media for obtaining health information during the COVID-19 pandemic. Using social media for health information was indirectly associated with depression, based on our analyses of perceived susceptibility, perceived severity, and worry due to COVID-19. With the growing popularity of social media as a source of health information, interventions are needed to equip pregnant women with the skills to properly identify and access useful information from social media, as well as to educate them about the potential negative impact that social media use may pose to their health.

Conflicts of Interest

None declared.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
2. World Health Organization. COVID-19 Weekly Epidemiological Update. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2021-02-14]
3. Jamieson DJ, Honein MA, Rasmussen SA, Williams JL, Swerdlow DL, Biggerstaff MS, et al. H1N1 2009 influenza virus infection during pregnancy in the USA. The Lancet 2009 Aug;374(9688):451-458. [doi: 10.1016/s0140-6736(09)61304-0]
4. Yang H, Wang C, Poon LC. Novel coronavirus infection and pregnancy. Ultrasound Obstet Gynecol 2020 Apr;55(4):435-437 [FREE Full text] [doi: 10.1002/uog.22006] [Medline: 32134165]
5. Castro P, Matos AP, Werner H, Lopes FP, Tonni G, Araujo Júnior E. Covid-19 and pregnancy: an overview. Rev Bras Ginecol Obstet 2020 Jul 19;42(7):420-426 [FREE Full text] [doi: 10.1055/s-0040-1713408] [Medline: 32559801]
6. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet 2020 Mar 14;395(10227):912-920. [doi: 10.1016/s0140-6736(20)30460-8]
7. Saulnier DD, Brolin K. A systematic review of the health effects of prenatal exposure to disaster. Int J Public Health 2015 Nov 23;60(7):781-787. [doi: 10.1007/s00038-015-0699-2] [Medline: 26298438]
8. Braunack-Mayer A, Tooher R, Collins JE, Street JM, Marshall H. Understanding the school community's response to school closures during the H1N1 2009 influenza pandemic. BMC Public Health 2013 Apr 15;13(1):344 [FREE Full text] [doi: 10.1186/1471-2458-13-344] [Medline: 23587175]
9. Poon L, Yang H, Kapur A, Melamed N, Dao B, Divakar H, et al. Global interim guidance on coronavirus disease 2019 (COVID-19) during pregnancy and puerperium from FIGO and allied partners: Information for healthcare professionals. Int J Gynaecol Obstet 2020 Jun;149(3):273-286. [doi: 10.1002/ijgo.13156] [Medline: 32248521]
10. Monk C, Lugo-Candelas C, Trumpff C. Prenatal developmental origins of future psychopathology: mechanisms and pathways. Annu Rev Clin Psychol 2019 May 07;15(1):317-344 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050718-095539] [Medline: 30795695]
11. Lee DT, Sahota D, Leung TN, Yip AS, Lee FF, Chung TK. Psychological responses of pregnant women to an infectious outbreak: a case-control study of the 2003 SARS outbreak in Hong Kong. J Psychosom Res 2006 Nov;61(5):707-713 [FREE Full text] [doi: 10.1016/j.jpsychores.2006.08.005] [Medline: 17084150]

12. Ng J, Sham A, Tang PL, Fung S. SARS: pregnant women's fears and perceptions. Br J Midwifery 2004 Nov;12(11):698-702. [doi: 10.12968/bjom.2004.12.11.16710]
13. Davenport MH, Meyer S, Meah VL, Strynadka MC, Khurana R. Moms are not OK: COVID-19 and maternal mental health. Front Glob Womens Health 2020 Jun 19;1:1. [doi: 10.3389/fgwh.2020.00001] [Medline: 34816146]
14. Hessami K, Romanelli C, Chiurazzi M, Cozzolino M. COVID-19 pandemic and maternal mental health: a systematic review and meta-analysis. J Matern Fetal Neonatal Med 2020 Nov 01:1-8. [doi: 10.1080/14767058.2020.1843155] [Medline: 33135523]
15. Durankuş F, Aksu E. Effects of the COVID-19 pandemic on anxiety and depressive symptoms in pregnant women: a preliminary study. J Matern Fetal Neonatal Med 2020 May 18:1-7. [doi: 10.1080/14767058.2020.1763946] [Medline: 32419558]
16. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
17. Siliquini R, Ceruti M, Lovato E, Bert F, Bruno S, De Vito E, et al. Surfing the internet for health information: an italian survey on use and population choices. BMC Med Inform Decis Mak 2011 Apr 07;11(1):21-29 [FREE Full text] [doi: 10.1186/1472-6947-11-21] [Medline: 21470435]
18. Gao L, Larsson M, Luo S. Internet use by Chinese women seeking pregnancy-related information. Midwifery 2013 Jul;29(7):730-735. [doi: 10.1016/j.midw.2012.07.003] [Medline: 22958935]
19. Javanmardi M, Noroozi M, Mostafavi F, Ashrafi-Rizi H. Internet usage among pregnant women for seeking health information: a review article. Iran J Nurs Midwifery Res 2018;23(2):79-86 [FREE Full text] [doi: 10.4103/ijnmr.IJNMR_82_17] [Medline: 29628953]
20. Larsson M. A descriptive study of the use of the internet by women seeking pregnancy-related information. Midwifery 2009 Feb;25(1):14-20. [doi: 10.1016/j.midw.2007.01.010] [Medline: 17408822]
21. Lagan B, Sinclair M, Kernohan W. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010;37(2):106-115. [doi: 10.1111/j.1523-536x.2010.00390.x]
22. Bert F, Gualano MR, Brusaferro S, De Vito E, de Waure C, La Torre G, et al. Pregnancy e-health: a multicenter Italian cross-sectional study on Internet use and decision-making among pregnant women. J Epidemiol Community Health 2013 Dec 01;67(12):1013-1018. [doi: 10.1136/jech-2013-202584] [Medline: 24072743]
23. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16(1):65-10 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
24. De Santis M, De Luca C, Quattrocchi T, Visconti D, Cesari E, Mappa I, et al. Use of the internet by women seeking information about potentially teratogenic agents. Eur J Obstet Gynecol Reprod Biol 2010 Aug;151(2):154-157. [doi: 10.1016/j.ejogrb.2010.04.018] [Medline: 20478650]
25. Pan Y, Xin M, Zhang C, Dong W, Fang Y, Wu W, et al. Associations of mental health and personal preventive measure compliance with exposure to COVID-19 information during work resumption following the COVID-19 outbreak in China: cross-sectional survey study. J Med Internet Res 2020 Oct 08;22(10):e22596 [FREE Full text] [doi: 10.2196/22596] [Medline: 32936776]
26. Cuello-Garcia C, Pérez-Gaxiola G, van Amelsvoort L. Social media can have an impact on how we manage and investigate the COVID-19 pandemic. J Clin Epidemiol 2020 Nov;127:198-201 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.06.028] [Medline: 32603686]
27. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
28. Seo M. Amplifying Panic and Facilitating Prevention: Multifaceted Effects of Traditional and Social Media Use During the 2015 MERS Crisis in South Korea. Journalism & Mass Communication Quarterly 2019 Jul 26;98(1):221-240. [doi: 10.1177/1077699019857693]
29. Choi D, Yoo W, Noh G, Park K. The impact of social media on risk perceptions during the MERS outbreak in South Korea. Comput Human Behav 2017 Jul;72:422-431 [FREE Full text] [doi: 10.1016/j.chb.2017.03.004] [Medline: 32288176]
30. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]
31. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020 Apr 16;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
32. Riello M, Purgato M, Bove C, MacTaggart D, Rusconi E. Prevalence of post-traumatic symptomatology and anxiety among residential nursing and care home workers following the first COVID-19 outbreak in Northern Italy. R Soc Open Sci 2020 Sep 30;7(9):200880 [FREE Full text] [doi: 10.1098/rsos.200880] [Medline: 33047047]
33. Hadjistavropoulos HD, Craig KD, Hadjistavropoulos T. Cognitive and behavioral responses to illness information: the role of health anxiety. Behav Res Ther 1998 Feb;36(2):149-164. [doi: 10.1016/s0005-7967(98)00014-x]

34. Garfin D, Silver R, Holman E. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
35. Liu X, Lo V. Media exposure, perceived personal impact, and third-person effect. Media Psychology 2014 Apr 25;17(4):378-396. [doi: 10.1080/15213269.2013.826587]
36. Determann D, de Bekker-Grob EW, French J, Voeten HA, Richardus JH, Das E, et al. Future pandemics and vaccination: public opinion and attitudes across three European countries. Vaccine 2016 Feb 03;34(6):803-808 [FREE Full text] [doi: 10.1016/j.vaccine.2015.12.035] [Medline: 26740251]
37. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec 30;25(8):661-669. [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
38. Sengupta S, Wang H. Information sources and adoption of vaccine during pandemics. Int J Pharm Healthc Mark 2014;8(4):357-370. [doi: 10.1108/ijphm-01-2014-0002]
39. Pérez-Fuentes MDC, Molero Jurado MDM, Martos Martínez Á, Gázquez Linares JJ. Threat of COVID-19 and emotional state during quarantine: positive and negative affect as mediators in a cross-sectional study of the Spanish population. PLoS One 2020 Jun 25;15(6):e0235305 [FREE Full text] [doi: 10.1371/journal.pone.0235305] [Medline: 32584897]
40. Smith CA, Ellsworth PC. Patterns of cognitive appraisal in emotion. J Pers Soc Psychol 1985;48(4):813-838. [doi: 10.1037/0022-3514.48.4.813]
41. Smith CA, Kirby LD. Putting appraisal in context: toward a relational model of appraisal and emotion. Cognition & Emotion 2009 Nov;23(7):1352-1372. [doi: 10.1080/02699930902860386]
42. Ding Y, Xu J, Huang S, Li P, Lu C, Xie S. Risk perception and depression in public health crises: evidence from the COVID-19 crisis in China. Int J Environ Res Public Health 2020 Aug 07;17(16):5728 [FREE Full text] [doi: 10.3390/ijerph17165728] [Medline: 32784792]
43. Han Q, Zheng B, Agostini M, Bélanger JJ, Gützkow B, Kreienkamp J, et al. Associations of risk perception of COVID-19 with emotion and mental health during the pandemic. J Affect Disord 2021 Apr 01;284:247-255 [FREE Full text] [doi: 10.1016/j.jad.2021.01.049] [Medline: 33602537]
44. Tran BX, Nguyen HT, Le HT, Latkin CA, Pham HQ, Vu LG, et al. Impact of COVID-19 on economic well-being and quality of life of the Vietnamese during the national social distancing. Front Psychol 2020 Sep 11;11:565153 [FREE Full text] [doi: 10.3389/fpsyg.2020.565153] [Medline: 33041928]
45. Tee ML, Tee CA, Anlacan JP, Aligam KJG, Reyes PWC, Kuruchittham V, et al. Psychological impact of COVID-19 pandemic in the Philippines. J Affect Disord 2020 Dec 01;277:379-391 [FREE Full text] [doi: 10.1016/j.jad.2020.08.043] [Medline: 32861839]
46. Bian C, Li C, Duan Q, Wu H. Reliability and validity of patient health questionnaire: depressive syndrome module for outpatients. Scientific Research and Essays 2011;6(2):278-282.
47. Zhang Y, Ting R, Lam M, Lam J, Nan H, Yeung R, et al. Measuring depressive symptoms using the Patient Health Questionnaire-9 in Hong Kong Chinese subjects with type 2 diabetes. J Affect Disord 2013 Nov;151(2):660-666. [doi: 10.1016/j.jad.2013.07.014] [Medline: 23938133]
48. Li Z, Dai J, Wu N, Gao J, Fu H. The mental health and depression of rural-to-urban migrant workers compared to non-migrant workers in Shanghai: a cross-sectional study. Int Health 2019 Oct 31;11(S1):S55-S63. [doi: 10.1093/inthealth/ihz081] [Medline: 31670817]
49. Kline R. Principles and practice of structural equation modeling (2nd edition). New York: Guilford; 2005.
50. Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Prevalence of depression during pregnancy: systematic review. Obstetrics & Gynecology 2004;103(4):698-709. [doi: 10.1097/01.aog.0000116689.75396.5f]
51. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
52. Chang C. News coverage of health-related issues and its impacts on perceptions: Taiwan as an example. Health Commun 2012;27(2):111-123. [doi: 10.1080/10410236.2011.569004] [Medline: 21843098]
53. Chung JE. A smoking cessation campaign on Twitter: understanding the use of Twitter and identifying major players in a health campaign. J Health Commun 2016 May;21(5):517-526. [doi: 10.1080/10810730.2015.1103332] [Medline: 27115179]
54. Islam M, Sarkar T, Khan S, Mostofa Kamal AH, Hasan S, Kabir A, et al. COVID-19-related infodemic and its impact on public health: a global social media analysis. Am J Trop Med Hyg 2020 Oct;103(4):1621-1629 [FREE Full text] [doi: 10.4269/ajtmh.20-0812] [Medline: 32783794]
55. Ho CS, Chee CY, Ho RC. Mental health strategies to combat the psychological impact of coronavirus disease 2019 (COVID-19) beyond paranoia and panic. Ann Acad Med Singap 2020 Mar 31;49(3):155-160. [doi: 10.47102/annals-acadmedsg.202043]
56. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]

57. Blackburn J, Fischerauer SF, Talaei-Khoei M, Chen NC, Oh LS, Vranceanu A. What are the implications of excessive internet searches for medical information by orthopaedic patients? Clin Orthop Relat Res 2019 Jul 31;477(12):2665-2673. [doi: 10.1097/corr.0000000000000888]
58. Norr AM, Capron DW, Schmidt NB. Medical information seeking: impact on risk for anxiety psychopathology. J Behav Ther Exp Psychiatry 2014 Sep;45(3):402-407. [doi: 10.1016/j.jbtep.2014.04.003] [Medline: 24818986]
59. Skinner H, Biscope S, Poland B, Goldberg E. How adolescents use technology for health information: implications for health professionals from focus group studies. J Med Internet Res 2003 Dec 18;5(4):e32 [FREE Full text] [doi: 10.2196/jmir.5.4.e32] [Medline: 14713660]
60. Cline R, Haynes K. Consumer health information seeking on the internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]

Abbreviations

CFI: comparative fit index
IFI: incremental fit index
MERS: Middle East respiratory syndrome
QR: quick response
RMSEA: root mean square error of approximation
SARS: severe acute respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SEM: structural equation modeling

Edited by C Lovis; submitted 24.02.21; peer-reviewed by X Hu, P Zheng; comments to author 05.04.21; revised version received 22.04.21; accepted 12.07.21; published 13.01.22.

Please cite as:
Wang Q, Xie L, Song B, Di J, Wang L, Mo PKH
Effects of Social Media Use for Health Information on COVID-19–Related Risk Perceptions and Mental Health During Pregnancy: Web-Based Survey
JMIR Med Inform 2022;10(1):e28183
URL: https://medinform.jmir.org/2022/1/e28183
doi:10.2196/28183
PMID:34762065

Original Paper

Current Status of the Health Information Technology Industry in China from the China Hospital Information Network Conference: Cross-sectional Study of Participating Companies

Zhongan Zhang[1*], BS; Xu Zheng[2,3*], MSc; Kai An[2,3], MSc; Yunfan He[4], BS; Tong Wang[5], BS; Ruizhu Zhou[6], BS; Qilin Zheng[7], BS; Mingfu Nuo[8], MS; Jun Liang[9,10], MS; Jianbo Lei[3,8,11], MD, PhD

[1]Department of Information Center, Affiliated Qingdao Central Hospital, Qingdao University, Qingdao, China

[2]Peking University Third Hospital, Beijing, China

[3]Center for Medical Informatics, Health Science Center, Peking University, Beijing, China

[4]School of Public Health, Zhejiang University, Hangzhou, China

[5]Department of Medical Informatics, School of Public Health, Jilin University, Changchun, China

[6]Department of Applied Mathematics, Northwestern Polytechnical University, Xian, China

[7]Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

[8]Institute of Medical Technology, Health Science Center, Peking University, Beijing, China

[9]Information Technology Center, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[10]Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education, School of Medicine, Zhejiang University, Hangzhou, China

[11]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, China

[*]these authors contributed equally

Corresponding Author:
Jianbo Lei, MD, PhD
Center for Medical Informatics
Health Science Center
Peking University
38 Xueyuan Rd
Haidian District
Beijing, 100191
China
Phone: 86 15201230935
Email: jblei@hsc.pku.edu.cn

Abstract

Background: The China Hospital Information Network Conference (CHINC) is one of the most influential academic and technical exchange activities in medical informatics and medical informatization in China. It collects frontier ideas in medical information and has an important reference value for the analysis of China's medical information industry development.

Objective: This study summarizes the current situation and future development of China's medical information industry and provides a future reference for China and abroad in the future by analyzing the characteristics of CHINC exhibitors in 2021.

Methods: The list of enterprises and participating keywords were obtained from the official website of CHINC. Basic characteristics of the enterprises, industrial fields, applied technologies, company concepts, and other information were collected from the TianYanCha website and the VBDATA company library. Descriptive analysis was used to analyze the collected data, and we summarized the future development directions.

Results: A total of 205 enterprises officially participated in the exhibition. Most of the enterprises were newly founded, of which 61.9% (127/205) were founded in the past 10 years. The majority of these enterprises were from first-tier cities, and 79.02% (162/205) were from Beijing, Zhejiang, Guangdong, Shanghai, and Jiangsu Provinces. The median registered capital is 16.67 million RMB (about US $2.61 million), and there are 35 (72.2%) enterprises with a registered capital of more than 100 million RMB (about US $15.68 million), 17 (8.3%) of which are already listed. A total of 126 enterprises were found in the VBDATA company library, of which 39 (30.9%) are information technology vendors and 57 (45.2%) are application technology vendors.

In addition, 16 of the 57 (28%) use artificial intelligence technology. Smart medicine and internet hospitals were the focus of the enterprises participating in this conference.

Conclusions: China's tertiary hospital informatization has basically completed the construction of the primary stage. The average grade of hospital electronic medical records exceeds grade 3, and 78.13% of the provinces have reached grade 3 or above. The characteristics are as follows: On the one hand, China's medical information industry is focusing on the construction of smart hospitals, including intelligent systems supporting doctors' scientific research, diagnosis-related group intelligent operation systems, and office automation systems supporting hospital management, single-disease clinical decision support systems assisting doctors' clinical care, and intelligent internet of things for logistics. On the other hand, the construction of a compact county medical community is becoming a new focus of enterprises under the guidance of practical needs and national policies to improve the quality of grassroots health services. In addition, whole-course management and digital therapy will also become a new hotspot in the future.

(JMIR Med Inform 2022;10(1):e33600) doi:10.2196/33600

KEYWORDS

medical informatics; China Hospital Information Network Conference; industry analysis; county medical community; smart hospital; cross-sectional study; digital therapeutic; information network; health care; hospital information; medical information; tertiary hospital

Introduction

With the Chinese government's strong push for health care reform in 2009, the informatization construction of China's tertiary hospitals has basically completed the primary stage of popularization. The Chinese government first proposed taking health information technology (HIT) as the key direction for motivating medical reform in March 2009 [1] and vigorously promoted electronic medical records (EMRs). After 20 years of construction, the informatization of Chinese hospitals has made phased achievements. The Hospital Management Research Institute of the National Health Commission issued the new edition of evaluation criteria and management measures, which divided the application level of the EMR system into 9 levels ranging from 0 to 8 in December 2018. The Chinese government required that all tertiary hospitals reach grade 3 or above by the end of 2019 and that all tertiary hospitals reach grade 4 or above, while secondary hospitals reach grade 3 or above by the end of 2020 [2]. In 2019, 7870 medical institutions completed the graded evaluation of the application level of the EMR system, and the average level was 2.08. A total of 1874 tertiary hospitals participated in the evaluation, with an overall participation rate of 99.36%, and the average level exceeded grade 3 [3]. In addition, 34% of tertiary hospitals and 24.3% of secondary hospitals received level 5 or above. There were 0 institutions that received level 8, 4 institutions that received level 7, 19 institutions that received level 6, and 100 institutions that received level 5 [4]. It can be said that China's hospital informatization construction has completed the infrastructure construction stipulated by the National Health Commission and is facing the initial stage of digital transformation. This year, the China Hospital Information Network Conference (CHINC) 2021 was held during this special period.

The purpose of the conference is mainly communication, and it is also the most important way to understand the current situation of a country's industry. Conferences on medical informatization can be divided into two categories: academic and industrial. The most famous academic conference is the American Medical Informatics Association (AMIA) annual symposium. In 2020, more than 2100 people attended the online conference, involving 111 academic topics [5]. In addition, the most famous industrial conference is the Healthcare Information and Management Systems Society (HIMSS) conference, with more than 45,000 participants and 1300 enterprises. There are 4 well-known conferences in China, including two academic conferences (the Chinese Medicine Information Association Annual Symposium [CMIAAS] and the China Proceedings of Medical Informatics [CPMI]), and two industrial conferences (CHINC and the China Health Information Technology Exchange Conference [CHITEC]) [6]. The scale of China’s medical informatization academic conferences is small, with fewer than 1000 participants. Studies have shown that medical informatics conferences in China and the United States have differences and similarities. From the scale point of view, as mentioned above, even the largest CHINC in China has only half the number of participants as the HIMSS. From the perspective of discussion themes, EMRs are the research hotspot and focus shared by medical informatics academia and industry worldwide [7]. In contrast, China is more application oriented: the implementation rate of EMRs in Chinese hospitals has been approaching and surpassing that of the United States in recent years [8], but theoretical research and educational discussions are advanced in the United States [9].

In contrast, the scale of industrial conferences is much larger. CHITEC lasted for 2 days, and the number of participants reached 230,000 in 2020. CHINC introduced in this paper has a larger scale, a longer duration, and more submeetings compared with CHITEC.

Chinese hospitals are in the transition period of informatization and digitization. Understanding the research direction in the next stage is of great guiding significance for developing the medical information field. Therefore, CHINC, which involves many cutting-edge ideas, plays a special role in the field of medical information at this stage. CHINC is sponsored by the Institute of Hospital Management of the National Health Commission and has been held annually in China since 1995. It is one of the most influential academic and technical exchange activities in the field of medical informatics and medical

informatization in China [6]. At present, it has successfully been held 25 times [10]. With the country's increasing attention toward public health and intelligent medical care, CHINC has attracted increasing attention. Only 6000 people attended conferences in 2016, but 17,000 people attended conferences in 2020. In addition, the number reached approximately 40,000 in 2021. Moreover, 2021 is the first year of the 14th 5-year plan. To implement the new requirements of the 14th 5-year plan and the Healthy China strategy for hospital construction and development, more than 400 experts gave wonderful lectures in 69 forums and academic activities, and 207 cooperative enterprises held roadshows to exchange and discuss new technologies, new achievements, and new experiences in hospital information construction to help the high-quality development of hospitals [11].

The main characteristics of CHINC include the organizer, history, cycle, holding time, number of participants, participating manufacturers, and conference forum. The number of participating enterprises and the main business of the enterprises are important factors reflecting the current situation of the industry. Therefore, we extracted the features of the enterprises participating in CHINC 2021. We analyzed the main concerns of the enterprises, gained insight into the current situation of China's medical information industry, and defined the future development direction. The conclusion can be used for reference by relevant experts in China and abroad.

Methods

Data Collection

First, the list of all participating enterprises (including enterprise name, exhibition booth, and keywords; Figure 1) was obtained from the conference's official website [10]. We compared the list with the on-site list on the participation day one by one to exclude enterprises that did not attend the conference. Second, we used the TianYanCha website [12] to search all exhibitors and obtain basic information about the enterprises, including a brief introduction, region, establishment time, personnel scale, financing rounds, registered capital, and listing. Finally, we used the VBDATA company library [13] to obtain deep-seated information, such as the industrial field, application technology, and company concept of each enterprise. All data collection was completed by May 6, 2021.

Figure 1. Information about enterprises provided on the official CHINC website. CHINC: China Hospital Information Network Conference.

Data Storage and Analysis

We used Microsoft Excel 2019 for data storage and analysis. Percentages, bar charts, Venn charts, and statistical charts were used to display the exhibitors' basic information, industrial and commercial information, classification, and grade data. Percentiles, medians, and quartile ranges were used to describe skew continuity data. We analyzed the data results according to the actual development of China's medical information industry.

Results

Exhibitors' Characteristics

A total of 207 cooperative enterprises were listed on the official website of CHINC. After checking each enterprise one by one on the conference day, the results indicated that 205 enterprises attended the conference. The participating enterprises were established from 1987 to 2021. A total of 61.9% (127/205) of enterprises were established from 2009, and 43.3% (55/127) of them were established from 2015 (Figure 2). A total of 79.02% (162/205) of the enterprises were from Beijing, Zhejiang, Guangdong, Shanghai, and Jiangsu Provinces, and 2 of them were from New Zealand (Figure 3). The enterprises' scale is shown in Figure 4. Most enterprises (62/205, 30.02%) have fewer than 50 members, followed by 54/205 (34%) enterprises with 101-500 members.

Figure 2. Establishment year of enterprises, displayed by year.

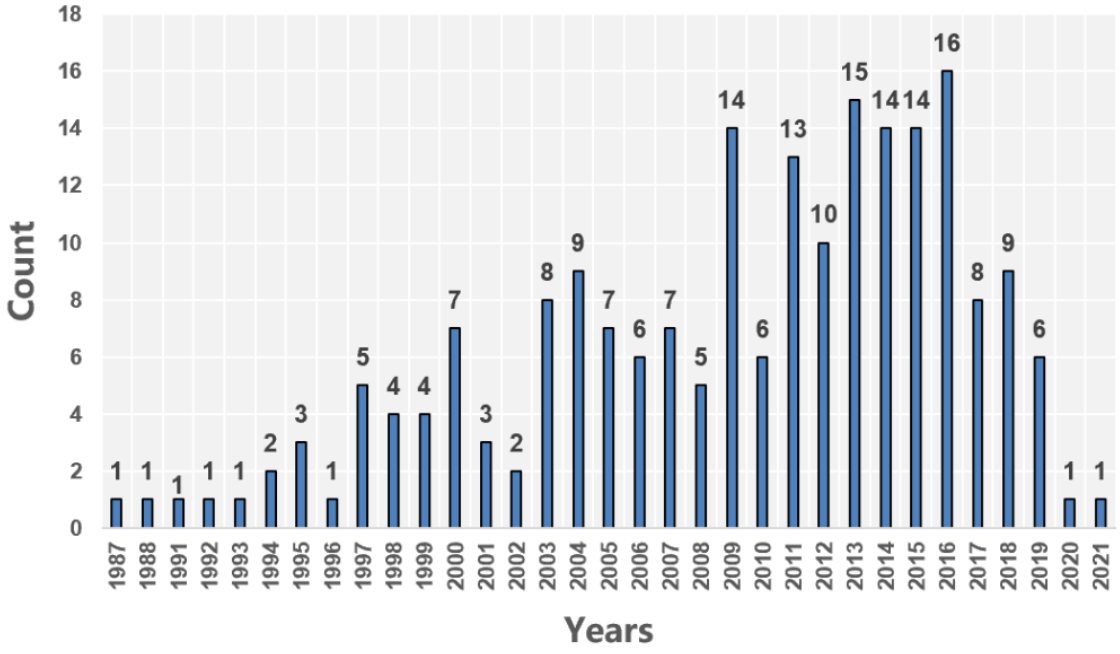


Figure 3. Establishment location of enterprises, displayed by province.

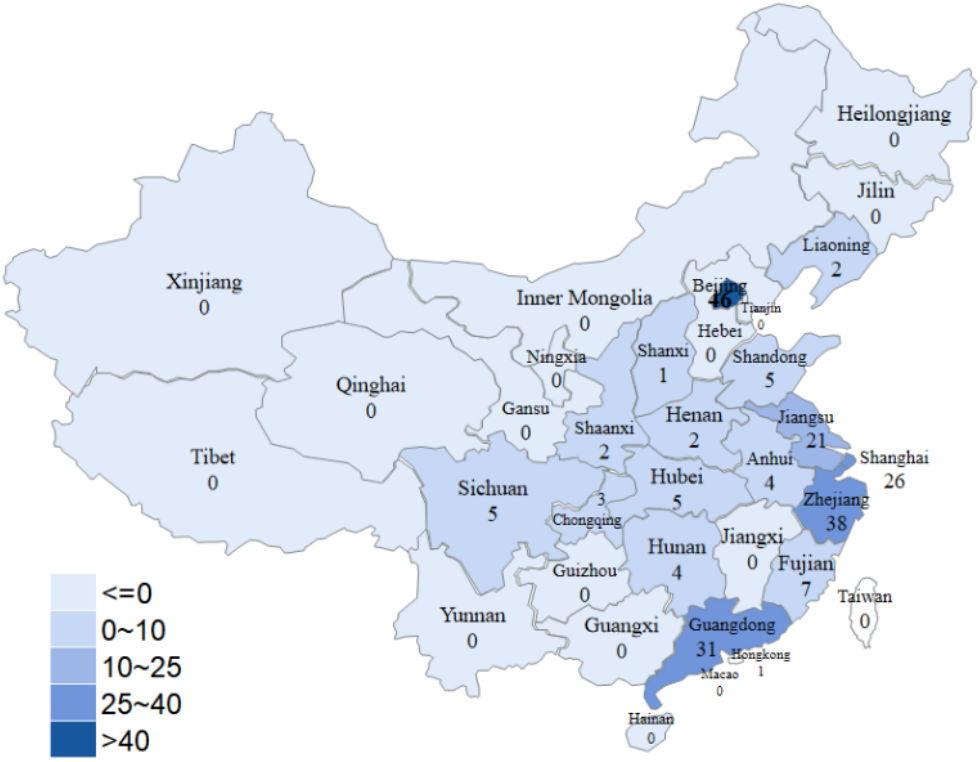
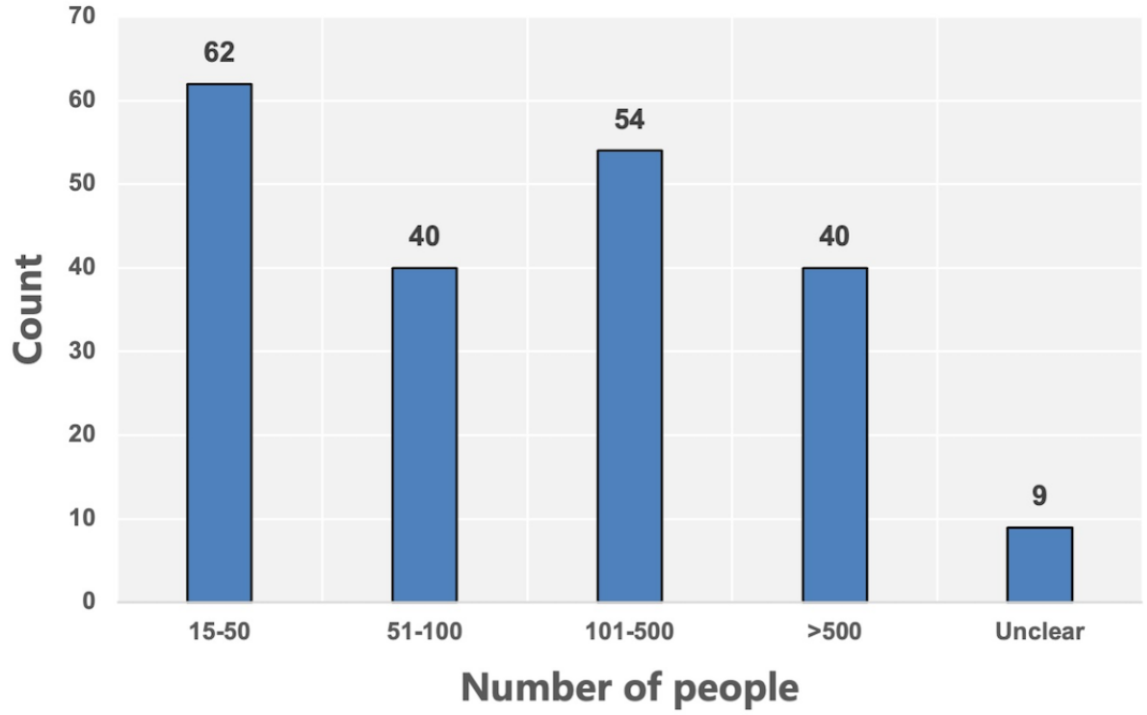


Figure 4. Establishment size of enterprises, displayed by scale.



Industrial and Commercial Information Analysis

The median registered capital of the 205 enterprises participating in the conference is 16.67 million RMB (a currency exchange rate of RMB 1=US $0.16 is applicable; IQR 10-60 million RMB, maximum value=30 billion RMB; Table 1). In addition, 35 (17.1%) enterprises have a registered capital of more than 100 million RMB; see Table 2. Furthermore, 99 (48.3%) enterprises provided their financing information, of which 17 (8.3%) were initial public offerings (Figure 5).

Table 1. Registered capital of enterprises (N=205).

Variable	Minimum	P5[a]	P25[a]	P50[a]	P75[a]	P95[a]	Maximum
Enterprise registered capital (10,000 RMB[b])	100	200	1000	1667	6000	86,941.15	30,000,000

[a]Px: Percentile occupied by the x-th position.

[b]A currency exchange rate of RMB 1=US $0.16 is applicable.

Table 2. Information about enterprises with more than 100 million RMB[a] of registered capital.

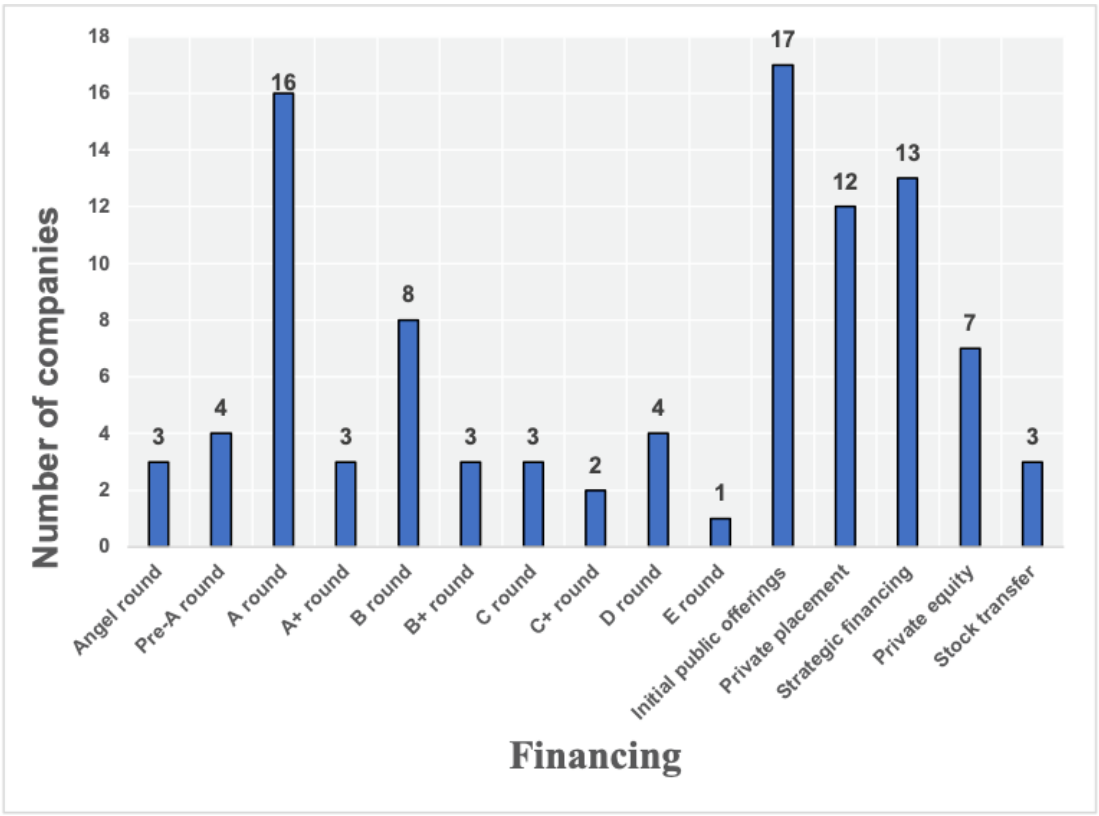
Company name	Brief introduction	Area	Date of establishment	Registered capital/10,000 RMB	Scale, n
Hangzhou Century Co., Ltd.	Smart health care service provider	Zhejiang	November 21, 2003	86,941.15	>500
DHC MediWay Technology Co., Ltd	IT service provider in the big health field	Beijing	May 1, 2012	13,000.00	101-500
B-Soft Co.,Ltd	Hospital information platform provider	Zhejiang	December 10, 1997	110,962.71	>500
Beijing Lenovo Wisdom Medical Information Technology Co., Ltd	Integrated smart medical solution provider	Beijing	February 6, 2016	19,246.86	101-500
Huawei Technologies Co.,Ltd	The world's leading information and communication (ICT) infrastructure and smart terminal provider	Guangdong	September 15, 1987	4,030,813.18	>500
Sangfor Technologies Inc	Security and cloud computing solution provider	Guangdong	December 25, 2000	40,901.47	>500
Goodwill Information Technology Co., Ltd	Information system development, sales, and service provider	Beijing	July 13, 2005	51,000.00	>500
Mediinfo I.t.Co.,Ltd	Medical information service provider	Zhejiang	September 6, 1999	10,224.00	101-500
Alibaba(China)Network Technology Co., Ltd	—[b]	Zhejiang	September 9, 1999	6,942,460.80	>500
Unicom (Guangdong) Industrial internet Co., Ltd	—	Beijing	June 18, 1994	10,481,551.96	>500
Neusoft Corporation	Internet and software product and service provider	Liaoning	June 17, 1991	124,237.03	>500
Winning Health Technology Group Co., Ltd	Medical and health information solution provider	Shanghai	April 7, 2004	164,100.58	>500
China Mobile Communications Group Co., Ltd	—	Beijing	July 22, 1999	30,000,000.00	>500
Ruijie Networks Co., Ltd	Informatization solution provider, China's leading brand of data communication solutions	Fujian	October 28, 2003	50,000.00	>500
Shanghai KingYee Information Technology Co., Ltd	Smart medical technology and service provider	Shanghai	August 7, 2009	11,776.88	>500
Heren Health Co., Ltd	Medical information provider	Zhejiang	October 26, 2010	11,719.05	>500
Dnake (Xiamen) Intelligent Technology Co., Ltd	Smart hardware developer	Fujian	April 29, 2005	12,000.00	>500
Baidu Online Network Technology (Beijing) Co., Ltd	Informatization solution provider in the medical field	Beijing	January 18, 2000	29,257.96	>500
Bringspring Science and Technology Co.,Ltd	Smart city, smart medical solutions, data center integration and operation, and maintenance services, financial IT outsourcing service provider	Liaoning	February 8, 2012	59,752.79	>500
Beijing Tianjian Yuan Da Tecnology Co., Ltd	Professional developer of medical information system	Beijing	August 9, 2005	17,091.03	>500
Lianren Health and Medical Big Data Technology Co., Ltd	Medical big data analysis service provider	Shanghai	November 18, 2019	200,000.00	51-100
Wanma Technology Co., Ltd	Medical information service and hardware provider	Zhejiang	January 28, 1997	13,400.00	>500
Honeywell Integrated Technology (China) Co., Ltd	Aviation products and services, building, home, and industrial control technology, automotive products, turbochargers, and special material R&D and manufacturer	Shanghai	January 1, 1988	19,289.54	>500

Company name	Brief introduction	Area	Date of establishment	Registered capital/10,000 RMB	Scale, n
NSFOCUS Technologies Group Co., Ltd	Enterprise-level information security service provider	Beijing	April 25, 2000	79,967.41	>500
Suzhou MedicalSystem Technology Co., Ltd	Comprehensive solution provider for clinical information systems and digital hospitals	Jiangsu	August 14, 2009	11,245.48	>500
Enjoyor Co.,Ltd	Mobile computing, intelligent identification, data fusion, and other technology developers	Zhejiang	November 13, 1992	65,578.91	>500
China Telecom Corporation Limited	—	Beijing	September 10, 2002	8,093,236.83	>500
Shanghai Aihui Healthy Technology Co., Ltd	Bedside information service provider	Shanghai	September 22, 2016	13,007.04	51-100
Edan Instruments, Inc	R&D, production, sales, and service provider of medical electronic equipment	Guangdong	August 2, 1995	58,172.18	>500
Nexans (Suzhou) cable solution Co., Ltd	—	Jiangsu	April 17, 2013	32,365.00	101-500
Yonyou Network Technology Co., Ltd	Data collection and business application solution provider	Beijing	January 18, 1995	324,872.13	>500
FUJIFILM (China) Investment Co., Ltd	Film R&D producer	Shanghai	April 12, 2001	21,339.70	101-500
Zhejiang Jandar Technology Co., Ltd	Software development, information system, integration service provider	Zhejiang	November 19, 1999	10,000.00	101-500
Newlink Technology Inc	—	Beijing	August 15, 2011	10,203.04	101-500
Dell (China) Company Limited	Electronic equipment manufacturer	Fujian	December 29, 1997	17,186.84	>500

[a]A currency exchange rate of RMB 1=US $0.16 is applicable.

[b]Not available.

Figure 5. Financing of enterprises.



Analysis of Industrial Fields, Applied Technologies, and Related Concepts

The nature of the enterprises, current hot concept directions, and the essence behind it could be analyzed by the keywords of exhibitors provided by the CHINC official website and the industrial field, applied technology, and concept of the enterprises provided by the VBDATA company library.

The conference's official website provided 151 enterprise keywords, which indicated the exhibition direction of the participating enterprises. We created word frequency statistics for the keywords. The keyword “smart hospitals” appeared 41 times at most and “internet hospitals” 22 times. The rest is shown in Table 3. We drew a cloud map according to the keyword frequency.

A total of 126 enterprises were found in the VBDATA company library. All (100%) companies disclosed their industrial field, 57 (45.2%) companies disclosed the technology used, and 114 (90.5%) companies provided the relevant concept labels. In the industry field, the number of information technology vendors was the largest (39/126, 30.9%), followed by EMRs (34/126, 26.9%). The distribution of industrial fields is shown in Figure 6. We also calculated statistics of the technologies used by enterprises. Of the 57 application technology enterprises, 16 (28.1%) use artificial intelligence (AI) technology and are ranked first, and 15 (26.3%) use the internet of things technology and are ranked second. The distribution of other technologies is shown in Figure 7. The VBDATA company library also provides conceptual labels of current mainstream products and innovative technologies of each enterprise, similar to the keywords given on the official CHINC website. We compared the top 15 concepts with the highest frequency by comparative analysis, and the results are shown in Table 3. Both smart medicine and internet hospitals were the focus of enterprises, and big data appeared most in VBDATA, but they were rarely mentioned at this conference.

Table 3. Enterprise concepts of VBDATA and enterprise keywords given by CHINC[a].

Rank	VBDATA concept	Enterprises, n	CHINC enterprise keywords	Appearance, n
1	Big data	34	Smart hospital	41
2	Smart health care	29	Internet hospital	22
3	COVID-19	27	Smart health care	19
4	SaaS	21	Medical community	17
5	Telemedicine	14	Big data	14
6	Cloud computing	14	Electronic medical records	13
7	Internet hospital	9	Smart services	11
8	Medical equipment	9	Artificial intelligence	10
9	AI device	7	Hospital informatization	10
10	Industrial internet	7	Internet of things	10
11	Medical device supplies	6	Medical cluster	9
12	Equipment	6	Integration platform	9
13	Consumer health care	5	Hospital information system	7
14	mHealth	5	Interoperability	7
15	Public health services	5	Internet health care	7

[a]CHINC: China Hospital Information Network Conference.

Figure 6. Industry field of enterprises.

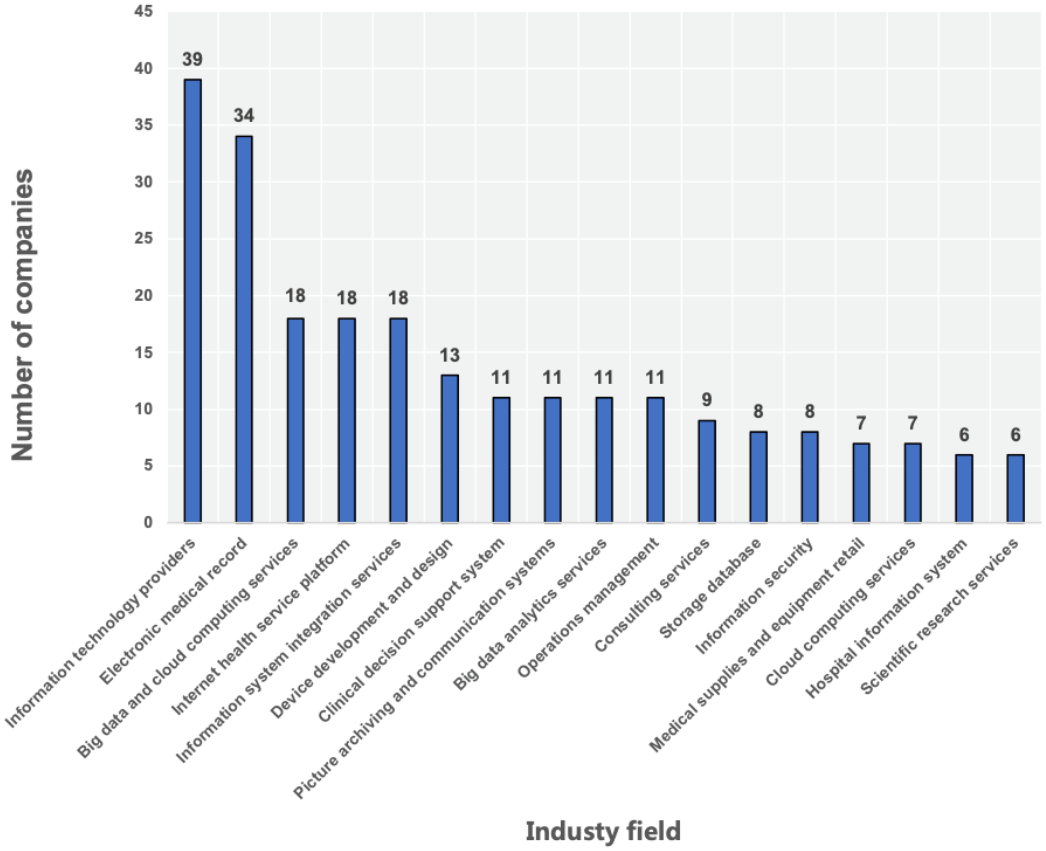
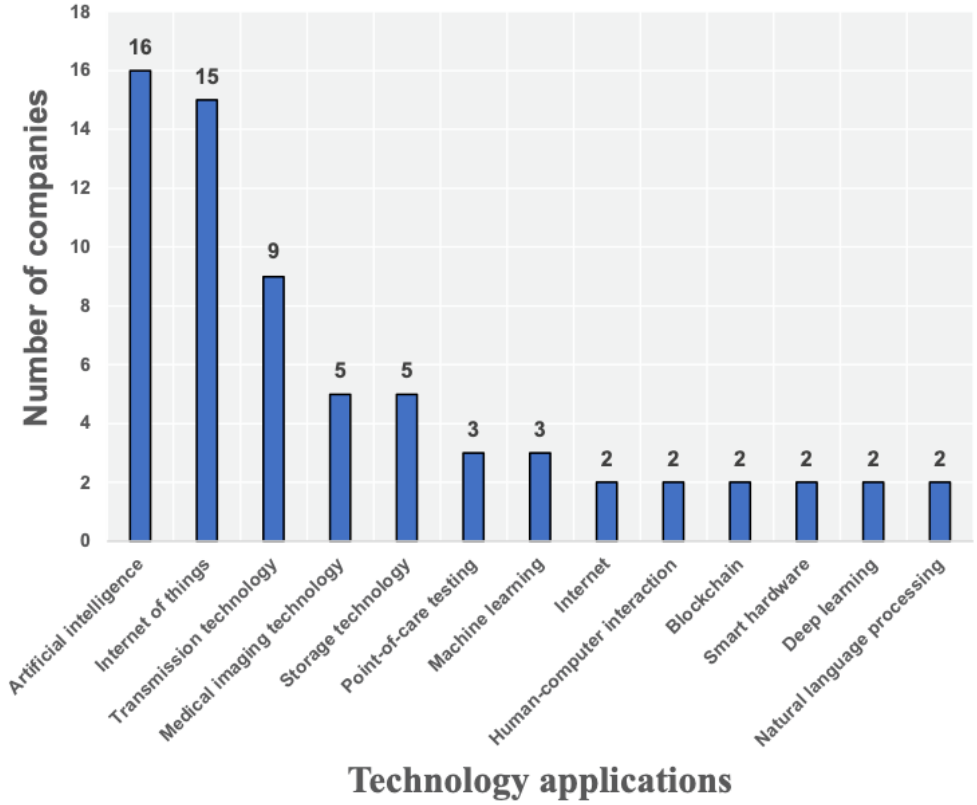


Figure 7. Technologies used by enterprises.



Discussion

Preliminary Results Have Been Achieved in the Process of Hospital Informatization in China

Many information manufacturers have emerged in the past 10 years, and hospital information has basically completed the primary stage of popularization. Our research group has predicted that by 2021, the popularization rate of EMRs in domestic secondary hospitals or higher hospitals may exceed 80%. The popularization rate of EMRs in tertiary hospitals may even exceed 95% [14]. According to official Chinese documents published by the National Health Commission, the rate of tertiary public hospitals participating in grading the EMR application level in 2019 was 99.36%, which confirmed our prediction. From the perspective of enterprises, approximately half of the information-based manufacturers were established in the past 10 years, of which approximately 25% were established in the past 6 years (see the Exhibitors' Characteristics section). With the help of innovative technologies, including AI and the internet of things (see the Analysis of Industrial Fields, Applied Technologies, and Related Concepts section), they have launched information-based solutions for various scenarios in hospitals.

It can be foreseen that the traditional basic content-related market in hospital information construction, for example, hospital information system (HIS), EMRs, laboratory information system (LIS), and picture archiving and communication system (PACS), has been gradually saturated. Therefore, the discussion of traditional hospital information systems and the exhibition of related products at this conference are not particularly ongoing. A new generation of hospital management systems began to appear in 2017, but few hospitals have the courage to carry out thorough information reform [15]. However, the National Health Commission of China has adopted the policy of linking EMR ratings with hospital performance appraisals in the past 3 years. COVID-19 will further strengthen the basic content of hospital informatization construction and gradually encourage new directions of construction. The reform will promote interconnection and high integration between information systems, provide high-quality management and treatment support by using operation data and clinical data, rebuild the hospital management system, and finally complete the digital and intelligent transformation of hospitals [16].

Smart Hospitals Are the Focus of Hospital Construction in the Future

Under the background that the informatization of tertiary hospitals has been basically completed and is steadily being carried out in secondary hospitals, the National Health Commission issued related official documents and further released a document revision in May 2020. At present, the complete definition and construction standards of smart hospitals have not been unified at the national level, and the focus and direction in the exploration of smart hospital construction are not the same among hospitals. The government pointed out that the scope of smart hospitals mainly includes three areas: *smart medical care* for medical personnel, *smart service* for patients, and *smart management* for hospital management. However, it is certain that different from the previous informatization handing over the paper process to the computer for processing, the essence of smart hospitals is to collect, use, and analyze the data inside and outside the hospital. The purpose of smart

hospitals is to provide basic support for hospital scientific research, clinical and management activities, and, finally, feedback to doctors and leaders for decision making.

Meanwhile, the government has put forward strict scoring requirements for the performance appraisal of 3-level public hospitals for several consecutive years. The performance appraisal includes data quality on the first page of EMRs, the application level of EMRs, comparability of clinical tests, the degree of intelligence of medical services, and the equality of rational use of liquid medicine. This is also an important factor in promoting the construction of smart hospitals.

The largest focus of enterprises in this conference is smart hospitals, which can be seen in Figure 6 and Table 3. Smart hospital solutions emerge one after another, mainly including the following points: First, we focus on scientific research systems and advocate data governance. Natural language processing helps the knowledge graph build a hospital special disease database and assist doctors in efficient scientific research to realize intelligent medical treatment. Related projects were carried out by Hangzhou Century Co Ltd., Shanghai Senyi Intelligent Technology Co Ltd., and Anxiang Medical Technology (Shenzhen) Co Ltd. Second, EMR quality control and diagnosis-related group (DRG)/big data diagnosis-intervention packet (DIP) are continuously hot. Automatic coding technology based on AI is introduced to promote exemplary management of hospitals and meet the quality of the first page of EMRs. Companies such as Hangzhou Firetree Technology Co Ltd. and BaseBit AI have designed intelligent operation systems. Third, the clinical decision support system (CDSS) still focuses on a single disease, and a knowledge graph is still the main technology. Deep learning methods are still used in medical image recognition. Companies such as Beijing Shenrui Bolian Technology Co Ltd, Beijing Airdoc Technology Co. Ltd., and Suzhou Mediston Medical Technology Co Ltd. are involved in this type of business. Fourth, we emphasize the ability of the middle platform and build an office automation (OA) system of smart hospitals. We also realize the personalization of different hospitals by using a middle platform and promote the office mode of a new generation of smart hospitals, on which DingTalk advocates and Xiniu Health Technology (Zhejiang) Co Ltd. is focusing. Fifth, the internet of things is hardware-upgraded to ensure intranet security and clinical efficiency. Hardware companies, such as Ruijie Networks Co Ltd. and Onco Information Technology (Shanghai) Co Ltd., have launched hospital dual-network routers based on Wi-Fi 6.0 and 5G to ensure strict internal and external network isolation, realizing intelligent wards.

The year 2021 was the first year of the 14th 5-year plan. Under the guidance of the above policies, the construction of smart hospitals has received more attention from hospitals. Starting from this demand, enterprises at conferences have launched customized smart hospital construction services.

The Construction of a Compact County Medical Community Has Become a New Focus of Enterprises

Compared with other countries, China faces more severe challenges in the distribution of medical services. Although China has the largest number of hospitals globally, the distribution of medical resources is extremely uneven: 80% of medical resources and patients are concentrated in large hospitals and 20% in community general clinics [17]. China introduced the market mechanism into the medical service system in the 1980s, and people can go to any level of medical institutions according to their wishes. Primary medical institutions no longer play the role of health gatekeepers. Many patients give priority to higher-level medical institutions when they need medical assistance. This has led to a large reduction in patients in grassroots hospitals, a decline in the level of grassroots health service personnel [18], and a rapid increase in medical expenditure [19]. The concentration of medical resources in high-level medical institutions further weakens the ability of grassroots health services, resulting in more detours and more waste.

The Chinese government is trying to solve the uneven distribution of medical services by promoting the integration of regional health services. With the Chinese government's strong push for health care reform in 2009, the first contact point between the hierarchical medical service system and grassroots medical institutions was proposed as a key task. The reform strategy notes that the construction of a regional medical consortium is the key to promoting hierarchical diagnosis and treatment [20]. At present, the construction methods of the medical union in China mainly include the Cross-Regional Professional Alliance, the Urban Medical Group, and the County Medical Community. The importance of a compact county medical community is particularly prominent for China’s large rural population. Therefore, carrying out the integration of health services, realizing cooperation between medical institutions at all levels in rural areas, and improving the quality of health services, treatment rate, and patient satisfaction is the fundamental way to truly enable the majority of grassroots residents to obtain health-centered, equal, homogeneous, and integrated health care services [21].

Information construction is an important basis for the construction of compact county medical communities. There are obvious information system breaks between different medical institutions in county medical communities. In the past, the township-level health institution system lacked unified development and business processes, data processes lacked unified norms, and the phenomenon of *information islands* was serious [22]. County medical communities require a high degree of entity integration and information interconnection to ensure the high continuity of medical services and truly realize the original intention of common service, common interests, common responsibilities, and common development.

The market scale of medical community information construction is huge. China officially issued a document to determine Shanxi and Zhejiang Provinces as pilot provinces of the medical community and 567 counties as pilot counties in September 2019. The compact county medical community is developing rapidly, and a unified, efficient, and easy-to-use regional health information system is one of the important information supports. According to national statistics, there are 2843 county-level divisions in China, and the informatization project of the medical community in each county is

approximately 40-80 million. It is roughly estimated that the market scale reaches approximately 113.7-227.4 billion RMB. A new county-level medical and health service system with clear objectives, clear rights and responsibilities, and division of labor and cooperation should be preliminarily established in 500 county-level units, gradually forming a community of services, responsibilities, interests, and management, and finally used throughout China [23]. At this conference, the county medical community informatization solution has become the struggle focus of major medical IT enterprises. Enterprises, such as YLZ Information Technology Co., Ltd., proposed establishing a regional information platform and achieving regional information interconnection, data sharing, and aggregation through unified data standards and service specifications. Big data mining and analysis technology can also be used to conduct intelligent analysis and judgment on operation management, providing intelligent auxiliary services for managers' scientific decision making.

Whole-Course Management and the Concept of Digital Therapy Are New Hotspots and Starting Points of HIT

It is worth mentioning that with many information manufacturers in hospitals and relatively mature solutions, there is a huge potential development space for targeting the market for out-of-hospital medical services, and several new technologies and solutions have emerged. These products are mainly named after the concepts of *special disease bank*, *scientific research follow-up*, *whole-course management*, and *digital therapeutics* (DTX). It is widely recognized that most chronic diseases need comprehensive management outside the hospital and cannot be cured by short-term drugs in the hospital. At present, the best intervention measure is to carry out various self-management measures of the patients' diet, exercise, and medication outside the hospital. This requires patients to have a certain reserve of medical knowledge, to grasp their own disease changes clearly, and to have high compliance. *Digital therapy* based on emerging technologies, such as mobile medicine, big data, and AI, is the potential best solution. DTX is an intervention program driven by software programs and based on evidence-based medicine that is used to treat, manage, or prevent diseases [24]. Digital therapy transforms the existing medical principles, medical guidelines, or standard treatment schemes into application software–driven interventions by digital means, which can effectively improve the compliance and accessibility of patients' chronic disease management. It is an innovative way to overcome the limitations of traditional drug treatment [25]. Compared with the application of assisted diagnosis, telemedicine, and all new technologies in health, digital therapy can be used alone or together with other therapies to promote disease remission [26].

At this conference, we can see that several companies, such as Weimai Technology Co., Ltd. and Hangzhou Zhuojian Information Technology Co., Ltd., mentioned the concept of digital therapy and proposed corresponding solutions. However, the clinical effect of such schemes has not been verified. This lack of progress may be related to several reasons. On the one hand, the landing effect of products is poor, and the products labeled with *digital therapy* are often simple technical upgrades of traditional business products. On the other hand, most products have not been clinically verified or recognized by peers. Whole-course management and digital therapy are mostly based on concepts. The process of scientific research and standardized verification based on inquiry medicine in the clinic should be accelerated. Due to the relatively mature informatization in hospitals and the large gap and imagination space of out-of-hospital medical services, the corresponding informatization has a large development space and many opportunities in the future, which may form new hotspots.

Conclusion

China's tertiary hospital informatization construction has basically completed the primary stage of popularization, showing two characteristics. First, China's medical information industry is focusing on the construction of smart hospitals. The most important directions include a smart system to support doctors' scientific research, a DRG smart operation system and an OA system to support hospital management, a single-disease CDSS to assist doctors in clinical practice, and the smart internet of things for logistics. Second, under the guidance of the practical needs for improving the quality of grassroots health services and national policies, the construction of a compact county medical community has become a new focus of enterprises. In addition, it can be foreseen that whole-course management and digital therapy will become new hotspots in the future. The process of scientific research and standardized verification should be accelerated.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos. 81771937 and 81871455) and partly supported by the Peking University (PKU)-Baidu Fund (project of intelligent auxiliary diagnosis using medical images; grant no. 2019BD008).

Conflicts of Interest

None declared.

References

1. The Central Peoples Government of the Peoples Republic of China. Opinions of the CPC Central Committee and State Council on Deepening the Reform of the Medical and Health System. 2009 Mar 17. URL: http://www.gov.cn/jrzg/2009-04/06/content_1278721.htm [accessed 2021-12-05]

2. The Central Peoples Government of the Peoples Republic of China. Notice on the Issuance of the Management Method (for Trial Implementation) and Evaluation Criteria (for Trial Implementation) for the Graded Evaluation of the Application Level of Electronic Medical Record System. 2018 Dec 09. URL: http://www.gov.cn/xinwen/2018-12/09/content_5347261.htm [accessed 2021-12-05]
3. The Central Peoples Government of the Peoples Republic of China. Briefing by the General Office of the National Health and Wellness Commission on the Situation Related to the National Monitoring and Analysis of the Performance Appraisal of Tertiary Public Hospitals in 2019. 2021 Mar 31. URL: http://www.gov.cn/xinwen/2021-03/31/content_5597121.htm [accessed 2021-12-05]
4. National Health and Health Commission Medical Administration. Letter on the Overall Situation of the Graded Evaluation of the Application Level of Electronic Medical Record System in Medical Institutions Nationwide in 2019. Beijing: National Health and Health Commission Medical Administration; Jun 29, 2020.
5. American Medical Informatics Association. AMIA 2020 Virtual Annual Symposium: Digital Collection. URL: https://www.amia.org/amia2020/digital-collection [accessed 2021-06-29]
6. Liang J, Wei K, Meng Q, Chen Z, Zhang J, Lei J. Development of medical informatics in China over the past 30 years from a conference perspective and a Sino-American comparison. PeerJ 2017;5:e4082 [FREE Full text] [doi: 10.7717/peerj.4082] [Medline: 29177118]
7. Jia Y, Wang W, Liang J, Liu L, Chen Z, Zhang J, et al. Trends and characteristics of global medical informatics conferences from 2007 to 2017: a bibliometric comparison of conference publications from Chinese, American, European and the Global Conferences. Comput Methods Programs Biomed 2018 Nov;166:19-32 [FREE Full text] [doi: 10.1016/j.cmpb.2018.08.017] [Medline: 30415715]
8. Liang J, Li Y, Zhang Z, Shen D, Xu J, Zheng X, et al. Adoption of electronic health records (EHRs) in China during the past 10 years: consecutive survey data analysis and comparison of Sino-American challenges and experiences. J Med Internet Res 2021 Feb 18;23(2):e24813 [FREE Full text] [doi: 10.2196/24813] [Medline: 33599615]
9. Liang J, Wei K, Meng Q, Chen Z, Zhang J, Lei J. The gap in medical informatics and continuing education between the United States and China: a comparison of conferences in 2016. J Med Internet Res 2017 Jun 21;19(6):e224 [FREE Full text] [doi: 10.2196/jmir.8014] [Medline: 28637638]
10. China Hospital Information Network Conference. URL: https://apps.ifaceyun.com/confer/responweb/index.php?id=663_432816 [accessed 2021-12-05]
11. National Health and Health Commission. About the First Round of the China Hospital Information Network Conference (CHINC). 2021 Mar 25. URL: http://www.niha.org.cn/hwaciis/news/publish/sevenInner?id=1127&title=%E6%90%9C%E7%B4%A2%E7%BB%93%E6%9E%9C [accessed 2021-12-05]
12. TianYanCha. URL: https://www.tianyancha.com/ [accessed 2021-12-05]
13. VbData. URL: https://vbdata.cn/companyList [accessed 2021-12-05]
14. Liang J, Li Y, Zhang Z, Shen D, Xu J, Yu G, et al. Evaluating the applications of health information technologies in China during the past 11 years: consecutive survey data analysis. JMIR Med Inform 2020 Feb 10;8(2):e17006 [FREE Full text] [doi: 10.2196/17006] [Medline: 32039815]
15. Zhao H. Smart Hospitals to Promote the Full Change of Hospital Architecture: A New Generation of Medical Management Information System Reconfiguration Is Underway. Chongqing, China: Chongqing Eggshell Institute; Apr 18, 2021.
16. Jiang K, Sun W, Liu Y. Smart Hospital Industry Trend Report: Digital Progression Is Imminent, the Construction of Smart Hospitals Ushered in a Golden Period. Beijing: iFenxi; Jan 11, 2021.
17. Li L, Chen Q, Powers D. Chinese healthcare reform. Modern China 2012 Sep 17;38(6):630-645. [doi: 10.1177/0097700412457913]
18. Uscher-Pines L, Pines J, Kellermann A, Gillen E, Mehrotra A. Emergency department visits for nonurgent conditions: systematic literature review. Am J Manag Care 2013 Jan;19(1):47-59 [FREE Full text] [Medline: 23379744]
19. Hsia RY, Friedman AB, Niedzwiecki M. Urgent care needs among nonurgent visits to the emergency department. JAMA Intern Med 2016 Jun 01;176(6):852-854. [doi: 10.1001/jamainternmed.2016.0878] [Medline: 27089549]
20. Huang P, Jiang SQ, Hao AM, Zhao X, Zhai S, Han ZJ, et al. Analysis on the two-way referral by leading hospitals of three health alliances. Chin J Hosp Adm 2015 Apr 02:31-36. [doi: 10.3760/cma.j.issn.1000-6672.2015.04.004]
21. Niu Y, Zhang L. Analysis in the development and problems of county medical alliance. China Health Economy 2020 Feb 01;39(02):22-25.
22. Liang J, Zheng X, Chen Z, Dai S, Xu J, Ye H, et al. The experience and challenges of healthcare-reform-driven medical consortia and regional health information technologies in China: a longitudinal study. Int J Med Inform 2019 Nov;131:103954. [doi: 10.1016/j.ijmedinf.2019.103954] [Medline: 31513943]
23. National Health Commission of the Peoples Republic of China. Notice on Promoting the Construction of a Tightly-Knit County Health Care Community. URL: http://bgs.satcm.gov.cn/zhengcewenjian/2019-05-30/9902.html [accessed 2021-12-05]
24. Digital Therapeutics Alliance. URL: https://dtxalliance.org/wp-content/uploads/2021/06/DTA_DTx-Overview_US.pdf [accessed 2021-12-05]
25. China Digital Therapy White Paper (2021). Chongqing, China: Chongqing Eggshell Institute; Apr 16, 2021.

26. Makin S. The emerging world of digital therapeutics. Nature 2019 Sep;573(7775):S106-S109. [doi: 10.1038/d41586-019-02873-1] [Medline: 31554990]

Abbreviations

AI: artificial intelligence
AMIA: American Medical Informatics Association
CDSS: clinical decision support system
CHINC: China Hospital Information Network Conference
CHITEC: China Health Information Technology Exchange Conference
CMIAAS: Chinese Medicine Information Association Annual Symposium
CPMI: China Proceedings of Medical Informatics
DIP: big data diagnosis-intervention packet
DRG: diagnosis-related group
DTX: digital therapeutics
EMR: electronic medical record
HIMSS: Healthcare Information and Management Systems Society
HIS: hospital information system
HIT: health information technology
LIS: laboratory information system
OA: office automation
PACS: picture archiving and communication system

Edited by C Lovis; submitted 16.09.21; peer-reviewed by A Hidki; comments to author 08.10.21; revised version received 09.10.21; accepted 14.11.21; published 11.01.22.

Please cite as:
Zhang Z, Zheng X, An K, He Y, Wang T, Zhou R, Zheng Q, Nuo M, Liang J, Lei J
Current Status of the Health Information Technology Industry in China from the China Hospital Information Network Conference: Cross-sectional Study of Participating Companies
JMIR Med Inform 2022;10(1):e33600
URL: https://medinform.jmir.org/2022/1/e33600
doi:10.2196/33600
PMID:35014959

Original Paper

Use of Clinical Data Interchange Standards Consortium (CDISC) Standards for Real-world Data: Expert Perspectives From a Qualitative Delphi Survey

Rhonda Facile[1], MSc; Erin Elizabeth Muhlbradt[2], PhD; Mengchun Gong[3,4], MD; Qingna Li[5,6,7], MSc; Vaishali Popat[8], MPH, MD; Frank Pétavy[9], MSc; Ronald Cornet[10], PhD; Yaoping Ruan[11], PhD; Daisuke Koide[12], PhD; Toshiki I Saito[13], MD, PhD; Sam Hume[1], DSC; Frank Rockhold[14], PhD; Wenjun Bao[15], PhD; Sue Dubman[1], MA; Barbara Jauregui Wurst[1], MSc, MD

[1]Clinical Data Interchange Standards Consortium, Austin, TX, United States
[2]Medical Science and Computing, a Guidehouse company, Rockville, MD, United States
[3]Digital Health China Technologies, Bejing, China
[4]Institute of Health Management, Southern Medical University, Guangzhou, China
[5]Institute of Clinical Pharmacology, Xiyuan Hospital of China Academy of Chinese Medical Sciences, Beijing, China
[6]Key Laboratory for Clinical Research and Evaluation of Traditional Chinese Medicine of National Medical Products Administration, Beijing, China
[7]National Clinical Research Center for Chinese Medicine Cardiology, Beijing, China
[8]Food and Drug Administration, Center for Drug Evaluation Research, Silver Spring, MD, United States
[9]European Medicines Agency, Amsterdam, Netherlands
[10]Department of Medical Informatics, Amsterdam Public Health Research Institute, Amsterdam University Medical Centers - University of Amsterdam, Amsterdam, Netherlands
[11]LinkDoc Technologies, Beijing, China
[12]Department of Biostatistics & Bioinformatics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan
[13]National Hospital Organization Nagoya Medical Center, Nagoya, Japan
[14]Duke Clinical Research Institute, Duke University Medical Center, Durham, NC, United States
[15]JMP Life Sciences, SAS Institute Inc, Cary, NC, United States

Corresponding Author:
Rhonda Facile, MSc
Clinical Data Interchange Standards Consortium
401 W 15th Street
Austin, TX, 78738
United States
Phone: 1 512 363 5825
Email: rfacile@cdisc.org

Abstract

Background: Real-world data (RWD) and real-world evidence (RWE) are playing increasingly important roles in clinical research and health care decision-making. To leverage RWD and generate reliable RWE, data should be well defined and structured in a way that is semantically interoperable and consistent across stakeholders. The adoption of data standards is one of the cornerstones supporting high-quality evidence for the development of clinical medicine and therapeutics. Clinical Data Interchange Standards Consortium (CDISC) data standards are mature, globally recognized, and heavily used by the pharmaceutical industry for regulatory submissions. The CDISC RWD Connect Initiative aims to better understand the barriers to implementing CDISC standards for RWD and to identify the tools and guidance needed to more easily implement them.

Objective: The aim of this study is to understand the barriers to implementing CDISC standards for RWD and to identify the tools and guidance that may be needed to implement CDISC standards more easily for this purpose.

Methods: We conducted a qualitative Delphi survey involving an expert advisory board with multiple key stakeholders, with 3 rounds of input and review.

Results: Overall, 66 experts participated in round 1, 56 in round 2, and 49 in round 3 of the Delphi survey. Their inputs were collected and analyzed, culminating in group statements. It was widely agreed that the standardization of RWD is highly necessary, and the primary focus should be on its ability to improve data sharing and the quality of RWE. The priorities for RWD standardization included electronic health records, such as data shared using Health Level 7 Fast Health care Interoperability Resources (FHIR), and the data stemming from observational studies. With different standardization efforts already underway in these areas, a gap analysis should be performed to identify the areas where synergies and efficiencies are possible and then collaborate with stakeholders to create or extend existing mappings between CDISC and other standards, controlled terminologies, and models to represent data originating across different sources.

Conclusions: There are many ongoing data standardization efforts around human health data–related activities, each with different definitions, levels of granularity, and purpose. Among these, CDISC has been successful in standardizing clinical trial-based data for regulation worldwide. However, the complexity of the CDISC standards and the fact that they were developed for different purposes, combined with the lack of awareness and incentives to use a new standard and insufficient training and implementation support, are significant barriers to setting up the use of CDISC standards for RWD. The collection and dissemination of use cases, development of tools and support systems for the RWD community, and collaboration with other standards development organizations are potential steps forward. Using CDISC will help link clinical trial data and RWD and promote innovation in health data science.

(JMIR Med Inform 2022;10(1):e30363) doi:10.2196/30363

KEYWORDS

real-world data; real-world evidence; clinical trials; Delphi survey; clinical data standards; regulatory submission; academic research; public health data; registry data; electronic health records; observational data; data integration; FAIR principles

Introduction

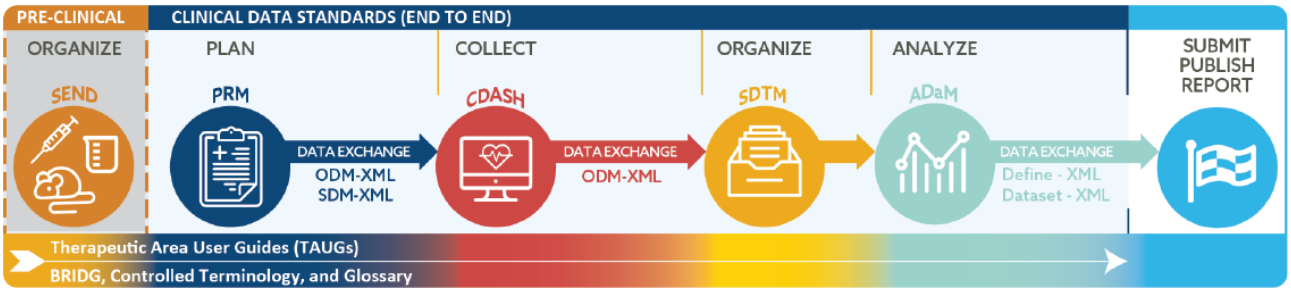
Background

Real-world data (RWD) and real-world evidence (RWE) have an increasingly important role in clinical research and health care decision-making in many countries [1-6]. To leverage RWD and generate reliable RWE, a framework must be in place to ensure that the data are well-defined and structured in a way that is semantically consistent across stakeholders to facilitate learning. The Clinical Data Interchange Standards Consortium (CDISC) RWD Connect Initiative was designed to better understand the barriers to implementing CDISC standards for RWD and to obtain a picture of what tools and guidance may be needed to implement CDISC standards more easily for this purpose.

In the world of traditional clinical trials, which are undertaken with the intent of submitting a new medical product or intervention to regulatory authorities such as the US Food and Drug Administration (FDA) or the Japanese Pharmaceutical and Medical Devices Agency for marketing authorization approval, a set of global data standards has been adopted and is being required by an increasing number of national and regional regulatory agencies. These standards were developed through CDISC, a global nonprofit organization that started >20 years ago to generate open-access platform-agnostic data standards for clinical research and its link to health care.

The CDISC standards span the clinical research process and include standards for the exchange of nonclinical data (SEND), data collection case report forms (CRFs; clinical data acquisition standards harmonization [CDASH]), aggregation and tabulation (study data tabulation model [SDTM]), Biomedical Research Integrated Domain Group (BRIDG) logical model, and operational data model (ODM) for transport (Figure 1). In collaboration with the National Cancer Institute's Enterprise Vocabulary Services (NCI-EVS) program, CDISC has developed a rich controlled terminology that is linked to other common research semantics through the NCI-EVS tools. These standards, presented in data models, implementation guides, and user guides, are globally recognized and heavily used by the biopharmaceutical industry and some academic institutions.

Figure 1. Clinical Data Interchange Standards Consortium standards in the clinical research process. ADaM: Analysis Data Model; BRIDG: Biomedical Research Integrated Domain Group; CDASH: clinical data acquisition standards harmonization; ODM: operational data model; PRM: Protocol Representation Model; SDM: Study Design Model; SDTM: study data tabulation model.



Although there are other standards developed and designed for different purposes (eg, health care data and observational studies), we believe that the benefits of using CDISC standards for purposes outside regulatory submission are many and include

improvements in data sharing, cross-study analysis, and meta-analysis of data for all clinical researchers, as well as streamlining the regulatory submission, review, and approval. Please see the Multimedia Appendix 1 of the full RWD Connect report for 4 supportive use cases (Infectious Diseases Data Observatory, Finger Lakes, Pan American Health Organization Hearts, and the Clinical Innovation Network) [7].

Currently, CDISC standards are required for electronic submissions of study data to the US FDA [8] and the Japanese Pharmaceutical and Medical Devices Agency [9] and are recommended by Chinese [10] and European regulators in rare instances where raw data are requested [4]. Government initiatives or centers that fund research also recommend and use CDISC standards, which include the Innovative Medicines Initiative [11], the US National Cancer Institute, and the National Institute of Allergy and Infectious Diseases. In addition, the Japan Agency for Medical Research and Development (AMED) has stated the following:

> *In the future, clinical trials including investigator-initiated studies will need to comply with the CDISC standards from the planning and implementation stages. Sooner or later, it is expected that we will require the use of CDISC standards for AMED's contract research* [11]

Although there are multiple definitions of RWD currently in use, the CDISC glossary has adopted the following:

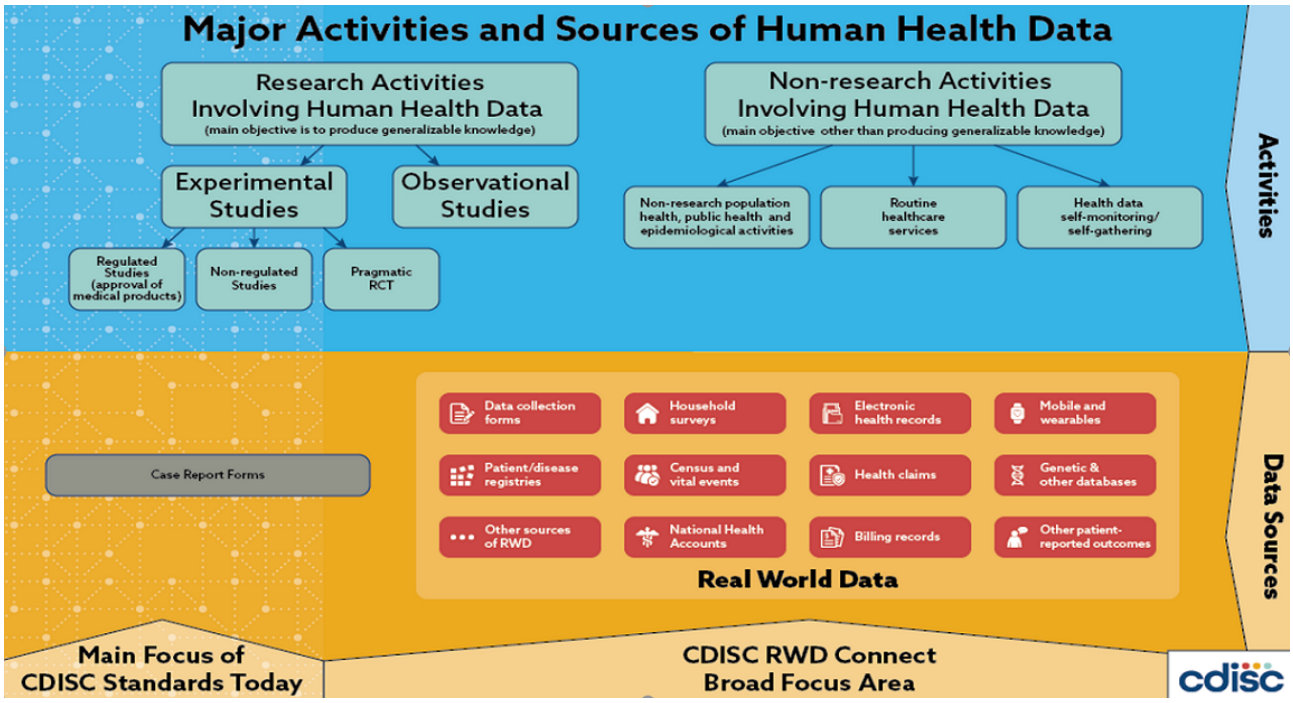
> *Data relating to patient health status and the delivery of health care routinely collected from sources other than traditional clinical trials. Examples of sources include data derived from Electronic Health Records (EHRs); medical claims and billing data; data from product and disease registries; biobanks; patient-generated data, including from in-home-use settings; and data gathered from other sources that can inform on health status, such as mobile devices* [15]

This definition of RWD is similar to the European Medicines Agency (EMA) definition, "routinely collected data relating to a patient's health status or the delivery of health care from a variety of sources other than traditional clinical trials [12]."

Figure 2 describes the data sources for RWD as they relate to research and nonresearch activities involving human health data [7]. This diagram was developed in collaboration with the Expert Advisory Board (EAB) members, with a majority consensus, and it is an oversimplification of reality. It would be impractical to attempt to cover all possible sources and types of RWD. Attempts were made to accommodate all suggestions, some of which contradicted each other. The diagram was meant to generate consensus on the main types of data that are considered RWD and their possible data sources. The FDA defines RWE as "the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD [1]". Therefore, if we have a consensus on the definition of RWD, then we believe that the FDA definition of RWE can be applied. Furthermore, we acknowledge that public health activities can involve research activities, which would then be included in the *research activities* on the left of the diagram. Research activities comprise activities using any kind of data, including public health sources and patient registries. The diagram shows that there are some research activities and many nonresearch activities that generate RWD.

There is no single definition of pragmatic randomized controlled trials. Pragmatic design elements exist on a spectrum [13]. Therefore, further discussion on the definition and scope of pragmatic clinical trials is needed to better understand where they fit in the realm of RWD.

Figure 2. Major activities and sources of human health data. CDISC: Clinical Data Interchange Standards Consortium; RCT: randomized controlled trial; RWD: real-world data.



Sherman et al [14] proposed the following working definition for RWE, "Information on health care that is derived from multiple sources outside typical clinical research settings, including electronic health records (EHRs), claims and billing data, product and disease registries, and data gathered through personal devices and health applications." The CDISC Glossary defines RWE as follows, "The clinical evidence derived from analysis of Real-World Data (RWD) regarding the usage and potential benefits or risks of a medical product [15]." The FDA issued a Framework for RWE [3] in December 2018 to announce a program that included demonstration projects, stakeholder engagement, and internal processes to evaluate RWE and promote shared learning and constituency, as well as guidance to assist in using RWD. This framework also states the following, "RWD sources can also be used for data collection and, in certain cases, to develop analysis infrastructure to support many types of study designs to develop RWE, including, but not limited to, randomized trials (eg, large simple trials, pragmatic clinical trials) and observational studies (prospective or retrospective)."

Similarly, the EMA is also exploring ways to leverage RWD in the generation of RWE. In a recent EMA paper, the authors imagined a future that leverages both regulated clinical trials and RWE to assess safety and effectiveness [16].

Insufficient data standardization in academic and public health settings hinders the use of RWD as part of a regulatory submission package. The use of RWD is increasingly being encouraged by regulatory authorities, given the potential of RWD to provide relevant evidence for new drug or product applications. As noted by Califf [17], RWD could complement and enhance the results of clinical trials. The FDA has expressed the need for new research paradigms to break down the barriers between RWD and clinical research so that evidence can be shared rapidly to improve both domains with increased validity and interoperability [18].

Despite their increasing acceptance as part of regulatory submissions, it is commonly acknowledged that RWD are not collected with research as their primary objective. Therefore, there are significant challenges in using and representing these data for research purposes, which can make the analysis of RWD difficult and resource intensive.

There are a number of disparate standards and systems currently in use to support the collection and analysis of RWD. The diverse panoply of common data models (CDMs; eg, Observational Health Data Science–Observational Medical Outcomes Partnership [OMOP], BRIDG, FDA Sentinel, Patient-Centered Clinical Research Network [PCORNet], and Informatics for Integrating Biology and the Bedside (i2b2)), data exchange standards (eg, Health Level 7 [HL7] Fast Health care Interoperability Resources [FHIR], Define-XML and extension CDISC ODM, and SAS V5 XPORT), and terminologies (eg, Systematized Nomenclature of Medicine–Clinical Terms [SNOMED-CT], Logical Observation Identifiers Names and Codes [LOINC], and Current Procedural Terminology coding) in health care settings across electronic health records (EHRs), insurance claims systems, and medical billing systems are all in varying degrees of development and may not be interoperable as they were not designed to meet the requirements of global regulatory submission [19]. A list of collaborations with other standards and initiatives is provided in Multimedia Appendix 2. Meanwhile, in most other academic and public health settings, data are usually collected in a nonstandard way using different formats and different terminologies [20], which do not allow for the data to be consolidated, compared, and shared. In cases where data are standardized, the variety of approaches, including openEHR, the US National PCORNet, Informatics for Integrating Biology and the Bedside (i2b2), OMOP, and HL7 FHIR, can lead to standard-specific silos. This disconnect creates an evidence gap that slows scientific and public health advances [21]. The need to coordinate across standards is clear, and organizations such as the ISO Joint Initiative Council provide forums to coordinate across standards development organizations; however, these need more support, participation, and adoption.

The benefits of the implementation of standards for RWD are potentially many and include better documentation of data collection, enabled analysis processes, and data sharing [22]. In response, multiple initiatives and tools have been developed in the last few years to seize the opportunity and tackle the challenges resulting from the sudden accessibility of massive amounts of information from multiple RWD sources. For example, in rare diseases where there are many small data collection efforts underway but large regulated clinical trials may not be feasible because of insufficient patient numbers and ethical issues, being able to combine or compare data from different sources becomes even more critical [23]. Cancer is another therapeutic area where there are efforts underway to pool and share data. The National Cancer Institute Cancer Research Data Commons (CRDC) is an infrastructure that connects data sets with analytics tools to allow users to share, integrate, analyze, and visualize cancer research data to drive scientific discovery.

Objective

With these potential benefits in mind and considering the increasing need and interest in data standardization beyond regulatory submissions, CDISC created the CDISC RWD Connect Initiative to develop a vision and strategy for the implementation of CDISC standards for RWD [7]. The first step of the CDISC RWD Connect Initiative was to invite international experts to join an EAB and to involve them in the Delphi survey process described in this paper to better understand what it will take to achieve CDISC standards implementation beyond regulatory submissions.

Methods

Overview

The goal of the RWD Connect initiative was to listen to the stakeholder community to better understand the barriers to implementing CDISC standards for RWD and to get a picture of what tools and guidance may be needed to more easily implement CDISC standards. The second phase focused on creating a strategy for fostering the consistent implementation of CDISC standards within the academic community. In addition, the initiative identified concrete examples of the use

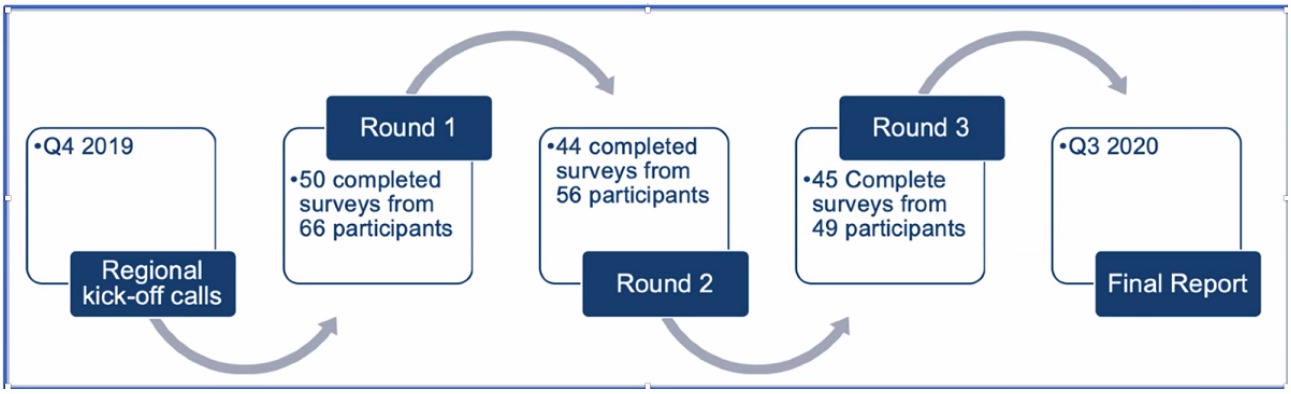
of CDISC standards for RWD and worked collaboratively with the implementers to document the use cases, their scope and characteristics, challenges, and lessons learned. With these goals in mind, we chose to conduct a qualitative Delphi survey to collect an array of different opinions about the use of CDISC standards for RWD and to assess the level of agreement or disagreement on key issues in an asynchronous, global manner. The results from the Delphi and the use cases were the foundation for the proposed vision and strategy described in this manuscript [7].

Qualitative Delphi

In September 2019, the CDISC RWD Connect formed an EAB with key stakeholders. The criteria were knowledge of CDISC standards (any level) and experience working with RWD. In selecting candidates, an effort was made to balance the different regions of the world to the extent possible and to include experts from academia, government, regulators, and health care settings. A list of EAB members is provided in Multimedia Appendix 3.

We identified an initial list of potential members who were either already CDISC partners or collaborators or had been referred by a partner or collaborator. We sent out email invitations to these 70 individuals inviting them to join the initiative, with a required commitment to participate in 3 qualitative Delphi rounds and a final web-based to discuss the results and agree on a way forward. Of the 77 experts invited to participate, 66 (86%) participated in round 1, 56 (73%) participated in round 2, and 49 (70%) participated in round 3 (Figure 3). All EAB members were invited to join the writing committee, and those who accepted are the coauthors of this paper.

Figure 3. Rounds and participants of the modified qualitative Delphi process.



From October 2019 to May 2020, we conducted a modified version of a 3-round qualitative Delphi survey based on published methodology [24]. The goal of the CDISC RWD Connect modified qualitative Delphi survey process was to answer the following questions: what are the priorities, needs, and challenges around the use of CDISC data standards outside regulated clinical trials; how can CDISC minimize the barriers to implementing CDISC standards for RWD; and what are the requirements for potential tools and educational materials for implementation support? The Delphi questionnaire is presented in Multimedia Appendix 4.

In November 2019, the first round of the qualitative Delphi survey was sent to the EAB. The survey comprised 2 sections: section 1 with questions for background information and section 2 with questions for the generation of group statements, as described in the CDISC RWD Connect: Report of Qualitative Delphi Survey [7]. During this first round, we received 50 answered surveys, which included perspectives and insights from at least 66 participants globally (at least 8 answered surveys had consolidated answers from multiple people within a team). From the responses obtained from the first round of the qualitative Delphi, we developed a summary of *group statements* containing the prevailing views of the EAB.

In February 2020, a second round of the qualitative Delphi survey was sent to the EAB. In it, participants were provided with group statements and were given a chance to state whether they agreed with each group statement and how they would modify it. We did not add any new questions. During the second round, we received 44 completed surveys from 56 participants.

In April 2020, the third and final round of the qualitative Delphi survey was sent to the EAB, and participants had a chance to review the final version of the group statements and share whether they strongly agreed, moderately agreed, or disagreed with each of the statements and their reasons for the same. During the final round, we received 45 completed surveys from 49 participants.

Use Cases

Examples are an effective way of showing how CDISC standards can be deployed in use cases outside regulated clinical trials. There are creative and innovative studies already being performed globally in various organizations that make use of CDISC standards. A key part of this study was to collect a number of these use cases from CDISC's existing network of partners and collaborators, as well as to ask for use case recommendations from the EAB.

To collect information on the selected use cases, we performed phone interviews and reviewed databases, presentations, and other documentation relevant to the experience of using CDISC standards for RWD [7].

Results

Overview

In total, 66 experts were included, and 139 feedback instances were collected and analyzed. There was broad agreement that the standardization of RWD is necessary, and the primary focus should be on its ability to improve data sharing and the quality of evidence. The RWD diagram shown in Figure 2 was extensively discussed by the EAB through the Delphi process. Approximately 49% (32/66) of the participants strongly agreed with the final version of the diagram, 42% (28/66) moderately agreed, and 9% (6/66) disagreed.

The priorities of data resources for the CDISC RWD Connect Initiative, as agreed strongly among the experts, included EHRs with a particular interest in data shared using the HL7 FHIR standard, data stemming from observational studies, and wearable devices and patient-reported data. The experts recommended that a gap analysis be performed, as there are different standardization stakeholders in these areas. An official mapping between CDISC and other standard terminologies and a common model to represent the data across different sources was considered necessary. Efforts have been undertaken to fill this void, such as the BRIDG model work group and the FDA CDM Harmonization project [25,26]. This work could be extended to use CDISC as a common model based on existing standards. The duplication of effort should be avoided where possible.

Participants' Background Information

During the first round of the qualitative Delphi survey, which was the most comprehensive and had the greatest impact on the results of this process, we received 50 answered surveys, which included the perspectives and insights from at least 66 participants globally (at least 8 answered surveys had consolidated answers from multiple people within a team). The respondents represented the following continents: Americas, 49% (32/66); Asia, 29% (19/66); Europe, 20% (13/66); and Africa, 2% (1/66). Regarding the represented institutions, 34% (22/66) of the participants represented universities, 24% (16/66) government organizations, 15% (10/66) research centers, 13% (9/66) nonprofit organizations, 6% (4/66) international organizations, and 8% (5/66) others, including hospitals, software companies, and other enterprises. Approximately 95% (63/66) of the participants had experience with RWD, with varying degrees of expertise (Table 1). The *Acknowledgments* section contains a list of institutions represented in the EAB.

Table 1. Expert advisory board participants' experience with real-world data (RWD; not mutually exclusive; N=66).

Participant experience	Participant, n (%)
I have conducted experimental research or academic studies using RWD that were not intended for regulatory submission.	27 (21)
I have conducted observational research studies (cohort study and case control etc).	24 (19)
I have worked with routine health care data.	24 (19)
I have worked with public health data (surveillance and public health programs).	20 (16)
I have worked with multiple RWD sources to conduct research around health care delivery.	17 (13)
I have not worked with RWD.	6 (5)
I attempted to use RWD data but gave up because of challenges.	1 (1)
Other	9 (7)

Benefits and Opportunities From Standardization of RWD

We also asked what participants saw as the primary benefits and opportunities from the standardization of RWD, and specifically, how they would make this case to their colleagues. Most participants (53/66, 80%) strongly agreed that the primary benefits and opportunities from RWD standardization focused on the ability to share data and improve the data quality. Specifically, they stressed that a CDM with no additional mapping was important. As one respondent stated, "achieving accurate results requires a common language, harmonization, and codified and structured data." Respondents acknowledged that implementing standards requires significant investment. However, the use of data standards and vocabularies could enable standard data collection, machine readability, automated data extraction from EHRs, data pooling, an increase in statistical power and scalability (especially for neglected or rare diseases), reproducibility, and allow long-term follow-up of a clinical trial. All of these benefits could be achieved while saving time and effort, enhancing productivity, and speeding the publication of results, which essentially enables findability, accessibility, interoperability, and reusability (FAIR) data principles [27,28].

Participants also noted that with the increased standardization of RWD, there might be an opportunity to better understand RWD and to improve or optimize the study design, which could facilitate more research studies being able to use RWE to support regulatory decision-making. Participants also noted that standards would be key to using data acquired via devices, especially in *Bring Your Own Device* research, and for leveraging other sources of data (eg, claims data). Standards can also increase consistency in clinical trial initiation and execution in both academic and industry settings, which could speed the development of new therapies and treatments. Others noted that standards could reduce the cost of archiving and long-term storage of data, allow for ethics and privacy protection

to be more strictly addressed, and contribute to the learning health care system [29,30].

Priority Components for CDISC RWD Connect

We asked participants to share which types of RWD CDISC should focus on first and why. Below is a summary of the participants' answers after 2 rounds of revisions based on the feedback received. Of note, 62% (41/66) of the participants strongly agreed with the following summary of the priorities and the rationale, and 38% (25/66) moderately agreed with them.

The responses to why CDISC should prioritize EHRs were as follows: EHRs are one of the most available and largest data sources; EHRs are already in electronic format; EHRs contain important and essential information directly relevant to the patients' health status; it would allow us to identify how EHRs could be improved to support better RWE; and EHRs will be the hardest to implement but the most important source of data for the generation of RWE.

The responses to how CDISC should prioritize the harmonization of their standards with HL7 FHIR were as follows: by harmonizing CDASH data elements with FHIR; by working with HL7 working groups to connect clinical research with health care, to update FHIR resources, or develop new FHIR resources needed for research; and by creating a canonical representation of FHIR in CDISC ODM as the electronic data capture vendors will likely be using ODM to ingest and share data from EHRs.

The responses for why CDISC should prioritize observational study data were as follows: observational study data are far less developed in terms of standard use compared with EHRs, and observational studies and pragmatic clinical trials collect similar data to randomized clinical trial data that can be leveraged to inform clinical or policy decision making [31]; CDISC should collaborate with Observational Health Data Science and Informatics (OHDSI)–OMOP CDM on observational research data; as standardized data can be shared and reused more broadly, observational studies using standards will have a greater impact; and the OMOP CDM is a standard-on-the-rise (for observational studies) that should be considered.

Secondary areas of focus should include data commons, registries, mobile health (including automatically generated data), billing records, and medical claims data.

The EAB also mentioned that before broadening the scope of CDISC, a gap analysis and insight into other standardization stakeholders should be conducted. There are already many standards for some of the areas above and often institutional standards as well. At a minimum, to help aggregate and analyze data from these different systems, a published mapping between CDISC controlled terminology and other standard terminologies used for the same data element might be useful; however, the potential lack of equivalence could be problematic. Given that many standard terminologies used in health care do not contain explicit definitions for the concepts contained therein, these mappings could potentially improve those terminologies as CDISC defines all of its controlled concepts. CDISC should also focus on the fundamentals of how to model and represent data and how to manage changes. Unless these are done well, building new additional standards on top of poor foundations will not necessarily bring any benefit. It was the opinion of some on the EAB that some CDISC models have underlying principles, mainly in the areas of data types and data modeling, that can make the implementation challenging. The EAB has recommended augmenting and extending CDISC standards with generalized forms and classes of RWD to address these issues.

Standards for Devices and Wearables

There are significant challenges related to implementing data standards for innovative data collection technologies, such as consumer wearables (eg, Fitbit [Google LLC], Apple Watch [Apple Incorporated], and other monitoring devices). The data itself suffers from credibility, accuracy, and reliability issues associated with proprietary, nonclinically tested algorithms that differ across vendors. This naturally leads to interoperability issues when comparing the same data across different devices; that is, given two different proprietary algorithms, one cannot say that a heart rate measurement is the same across two different devices. Concern was also expressed around privacy, data ownership, and inequitable access, which may leave certain populations out of the analysis. Finally, the current direct-to-consumer marketing approach ensures that there is very little incentive for competing companies to standardize and harmonize among each other.

Patient Perspectives in RWD

There was general agreement that the perspective of the patient, with respect to the collection and use of RWD, is vitally important to ensure the ethical use of the data. However, there was no consensus as to whether the patients' perspective regarding the use of data standards was relevant. At the very least, it was thought that data standards should enable data sharing with the patients themselves and help clinicians make decisions about patient care. Collaboration with professional organizations such as clinical medical societies and disease foundations, as well as patient advocacy groups, was thought to be of value in this effort. A good place to start with respect to patient-valued data standardization was with the standardization of patient-reported outcome data models and measures. Another potential resource currently under development is the Critical Path Institute's *Best Practice Recommendations for ePRO Dataset Structure and Standardization to Support Drug Development*, which uses CDISC standards.

In addition, patient groups should be educated about the benefits of data standards and how this can lead to better and more efficient data sharing. Increasing patients' awareness of the usefulness of the data for themselves and for knowledge generation would ensure strong, patient-lead advocacy groups that promote data standards.

Making the Case for Using CDISC Standards for RWD

Participants were asked what they saw as the main challenges in academic clinical research that could be overcome with the increased standardization of RWD. Their responses focused on issues related to the different sources of data, inconsistency in data collection, inconsistencies in the data, text fields, poor

integration and interoperability, too many standards used or none at all, no standards analysis or meta-analysis tools that results in the development of in-house standards, mapping and the accompanying loss of data or errors, and finally, lack of awareness regarding standards and harmonization of clinical trial initiation and conduct across academic clinical research sites, all of which contribute to the creation of data silos.

Tools or Support Needed

We asked participants what tools or support would be helpful in the implementation of CDISC standards to support RWD in academic settings. The responses focused on providing data collection templates, CDMs, standard user guides, and dictionaries. It was reported that data collection templates containing preannotated fields that link data collection activities to CDISC standards would be useful. In addition, CDISC standards would need to be expanded to include those elements commonly collected and analyzed in observational studies. Finally, educational and training opportunities for CDISC standards will be required to support those working in academic research.

Robust software tooling would also be needed to enable efficient data collection, mapping, quality control or validation, integration, transformation, and analysis. Ideally, software tooling should be open-source, easy to use, flexible, and web-based, containing CRFs and ODM files with built-in CDASH and SDTM coding. Given the heterogeneity of systems used across health care and academic institutions, novel software tooling should be able to interact with the existing standards, such as HL7 FHIR. Mapping across data elements and dictionaries to marry in-house standards with CDISC-standard variables and terminology would also be a useful feature in any software tool. Terminology and metadata validation tools based on open-source Export, Transform, and Load (ETL) tooling may help with quality control issues. These tools would also need to be usable and supported by regulatory agencies.

Building Knowledge and Expertise on CDISC Implementation

We also asked about the most effective ways to build knowledge and expertise on the implementation of CDISC standards in academic institutions. The responses included providing funding for capacity building (eg, award grants to academic institutions and fund institutional roles to support implementation). One respondent noted that CDISC has a role in communicating with research investors or funders to streamline requirements and competing standards across funder organizations. Other recommendations included collaboration and compromise with and among institutions, creation of a certification program, development of simple, free web-based tools (eg, templates for CRFs, data dictionaries, and data sets based on real-world scenarios), documenting and highlighting the use cases and demonstrations, and providing web-based and on-site training.

Discussion

Principal Findings

Existing standards support many facets of human health activity–related data and clinical research. However, there is a lack of standardization for the process to derive RWE from RWD, which results in limited use of RWD in clinical medicine and therapeutics development. CDISC standards have been successfully used in trial-based data management for regulated research worldwide. CDISC aims to extend its standards to support RWD to bridge the existing gaps. However, the complexity of CDISC standards, lack of awareness and incentive to use a new standard, and insufficient training and implementation support were reported to be barriers to setting up standards for RWD following the CDISC methodology, although CDISC has been successful in the trial-based data area. As commented, potential solutions include building use cases for using CDISC for RWE studies, developing tools and support systems, and collaborating with other standards and initiatives (Multimedia Appendix 2).

Barriers to the Use of CDISC Standards for RWD

EAB participants identified the most significant barriers to using CDISC standards in academic settings for RWD. First, it was reported that CDISC standards were considered to be more complex than those used currently for RWD and that their implementation in an academic setting might be burdensome because of unstructured data. There are likely insufficient financial and trained human resources within academic institutions to put toward an implementation. Granting agencies should consider including resource allocation for the use of data standards within their awards. Free, open-source, and easy-to-use tools that incorporate CDISC standards, as well as free or reduced-price training, could also be used to support the implementation of data standards within academic institutions.

Second, there are real gaps in CDISC standards related to RWD that prevent their use in fully supporting RWD at this time. It was the opinion of some on the EAB that some CDISC models have underlying principles, mainly in the areas of data types and data modeling, that can make their implementation challenging. For example, data elements related to longitudinal, prospective, and observational study designs are not sufficiently modeled in CDISC standards currently. The EAB recommended augmenting and extending CDISC standards with generalized forms and classes of RWD to address these issues. A gap analysis between CDISC and OMOP data elements could be the first step in reducing the disparity.

Third, there may be insufficient knowledge of the value of data standards, and more specifically, CDISC standards, coupled with a lack of real and perceived incentives for using standards within institutions such that implementation of CDISC standards may be considered a low priority. In addition, the value of the use of CDISC standards, which has been established in certain sectors (eg, pharmaceutical industry), might not be as well known outside of the regulated research context. An increase in public presentation and publication of case studies showing the enormous value of CDISC standards would go a long way toward educating groups outside of the pharmaceutical industry.

Finally, RWD is currently supported by a number of disparate CDMs, standards, and terminology in use by EHRs, insurance claims systems, and medical billing systems in varying degrees of development; however, they are not connected to one another. These data models, standards, and terminologies are not usually

the same as CDISC, which would require harmonization or mapping to remedy. Currently, there is little incentive for EHR vendors and health care providers to adopt data standards. Furthermore, academic institutions may lack CDISC-trained human resources, which would require financial and temporal resources to remedy. In addition, academic institutions may use multiple disparate systems within and across organizations that would disallow standardization even within a single institution. There is also insufficient knowledge on the importance of data standards and, more specifically, a perceived lack of benefit to using CDISC standards beyond reporting to regulatory agencies. For example, journal publication of results does not require the use of data standards of any kind.

The Future of RWD and CDISC Standards

CDISC initiated the CDISC RWD Connect Initiative with the aim of developing a vision and strategy for the use of CDISC standards for RWD. The following is a list of the key requirements and steps to achieve this goal:

- Simple and flexible tools (eg, templates, plug-and-play tooling, master user guide for mapping and terminology, and open-source file formats)
- Free or affordable training and education (eg, quick start for academics, one-to-one training to create new resources or apps, or registries using CDISC standards)
- Support for standardization of EHR data (eg, decrease the use of open text fields in EHRs to facilitate artificial intelligence data extraction from physician's notes, use new terminologies, and collaborate with health care standards experts and vendors to align and design systems that bridge the health care to research gap) while being mindful of the fact that the primary role of EHRs is patient care, and this process should, therefore, minimize the impact on providing that care
- Publication of use cases that demonstrate the value in the use of CDISC standards outside regulated clinical trials
- Standardization across terminologies used by health care and research
- Simplification where possible and minimizing the number of standards
- Regulation and requirements; specifically, where data cannot be standardized at collection, regulatory requirements must be established to confirm the validity of the mapped data
- Ongoing support for implementation (eg, information technology staffing, 24-hour support, data standards experts, and data warehouse expertise in staff to help implementers)
- Champions and key opinion leaders to support or influence the use of standards and cooperation
- Development of a well-defined purpose and scope for the use of CDISC standards for RWD
- Financial support for development, maintenance, and implementation; specifically, resources are needed for implementation support in the form of educational programs and consulting services
- Incentives in the form of grants to consortiums implementing CDISC standards, free education, free CDISC membership, and granters allowing budget lines for the use of standards and other funding mechanisms can also help encourage the use of CDISC standards

Areas of Nonagreement During the Qualitative Delphi Process

For the most part, EAB participants were able to reach a consensus on the main areas of discussion. However, there were some specific issues on which consensus was not reached. Participants had different ideas regarding the types and sources of data that could be considered RWD. Most participants agreed that RWD standardization efforts should focus on EHRs as a priority. However, the few who strongly disagreed explained that the implementation of CDISC standards in EHRs would be difficult and that HL7 FHIR was addressing the EHR space. Registries were another area of nonagreement, with some participants prioritizing registry data standardization and others saying it should not be a priority. Finally, some participants maintained that CDISC standards should be made easier to use before attempting to expand their scope, whereas others proposed improving the standards in parallel with exploring and testing the expansion of use for RWD.

Limitations

This survey was sampled by convenience; therefore, we were not able to generalize the results of the survey to all settings of RWD generation and use. This project was also geographically limited, as most participants originated from North America and Europe and to a lesser extent from Asia and Africa. We note that the risk for bias is present because of the reasons for which people chose to take part in the Delphi survey.

Conclusions

The CDISC RWD Connect project sought to better understand the barriers to implementing CDISC standards for RWD and to articulate steps toward making CDISC standards easier to use in settings outside regulated clinical trials. Recommendations included identifying the tools and guidance needed for consistent implementation and the expansion of CDISC standards to accommodate data stemming from observational studies, which account for a large amount of available clinical data. Other potential standards development focus areas included data commons, registries, mobile health, and billing and medical claims.

Other practical steps included bringing the standards up to date with current data science technology, making implementation guides easier and more intuitive to be implemented by novice users, and creating a number of tools, strategies, and adaptations to facilitate and promote the use of RWD. Examples included augmenting the SDTM with generalized forms and classes of RWD, creating simpler and more flexible templates and tools, providing free or affordable training and education, increasing regulations and requirements for RWD standards, encouraging champions and financial support, and disseminating concrete examples of the implementation of CDISC standards for RWD. Underpinning these steps, CDISC should support a community of practice that highlights successful implementations and shares their experience by publishing use cases and presenting at conferences and connectathons. Finally, global regulatory

support and mandates from funders of academic studies were also cited as key factors in fostering implementation.

There is a unique opportunity for CDISC to broaden the scope of its suite of data standards to accommodate and connect with RWD to better facilitate RWD sharing. We believe that CDISC standards can provide FAIR structure and semantics for common clinical concepts and domains and help bridge the gap between RWD and clinical trial–generated data for the benefit of all stakeholders. CDISC will use the findings and recommendations from the RWD Connect initiative as inputs to their strategic plan and take the next steps toward developing standards, tools, and guidance for the use of RWD in global regulatory submissions.

Acknowledgments

The writing committee would like to acknowledge the Expert Advisory Board (EAB) for their work on the Clinical Data Interchange Standards Consortium Real World Data (CDISC RWD) Connect Delphi survey project, and Rebecca Daniels Kush, PhD, CDISC Founder and President Emeritus, and Meredith Nahm Zozus, PhD, Professor, Division Chief and Director of Clinical Research Information at the University of Texas Health Science Centre at San Antonio, for reviewing this manuscript.

The views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the regulatory agency/agencies or organizations with which the author(s) is/are employed/affiliated.

Conflicts of Interest

FR is the chairman of the Board Frontier Science and Technology Research Foundation and employee of Duke University and Duke Clinical Research Institute.

Multimedia Appendix 1
Clinical Data Interchange Standards Consortium Real-world Data Connect Report of Final Qualitative Delphi Survey Consultation to Expert Advisory Board.
[DOCX File , 2833 KB - medinform_v10i1e30363_app1.docx]

Multimedia Appendix 2
Collaborations with other standards and initiatives.
[DOCX File , 16 KB - medinform_v10i1e30363_app2.docx]

Multimedia Appendix 3
List of expert advisory board members.
[DOCX File , 18 KB - medinform_v10i1e30363_app3.docx]

Multimedia Appendix 4
Delphi questionnaire.
[DOCX File , 18 KB - medinform_v10i1e30363_app4.docx]

References

1. Use of real-world evidence to support regulatory decision-making for medical devices: guidance for industry and food and drug administration staff. US Food and Drug Administration. 2017. URL: https://tinyurl.com/bd3js9pm [accessed 2021-11-09]
2. Real-world evidence to support drug development and review (trial version). Center for Drug Evaluation and National Medicinal Products Agency, Beijing. 2020. URL: https://www.cde.org.cn/zdyz/domesticinfopage?zdyzIdCODE=db4376287cb678882a3f6c8906069582 [accessed 2021-11-09]
3. Framework for FDA's real-world evidence program. US Food & Drug Administration. 2018. URL: https://www.fda.gov/media/120060/download [accessed 2021-11-09]
4. HMA-EMA Joint Big Data Taskforce – summary report. European Medicines Agency. 2017 Mar. URL: https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf [accessed 2021-11-09]
5. GetReal Project No. 115546, WP1: Deliverable D1.3 - Glossary of definitions of common terms (including comments and replies from consultation rounds). Innovative Medicines Initiative. 2017. URL: https://tinyurl.com/bd3js9pm [accessed 2021-10-28]
6. Uyama Y. Utilizing real world data: a PMDA perspective. Poceedings: DIA 2018 Global Annual Meeting. 2018. URL: https://globalforum.diaglobal.org/issue/august-2018/utilizing-real-world-data-a-pmda-perspective/ [accessed 2021-11-09]
7. Facile R, Wurst BJ. CDISC RWD Connect: Report of Qualitative Delphi Survey. CDISC. 2020. URL: https://www.cdisc.org/sites/default/files/2021-05/CDISC_RWD_Connect_Report.pdf [accessed 2021-11-09]

8. Study data technical conformance guide : guidance for industry providing regulatory submissions in electronic format – standardized study data. US Food & Drug Administration. 2020. URL: https://www.fda.gov/media/136460/download [accessed 2021-11-09]
9. Utilization of real world data - PMDA's approaches. Pharmaceuticals and Medical Devices Agency. 2021. URL: https://www.pmda.go.jp/english/about-pmda/0004.pdf [accessed 2021-11-09]
10. Guideline on the submission of clinical trial data. National Medical Products Administration. 2020. URL: https://www.cde.org.cn/main/news/viewInfoCommon/d238c109af91307fb4f16c4da86506b6 [accessed 2021-11-09]
11. Mansell P. IMI, CDISC hook up for standardised data management. PharmaTimes. 2011. URL: http://www.pharmatimes.com/news/imi,_cdisc_hook_up_for_standardised_data_management_979864 [accessed 2020-12-15]
12. Cave A, Kurz X, Arlett P. Real‐world data for regulatory decision making: challenges and possible solutions for Europe. Clin Pharmacol Ther 2019 Apr 10;106(1):36-39 [FREE Full text] [doi: 10.1002/cpt.1426] [Medline: 30970161]
13. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. Br Med J 2015 May 08;350:h2147 [FREE Full text] [doi: 10.1136/bmj.h2147] [Medline: 25956159]
14. Sherman RE, Anderson SA, Dal PG, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016 Dec 08;375(23):2293-2297 [FREE Full text] [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
15. CDISC Glossary, 2020-12-18. NCI EVS Terminology Resources. 2018. URL: https://evs.nci.nih.gov/ftp1/CDISC/Glossary/CDISC%20Glossary.pdf [accessed 2021-11-09]
16. Eichler H, Pignatti F, Schwarzer-Daum B, Hidalgo-Simon A, Eichler I, Arlett P, et al. Randomized controlled trials versus real world evidence: neither magic nor myth. Clin Pharmacol Ther 2021 May 12;109(5):1212-1218 [FREE Full text] [doi: 10.1002/cpt.2083] [Medline: 33063841]
17. Shattner E. Dr. Robert Califf shares ideas about real-world evidence and health data. Forbes. 2017. URL: https://tinyurl.com/22svcndv [accessed 2021-11-09]
18. Statement by FDA Commissioner Scott Gottlieb, M.D., on new strategies to modernize clinical trials to advance precision medicine, patient protections and more efficient product development. US Food & Drug Administration. 2019. URL: https://tinyurl.com/534dun9n [accessed 2021-11-09]
19. Sun X, Tan J, Tang L, Guo JJ, Li X. Real world evidence: experience and lessons from China. Br Med J 2018 Feb 05;360:j5262 [FREE Full text] [doi: 10.1136/bmj.j5262] [Medline: 29437644]
20. Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I, Joint ISPE‐ISPOR Special Task Force on Real World Evidence in Health Care Decision Making. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies v1.0. Value Health 2017 Sep;20(8):1009-1022 [FREE Full text] [doi: 10.1016/j.jval.2017.08.3018] [Medline: 28964431]
21. Pétavy F, Seigneuret N, Hudson L, Jauregui B, Becnel L, Galvez J, et al. The turning point for clinical research: global data standardization. Appl Clin Trials. 2019. URL: https://www.appliedclinicaltrialsonline.com/view/turning-point-clinical-research-global-data-standardization [accessed 2021-11-12]
22. Hudson LD, Kush RD, Almario EN, Seigneuret N, Jackson T, Jauregui B, et al. Global standards to expedite learning from medical research data. Clin Transl Sci 2018 Jul 26;11(4):342-344 [FREE Full text] [doi: 10.1111/cts.12556] [Medline: 29697200]
23. Mezher M. Patient groups, industry seek changes to rare disease drug guidance. Regulatory Affairs Professionals Society (RAPS). 2019. URL: https://www.raps.org/news-and-articles/news-articles/2019/4/patient-groups-industry-seek-changes-to-rare-dise [accessed 2021-11-09]
24. Sekayi D, Kennedy A. Qualitative Delphi Method: a four round process with a worked example. Qual Rep 2017 Oct 20:2755-2763. [doi: 10.46743/2160-3715/2017.2974]
25. Becnel L, Hastak S, Ver HW. BRIDG: a domain information model for translational and clinical protocol-driven research. J Am Med Inform Assoc 2017 Sep 01;24(5):882-890. [doi: 10.1093/jamia/ocx004]
26. Common Data Model Harmonization (CDMH) and Open Standards for Evidence Generation. US Food & Drug Administration and National Institutes of Health. 2020. URL: https://aspe.hhs.gov/system/files/pdf/259016/CDMH-Final-Report-14August2020.pdf [accessed 2021-11-09]
27. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
28. Kush R, Warzel D, Kush M, Sherman A, Navarro E, Fitzmartin R, et al. FAIR data sharing: the roles of common data elements and harmonization. J Biomed Inform 2020 Jul;107:103421 [FREE Full text] [doi: 10.1016/j.jbi.2020.103421] [Medline: 32407878]
29. Budrionis A, Bellika JG. The learning healthcare system: where are we now? A systematic review. J Biomed Inform 2016 Dec;64:87-92. [doi: 10.1016/j.jbi.2016.09.018] [Medline: 27693565]
30. Institute of Medicine, Roundtable on Evidence-Based Medicine. In: McGinnis J, Aisner D, Olsen L, editors. The Learning Healthcare System: Workshop Summary. Washington (DC): National Academies Press (US); 2007:1-374.

31. Lentz TA, Curtis LH, Rockhold FW, Martin D, Andersson TL, Arias C, et al. Designing, conducting, monitoring, and analyzing data from pragmatic randomized clinical trials: proceedings from a multi-stakeholder think tank meeting. Ther Innov Regul Sci 2020 Nov 08;54(6):1477-1488. [doi: 10.1007/s43441-020-00175-7] [Medline: 32514736]

Abbreviations

AMED: Japan Agency for Medical Research and Development
BRIDG: Biomedical Research Integrated Domain Group
CDASH: clinical data acquisition standards harmonization
CDISC: Clinical Data Interchange Standards Consortium
CDM: common data model
CRDC: Cancer Research Data Commons
CRF: case report form
EAB: expert advisory board
EHR: electronic health record
EMA: European Medicines Agency
ETL: Export, Transform, and Load
FAIR: findability, accessibility, interoperability, and reusability
FDA: Food and Drug Administration
FHIR: Fast Health care Interoperability Resources
HL7: Health Level 7
i2b2: Informatics for Integrating Biology and the Bedside
LOINC: Logical Observation Identifiers Names and Codes
NCI-EVS: National Cancer Institute's Enterprise Vocabulary Services
ODM: operational data model
OHDSI: Observational Health Data Science and Informatics
OMOP: Observational Medical Outcomes Partnership
PCORNet: Patient-Centered Clinical Research Network
RWD: real-world data
RWE: real-world evidence
SDTM: study data tabulation model
SEND: standards for the exchange of nonclinical data
SNOMED-CT: Systematized Nomenclature of Medicine–Clinical Terms

Edited by C Lovis; submitted 01.06.21; peer-reviewed by Y Yu; comments to author 25.07.21; revised version received 17.09.21; accepted 09.10.21; published 27.01.22.

Please cite as:
Facile R, Muhlbradt EE, Gong M, Li Q, Popat V, Pétavy F, Cornet R, Ruan Y, Koide D, Saito TI, Hume S, Rockhold F, Bao W, Dubman S, Jauregui Wurst B
Use of Clinical Data Interchange Standards Consortium (CDISC) Standards for Real-world Data: Expert Perspectives From a Qualitative Delphi Survey
JMIR Med Inform 2022;10(1):e30363
URL: https://medinform.jmir.org/2022/1/e30363
doi:10.2196/30363
PMID:35084343

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org