Clinical Informatics Researcher's Desiderata for the Data Content of the Next Generation Electronic Health Record

Timothy I. Kennell Jr.[1]   James H. Willig[1,2]   James J. Cimino[1,2]

[1] Informatics Institute, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States
[2] Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States

Address for correspondence Timothy I. Kennell Jr., BS, Informatics Institute, University of Alabama at Birmingham, Tinsley Harrison Towers 103, 1900 University Boulevard, Birmingham, AL 35233, United States (e-mail: tikenn@uab.edu).

Abstract

Objective Clinical informatics researchers depend on the availability of high-quality data from the electronic health record (EHR) to design and implement new methods and systems for clinical practice and research. However, these data are frequently unavailable or present in a format that requires substantial revision. This article reports the results of a review of informatics literature published from 2010 to 2016 that addresses these issues by identifying categories of data content that might be included or revised in the EHR.

Materials and Methods We used an iterative review process on 1,215 biomedical informatics research articles. We placed them into generic categories, reviewed and refined the categories, and then assigned additional articles, for a total of three iterations.

Results Our process identified eight categories of data content issues: Adverse Events, Clinician Cognitive Processes, Data Standards Creation and Data Communication, Genomics, Medication List Data Capture, Patient Preferences, Patient-reported Data, and Phenotyping.

Discussion These categories summarize discussions in biomedical informatics literature that concern data content issues restricting clinical informatics research. These barriers to research result from data that are either absent from the EHR or are inadequate (e.g., in narrative text form) for the downstream applications of the data. In light of these categories, we discuss changes to EHR data storage that should be considered in the redesign of EHRs, to promote continued innovation in clinical informatics.

Conclusion Based on published literature of clinical informaticians' reuse of EHR data, we characterize eight types of data content that, if included in the next generation of EHRs, would find immediate application in advanced informatics tools and techniques.

Keywords
► electronic health records
► information storage and retrieval
► health system
► clinical informatics research
► data quality

Background and Significance

Clinical informatics researchers often depend on the reusability of electronic health record (EHR) data to design many of the new methods and systems that improve clinical practice and research. For example, innovations such as those that streamline research subject selection from patient populations require access to patient data.[1] Other applications that provide precision medicine at the point of patient care must compute solutions using these data.[2] This research encompasses a wide variety of data reuse, including retrospective analyses, experimental system development, and data modeling. However, EHRs often fall short of this need for reusable data, either

lacking the information entirely or storing it in a format that requires time-consuming revisions for machine interpretability.[3,4] Clinical informaticians are reporting some of these deficiencies,[5] while making do with extracting and inferring patient information from current EHRs for various purposes.

Given broader discussions surrounding redesigning the EHR are taking place,[6,7] it is timely to examine the systems current state in order to advance the EHR and address its glaring shortcomings for clinical informatics researchers. Studies show EHRs continue to miss important patient data[8] or provide other information in a form that is not machine-processable, complicating data analysis.[3] Both shortcomings are critical to overcome for clinical informatics research and suggest that the data content of these systems requires attention. It is imperative to appropriately store information according to data storage standards and properly capture data types. However, updating the information captured by the EHR and revising storage and retrieval methods will be important to advance the health systems for use as a learning health system.[9]

To catalogue EHR shortcomings that limit data reusability, we have conducted a scoping review of the research informatics literature to identify categories of data content that are inadequate or need revision. These categories can serve as a foundation for establishing some of the data requirements for the next generation of EHRs.

Materials and Methods

We conducted a review of the informatics literature to identify discussions regarding the limited reusability of EHR data and to group these discussions into meaningful categories. To accomplish this task, we first performed a broad, preliminary search to locate journals that most frequently contained articles with this type of content. We selected the journals by using the broad search term "electronic health records" in PubMed without any other limitation and scanning the first 2 years of the results. After limiting the focus of the search based on the preliminary investigation, we then used a standardized search strategy and an iterative expert review process (discussed later) to identify the data content categories through consensus. The iterative process was used to ensure as uniform category creation and article categorization as possible. This study design was selected to encompass a broad overview of the discussions in the literature and provide new perspective on areas in EHR data that need to be addressed for reusability. While the study design primarily follows the strategy set forth in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline,[10] several of the elements in the PRISMA checklist were not relevant to the present study as shown in the ►**Supplemental Material**, available in the online version.

Search Strategy

We searched the PubMed database for informatics literature published between January 2010 and March 2016 to find recent, relevant English-language articles that addressed current limitations of EHR data. Our broad, preliminary search through PubMed yielded several journals with greater

informatics relevant to the topic and we therefore limited our search to the following journals:

- Journal of American Medical Informatics Association.
- New England Journal of Medicine.
- Journal of Pathology Informatics.
- Methods of Information in Medicine.
- International Journal of Medical Informatics.
- Journal of Medical Internet Research.
- Proceedings of the AMIA Annual Fall Symposium.

Additionally, the following inclusion criteria were created to limit the scope of the review to address data content in the EHR and the specific limitations currently present in the system:

1. Data not stored in the EHR that would be useful for patient care or research.
2. Data that would be more efficient to store in the EHR but are typically derived from the information already stored.
3. Data stored in an inefficient manner for downstream use (e.g., research or clinical decision support [CDS]).

Articles were excluded from the review if they met one of the following exclusion criteria:

1. No discussion of EHR data.
2. Only discussion of user interface modifications to the EHR for data already present.
3. Only discussion of derived data that, for efficiency reasons, should remain derived rather than stored (e.g., data subject to constant change).

The search pattern began with the broad search term "(electronic health record) OR (electronic medical record) OR (electronic health records) OR (electronic medical records)" to initially capture a wide variety of articles for review. One reviewer (T.I.K.) surveyed the literature from newest to oldest three times to locate articles matching the inclusion criteria. Per iteration, the reviewer evaluated all the articles listed in ►**Table 1** first by title, then by abstract if the title did not clearly include/exclude the article, and finally by full text if necessary. Following each pass through the literature, all authors served as evaluators of a subset to review the articles selected as meeting the inclusion criteria, the categories of data content created, and the placement of articles into each of the categories. Disagreements were resolved with discussion to reach a consensus.

Analysis

We used consensus as the primary mechanism for moving forward after each standardized iteration through the literature (described below). However, we also calculated an estimated inter-rater reliability (IRR) after each one. As each evaluator was allowed to classify an article into multiple categories, Kraemer's modified kappa coefficient was used.[11] Briefly, for each iteration, we chose a random subset of 30 articles as a representative yet efficient sample from the total articles reviewed per iteration (►**Table 1**). Each evaluator placed the articles into the categories created for the current iteration or chose to create additional ones. The categories were consolidated based on similarity and Kraemer's modified

Table 1 Results from the iterative review process

Evaluation step	Articles reviewed	Articles categorized	IRR	Categories postevaluation
Evaluation 1	655	71	0.4401	8
Evaluation 2	1,062	153	0.5567	8
Evaluation 3	1,215	165	0.6864	8

Notes: Each evaluation step indicates the number of articles that were reviewed, the number that were put into categories of data content for expanding the EHR, the inter-rater reliability (IRR) after all evaluators reviewed a sample of 30 articles, and the final categories after each evaluation.

kappa coefficient was calculated by first computing Fleiss's kappa for multiple categories and raters[12] and adding Kraemer's correction for each rater potentially choosing multiple categories.[11]

Article Classification

Our search strategy's goal was to characterize potential areas for expanding the current data content of the EHR. Our three passes through the literature included several article classification steps, followed by evaluation steps to achieve consensus on the categories of data content and the classification of articles into those categories (see ►**Fig. 1**). We used Zotero 4.0 to manage categorization and access to the articles throughout each pass.

During the first pass through the literature, the reviewer (T.I.K.) created general categories for the articles that met the inclusion criteria. This pass produced the first version of categories to be used for the remaining literature searches. Following review of a set number of articles, we selected a subset of 30 articles for all evaluators to review and independently classify into the list of categories created. During this evaluation, the evaluators reviewed the appropriateness of the articles based on the inclusion criteria, the appropriateness of the categories created and their respective definitions, and the classification of each article. After the evaluation, the evaluators discussed the decisions made regarding each of these topics and made changes based on consensus.

The second pass through the literature included all articles in the chosen time frame. The reviewer classified the remaining articles according to the second version of the categories (see ►**Fig. 1**). We then chose a second sample of 30 articles for all evaluators to review and classify using the revised set of categories. Again, we resolved disagreements regarding categories, category definitions, and article classification through discussion and created a third version of categories.

The reviewer then performed the third pass through all articles and classified them according to the third version of the categories. We performed an evaluation of 30 articles and discussed the previously mentioned topics to resolve disagreements by consensus.
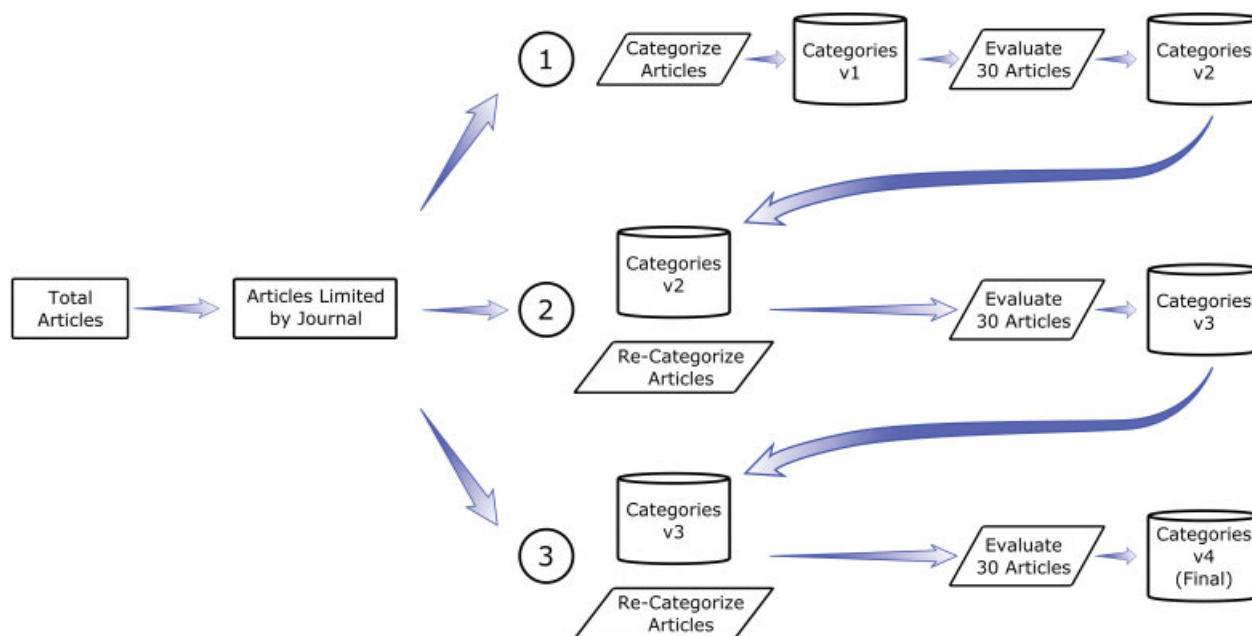

Fig. 1 Literature search and article classification strategy for identifying potential areas of expansion for electronic health record data content. Articles returned by the search term "(electronic health record) OR (electronic medical record) OR (electronic health records) OR (electronic medical records)" were filtered by journal and the remaining articles matching the inclusion criteria were placed into general categories . A sample of 30 articles and their respective categories from this first pass through the literature were then evaluated for both the appropriateness of the article, the appropriateness of the categories, and the relevance of the articles to the categories (). A second version of categories was then created and the process was repeated two more times (and) using the previous set of categories from each pass yielding a final version of categories.

Results

Overview

Our preliminary search for articles published between January 2010 and March 2016 retrieved 26,031 citations. Using the journals identified in the preliminary search, the standardized search returned 1,215 articles that were then reviewed based on the inclusion and exclusion criteria. The method of consensus used to select articles for each iteration then resulted in a different number of articles selected from the 1,215 articles per iteration as discussed below. Additionally, each iteration, except the final, reviewed only a subset of the 1,215 articles to maintain consensus on category creation and article selection (see ►Table 1).

For each iteration of the categorization process (see ►Table 1), we reviewed a set number of articles for categorization. The first evaluation did not complete the entire time span from January 2010 to March 2016 resulting in fewer number of articles. Additionally, several more articles in the selected journals were added to the literature after evaluation 2 as well. The IRR increased for each iteration, but the number of categories remained constant (see ►Table 1). While the number of categories did not change, we made modifications to the names of the categories and their definitions. For example, from the second round of classification, the authors agreed to change the category "Drug Monitoring" to "Medication List Data Capture" and also changed the definition from:

- Articles describing a need for or the creation of an algorithm or data model to monitor patient use of drugs whether through abuse, adherence (medication compliance), or incidental use.
- More robust medication data storage (e.g., medications prescribed by other hospitals, medication and/or illicit drug abuse information), including additional drug metadata (e.g., adherence to medication schedule) that would allow clinicians to easily determine a patient's medication status along with storage of patient medication information in medication list rather than free text.

The categories in ►Table 2 represent the final consensus categories and definitions chosen by the evaluators after iteratively classifying the articles. We believe that these categories represent some of the major concerns raised in the literature regarding shortcomings of the EHR to provide data for reusability.

►Table 3 presents an overview of the classification process and shows an example of how an article was grouped into each category. The information is presented for the final iteration.

Categories for EHR Data Content Expansion

Adverse Events

An adverse event (AE) in medicine is any undesirable event that occurs during or as the result of treatment, including falls, adverse drug events (ADEs), and food allergies. While these events may be recorded in the EHR, this information is typically not stored in a structured location and detection of these events post occurrence typically requires searching both structured and unstructured data. Although some studies employ techniques such as rule-based detection[13] and pattern matching in free text[14] as in phenotyping, many more studies utilize a combination of natural language processing (NLP) and machine learning (ML)[15–18] to search the free text. One study specifically pointed out that while pattern-matching methods could easily extract common side effects of medications, ML (decision trees in this specific study) was more useful for extracting more complex symptomatologies as ADEs.[14] Once extracted, the information has multiple downstream applications including research, reporting, quality improvement, and prevention.

Clinician Cognitive Processes

The clinician's cognitive process during patient care is the reasoning behind decisions made in diagnosing and treating patients. While the category is broad, studies focus on two primary aspects: alert overrides and patient handoffs. A list of reasons for overriding an EHR alert that is customized to be more relevant to a clinician during patient care has been shown to improve the appropriateness of the one chosen.[19] Other studies have repeatedly shown that structured data elements reflecting clinical reasoning are important during the hand-off process, suggesting that clinical reasoning is vital during communication.[20,21] Storing cognitive maps that diagram the thought process during patient care is one effective mechanism of storing this reasoning for use during hand-offs.[22] Additionally, one specific study attempted to manage the large amount of information to be analyzed during transfer by studying the effectiveness of a handoff tool that automatically imported relevant information.[23]

Data Standards Creation and Data Communication

This category focuses on the ability to unify the data content across multiple EHRs. Many studies discuss two similar methodologies of unifying the data: (1) standardizing the storage of the data in a component of the EHR itself[24–26] or (2) extracting the data from the EHR in a standardized format agnostic of the underlying data structure.[27–30] While a handful of these studies, specifically those focused on data communication, attempt to solve the issue of general data unification, other studies attempt to solve only a smaller area, such as oncology,[31] rare diseases,[25] or family health history.[32] Multiple standards have been employed, including HL7, HL7's Fast Healthcare Interoperability Resource (FHIR), the Consolidated Clinical Document Architecture (C-CDA), and the Web Ontology Language (OWL), with varying success. However, the flexibility provided by many of these standards prevents complete interoperability between systems.

Other studies have focused on some of the difficulties in the data modeling process for purposes of standardization. One study evaluated several of the tools used for modeling clinic workflow and suggested that the tools are not mature enough to appropriately handle all modeling requirements.[33] Another study described an application that provided clinicians with feedback regarding data quality and similarity of reporting across multiple hospitals in the Netherlands.[34] While this tool

Table 2 Number of articles per category after the final iterative review through the literature

Category	Description of a need for or the creation of an algorithm to detect or a data model for...	Number of articles (citations)
Adverse events	The potential for or the occurrence of unexpected and/or undesirable medical events such as drug allergies, drug side effects, falls, unexpected diseases, or other treatment-related injury	22[13–18,82–97]
Clinician cognitive processes	The clinician's reasons for decisions made in the EHR regarding patient care, including alert overrides and handoffs	12[19–23,98–104]
Data standards creation and data communication	Storage of data for medical fields or aspects of medical fields in a standard medical format (e.g., HL7, C-CDA, or an author-specific format) or mapping of data models of commonly used resources (e.g., Web sites or apps) to standard medical data formats for the purpose of EHR interoperability among other EHRs and external applications	29[24–34,105–122]
Genomics	Patient genetic information including WGS, whole-exome, SNP, and other genetic data from other tests not listed. The genetic data can be utilized for the purpose of diagnosis, prescription (pharmacogenomic information), or other medically relevant purposes	12[35–41,123–127]
Medication list data capture	More robust medication data storage (e.g., medications prescribed by other hospitals, medication and/or illicit drug abuse information), including additional drug metadata (e.g., adherence to medication schedule) that would allow clinicians to easily determine patients' medication status along with storage of patient medication information in medication list rather than free text	14[42–54,128]
Patient preferences	Storage of patient's desires for treatment, therapy, or lack thereof for health events such as end-of-life care, diseases, or AEs	2[57,129]
Patient-reported data	An outcome of a health event (e.g., disease, risk factor) or therapy (medication schedule, treatment plan) that is reported by and directly relatable to the patient. Quantification is sometimes done through the abstract score of quality of life. Patient-reported data may not be correlated with medically defined outcomes (increased FEV_1 in COPD patients does not always result in improved QOL for a patient)	13[58–64,130–135]
Phenotyping	Identifying a specific, medically relevant, physical characteristic (e.g., disease state, current treatment, or physical trait) by utilizing the presence of clinical data in the medical record (e.g., laboratory test results, clinical notes analyzed through NLP, or physical exam findings)	61[13–18,20,24,25,31,32,35, 41–43,45,46,60,64,66,68–187]

Note: The categories and their definition of data content described in the literature that could be used to expand the EHR.

improved data quality, the provision of relevant diagnostic codes still varied greatly between 30 and 100%.

Genomics

Genomic information can include anything from a basic genetic test designed to identify a single-nucleotide polymorphism (SNP) to whole-genome sequencing (WGS). Overall, several studies have highlighted a need to improve the infrastructure used to store genetic information and the need to implement standard ontologies and semantics regardless of a genetic test's originating laboratory.[35–37] The results of a genetic test can be used for CDS, provided that metadata regarding actionability of the test results are available.[38] Pharmacogenetic data (drug-related genetic data) have many similarities to their parent category, genetic data, including the need for standardization[39] and metadata for CDS.[40] However, the relationship to medications and

prescriptions opens the potential to creating an association between a prescribed medication and a genetic test, thereby annotating reasoning behind the medication choice or dosage.[41]

Medication List Data Capture

Medication lists are a common component of any EHR today and maintain an active record of a patient's medications, both current and previous. Several studies focused on improving the accuracy of this list utilize both NLP and ML to retrieve additional information from the free text. One study showed the effectiveness of these techniques in monitoring opioid use in patients who did not have opioid use frequency recorded in their chart.[42] Another used NLP to detect antidepression medication use in patients.[43] Other studies have focused on more general omissions in a patient's medication list with one predicting missing medication using an ML

Table 3 Example articles from each category and the description of their categorization process

Category	Title of article in category	Description of categorization
Adverse events	A long-term follow-up evaluation of EHR prescribing safety	Article discusses the analysis of prescription error rates when transitioning between EHRs. Information retrieval for this study required data derivation from a chart review
Clinician cognitive processes	A novel use of the discrete templated notes within an EHR software to monitor resident supervision	This article discusses the specific documentation of resident procedures outside formal procedures allowing monitoring of resident training and potentially cognitive reasoning behind procedures
Data standards creation and data communication	A methodology for a minimum dataset for rare diseases to support national centers of excellence for health care and research	Specifically discusses standard data elements that could be used for rare diseases for epidemiology studies
Genomics	An EHR-driven algorithm to identify incident antidepressant medication users	Discusses a pharmacogenomics platform that focuses on harvesting this type of data for CDS and reporting. The article specifically deals with genomic data for CDS
Medication list data capture	Creating a scalable clinical pharmacogenomics service with automated interpretation and medical record result integration—experience from a pediatric tertiary care facility	Discusses the design an algorithm for derivation of antidepressant users from the EHR data. Indicates missing information on patient medication lists
Patient preferences	An information model for automated assessment of concordance between advance care preferences and care delivered near the end of life	Discusses storage of advance care preference (a patient preference) information in the EHR in an easier-to-retrieve format
Patient-reported data	Assessing older adults' perceptions of sensor data and designing visual displays for ambient environments	Studies the perceptions of elderly patients toward the use of in-home sensors for the collection of medical data (patient-reported due to sensor collection directly from the sensors). Addresses the collection of this information
Phenotyping	A collaborative approach to developing an EHR-phenotyping algorithm for drug-induced liver injury	Discusses the creation of a phenotyping algorithm designed to identify patients in the EHR with drug-induced liver injury

Abbreviations: CDS, clinical decision support; EHR, electronic health record.
Note: Each article's content is described in relation to the reason that it fits in the category (i.e., why it was placed in the category listed).

algorithm[44] and others highlighting missing medications at discharge.[45,46]

A related area of study is the retrieval of both the reason for a medication's prescription and the duration of taking a medication. In the 2009 i2b2 challenge, methods for identifying the reasons for and duration of prescriptions employed pattern matching (such as regex) combined with NLP, heuristics,[47,48] and ML.[49] All studies concluded that this type of information was the most difficult to extract.[47–54]

Patient Preferences

While patient preferences for end-of-life care can be found in many EHRs, locating these preferences in the system can be challenging.[55] The difficulty in locating this information adds to the lack of a mechanism for determining if the patient's preferences have been met.[56] One study investigated the use of 15 end-of-life data elements that could easily determine the status of end-of-life care with regard to patient preferences.[56]

Patient preferences also extend to other contexts, including the ability to send a reason for a nurse's call. One study explored the ability to transfer contextual information regarding patients' desires along with the nurse's call that allowed the nurses to more appropriately respond to the patient's request.[57]

Patient-Reported Data

Patient-reported data (PRD) falls into two large subcategories: (1) data that are directly reported by a patient to a clinician as something that is relevant to his or her health and (2) sensor data that are passively collected from the patient through devices such as smartphones or home-based devices. Regarding directly reported data, the literature currently suggests that PRD be collected and stored in a standardized method. Many studies provide possible discrete data elements to be used, including elements of a personal profile, goals for overall health and clinic visits, and quality of life.[58] Other studies focus on incorporating social and behavioral determinants of health as facilitators for patient care.[59,60]

Several studies have actually shown feasibility of collecting PRD from sensor data as well. By tying a continuous glucose monitor to a smartphone and eventually to the EHR, one group

showed that passive glucose levels could be captured and stored in a patient's chart.[61] Other studies have experimented with using smart home sensors for monitoring elderly patients' health.[62–64]

Phenotyping

Phenotyping is the process of identifying cohorts of patients with desired characteristics (typically disease states). This output may then be used for downstream purposes in research. The most common approaches to phenotyping have been rule based[65] and are the simplest methods of this type potentially utilizing only the diagnosis codes and problem lists in the EHR. However, due to reasons such as missing diagnostic codes from a patient's record or diagnoses absent from the problem list, rule-based algorithms include other information, such as medication lists, laboratory values, and chart reviews to increase the accuracy of phenotyping patients.[66,67] Once constructed, the rule-based algorithms are validated and then may be submitted to one of several public repositories for general use such as PheKB eMERGE (Phenotype Knowledge Base: Electronic Medical Records and Genomics), OMOP (Observational Medical Outcomes Partnership), and others.

Rule-based techniques have been frequently used in the past. However, the methods needed to construct, validate, and implement them are time consuming, especially if the rule requires the interpretation of data found in the free text of patient's charts.[68,69] As a result, alternative methodologies have been employed to speed up the phenotyping process, including the use of NLP and ML. NLP, which uses linguistic knowledge to allow computers to gather knowledge from language (speech, text, etc.), has been used with rule-based algorithms to assist with mining free text[70,71] and in conjunction with ML, a probabilistic modeling technique, to identify patients using large feature sets.[72,73]

Because all that is needed is labeled data, both techniques decrease the time needed to validate and implement the algorithm. Unfortunately, the initial construction has a large upfront time cost due to the expense of labeling data to train a ML algorithm. However, there have been recent attempts to decrease the time cost by automating the labeling process.[72,74]

Of final note is the accuracy of phenotyping algorithms. All of the aforementioned techniques utilize inferential techniques to phenotype patient cohorts due to the imperfect capturing of patient phenotypes in the EHR. As a result, those wishing to use the phenotyping results for a downstream application must take into account these inaccuracies when deciding on the phenotyping algorithm to use.[75]

Discussion

Clinical informatics research focuses on creating new methods and systems that depend on diverse, high-quality EHR data, which are not always present. The purpose of our study was to identify categories of data content that would promote the reusability of data in the EHR for clinical informatics research. It was not our intent to create an exhaustive list of all data that should be considered for addition as EHRs undergo natural

evolution, but rather to focus on those data for which reuse application in informatics are ready and waiting.

One proposed goal for advancement of medical systems is to create the learning health system.[76,77] This system would incorporate data from patients, clinicians, laboratories, and many other information sources to translate information to knowledge. Part of the translation process will require the appropriate data to be available in a reusable format. The eight categories we found should be considered a starting point for revising the next generation of EHRs, built with the ability to allow data reuse for clinical informatics research and the advancement of the learning health system.

The categories are not intended to be a classification system for published articles; so, we were not concerned about the occasional differences in how an individual article was classified. Although the IRR in classification increased as we revised our categories with each iteration in the evaluation, it is more important to note that all evaluators ultimately agreed on the definitions of these categories and that they were sufficient to classify all articles that met our inclusion criteria.

The presence of each of these categories in the literature unifies them as pieces of the larger problem of EHR data reuse for clinical informatics research. However, while each of these categories makes up a component of the discussion, they also vary in the scope of data content that they cover. For example, Medication List Data Capture focuses specifically on the medications and associated metadata captured by the EHR, whereas phenotyping covers a broader aspect of data content. This discrepancy is expected due to the varying levels at which EHR data can be reused. Medications cover a focused area of data content, yet represent a major aspect of patient treatment. In contrast, phenotyping, which focuses on identifying disease states, employs a large portion of the EHR's data and is naturally larger than other categories.

Additionally, some categories represent novel information that is not captured by the EHR, such as explicit expression of clinicians' cognitive processes. If this information is present in the EHR, it is typically found only in narrative-free documents such as clinical notes and hand-off documentation. Other categories focus on information captured in a form that needs a revision. For example, phenotyping information is abundant, yet effective use of this information requires complex predictive algorithms that can never be 100% accurate.

It is important to note that the categories proposed have varying degrees of actionability in the clinic, currently. For example, categories such as Adverse Events, Medication List Data Capture, and Patient Preferences can typically be immediately acted on with current medical knowledge and standards of practice. However, other categories, such as PRD and Genomics, may still require more research to make the data use more effectively in the clinic. This last statement, however, highlights the need to make this type of data that are already in the EHR reusable.

While there are many ways of addressing the capture or storage structure of the data content in each category, it is our general observation that most of the projects in each of the categories described in our literature set would benefit from a data storage that used a standardized terminology.

However, not all projects were so bold to request this explicitly. Questions remain on how to acquire these data and how to store them appropriately. The current predominant method for accomplishing this is to require clinicians (typically nurses and physicians) to take on new structured data entry tasks (diagnoses, problem lists, medications, allergies, etc.)—often replicating efforts they had already expended in writing their notes and reports. Enthusiasm for such additional responsibilities is often lacking.

Several solutions for each category might exist (see ►Table 4). Solutions that might alleviate the need for excessive data entry will require advances in clinical informatics research and possibly policy changes. For example, most of the categories would benefit from a unified medical record, whether the data are centralized at a single institution or distributed at the patient level.[78] This system would unify and relate medical information across all patients, regardless of the institution providing care. Such a system would require the use of a standardized, research-controlled terminology mentioned earlier and would also necessitate

Table 4 Potential solutions to each of the categories of data content being discussed in the literature as missing from the EHR or needing revision

Category	Solutions
Adverse events	• Unified medical record • Voice recognition • NLP • Patient portals
Clinician cognitive processes	• Voice recognition • NLP
Data standards creation and data communication	• Unified medical record • HL7 development and standardization • Voice recognition • NLP
Genomics	• Unified medical record • Automated laboratory data transfer
Medication list data capture	• Unified medical record • Voice recognition • NLP
Patient preferences	• Unified medical record • Voice recognition • NLP • Patient portals
Patient-reported data	• Unified medical record • Voice recognition • NLP • Patient portals
Phenotyping	• Unified medical record • Voice recognition • NLP

Abbreviation: NLP, natural language processing.
Note: Some of the proposed solutions are currently implemented in some EHRs but are not developed enough to solve the issue. Others will require significant research and further development of the technologies. Still others might require policy changes.

the standardization of a data model for storage. However, both requirements would enforce interoperability. Additionally, using the categories presented in this study as a starting point might guide the creation of an information model that would allow data reusability for researchers with permissions for access. This system would also prevent duplication of data when patients visit multiple health care institutions. Coupled with continued development of data communication standards, such as HL7, a unified medical record would allow data transfer rather than requiring data entry.

Additionally, other solutions, such as voice recognition coupled with NLP, could lower some of the barriers perceived by those charged with the data entry. While benefitting multiple categories, transfer and storage of clinician cognitive processes would especially benefit from technology. For some categories, such as AEs, PRD, and patient preferences, information could be captured directly from the patient through patient portals, which are currently in use in limited circumstances, or monitoring devices such as wearables. However, for all data entry methods, barriers could be reduced by making clear the immediate and long-term advantages of new data capture. For clinicians, this might come in the form of intelligent decision support that automates workflow processes. For patients, it might be clear indications of the medical benefits and progress tracking over time that new data capture provides.

It is noted that NLP and voice recognition form a solution for most of the categories that have been mentioned. The exception, genomics, can be addressed through an automated laboratory data transfer. As previously discussed, these technologies might alleviate the burden of data entry. Additionally, these technologies might provide an initial method of implementing a standard from free text rather than requiring rigid data entry. The key point is that both technologies would alleviate the strain on data entry, allowing it to be more flexibly submitted while potentially maintaining storage in a standardized way. These technologies would form an additional benefit to those already mentioned for the unified medical record that also would benefit most the categories through standardizations.

Moving forward, it is important to recognize that modifications to the EHR to address the categories presented here will have a cost associated with them. These costs include developers' time, physicians' time, licensing, and others. The cost might depend on the current status of a category's development in the EHR. For example, genetic information might have a higher cost as it is still in its beginning stages of addition to the EHR compared with other categories. Therefore, it will be important moving forward to justify costs carefully and consider that continued development in one area (such as phenotyping) might assist in some way with others (most genetic information has phenotypic implications).

There have been other studies highlighting the shortcomings of the EHR and suggesting changes. For example, Liaw et al focus on the accuracy of the data present in the EHR, including the completeness of the data as one aspect of that metric.[79] Dixon et al suggest a general framework for addressing data quality in the EHR without discussing specific areas of the data content needing attention.[80] Finally, Cusack et al focus on

general recommendations highlighting data collection with respect to patient care as needing the main focus.[81] Each of these studies highlights a different aspect of the shortcomings in the EHR. Our study focuses on specific data content areas that need attention for the clinical informatics researcher in the next-generation EHR. Specifically, we address data reusability in our study. We, therefore, believe that we have highlighted another important area for improvement moving forward.

There are a few limitations of our study; however, we do not believe these have affected our conclusions. Our literature search was limited to articles published in 2010 or later, because we sought categories describing the current state of the EHR and its shortcomings regarding data use for clinical informatics research. Surveying older literature might have identified categories that have already been resolved. Our broad, preliminary search leading to a focus on prominent informatics journals indexed in PubMed may have excluded some articles that may have introduced additional categories. However, we believe it is unlikely that a popular article would completely evade the mainstream literature. Finally, although there may be many other areas of data content that could be added to the EHR, we believe that the categories that we have identified focus on the major areas of discussion in the literature that surround the use of EHR data for clinical informatics research.

Conclusion

Despite 50 years of development, EHRs still remain inadequate for many intended tasks, including clinical informatics research. As the next generation of informaticians takes on the task of developing the next generation of EHRs, we recommend that their plans incorporate new data types and structures guided in part by the eight desirable categories we have distilled from current clinical informatics literature. Although creative approaches will be needed to accomplish this, many promising applications stand ready to exploit these data to improve the care of individual patients and, through a "learning health system,"[9] the health of humankind.

Clinical Relevance Statement

The reusability of electronic health record data provides clinical informatics researchers the ability to create innovative applications for clinical applications. The revisions and additions to the EHR data content that we have discussed will streamline these innovations, providing faster development of these applications for use by clinicians. Additionally, many of the improvements discussed, such as genetic data contend, would affect the ability of the clinician to store, retrieve, and interpret a patient's genetic information in a clinical context.

Multiple Choice Questions

1. Of the following, which solution for data storage would offer the most uniform structure for storage and retrieval of the medical information for both research and clinical practice?

A. Unified medical record
B. Universal medical record number
C. Data communication standards
D. Natural language processing

Correct Answer: The correct answer is A, unified medical record. A "Universal medical record number" would allow information to be linked across EHR systems that adhered to the universal medical record number; however, this would not ensure data storage or retrieval in each of these systems would be similar past this relationship.

Data communication standards may provide universal retrieval of data, but will not ensure universal storage. This answer, therefore, is also incorrect.

Finally, natural language processing is an information retrieval methodology that might be able to form a layer between data entry and retrieval to standardize data flow in either direction (storing free text as standardized structured information or retrieving free text as structured information). However, the best implementations of this method are inferential and will always have additional error beyond error in data entry and retrieval themselves.

2. Which of the following categories of information in the electronic health record has the greatest impact on natural language processing in terms of information retrieval?

A. Patient preferences
B. Phenotyping
C. Data standards creation and data communication
D. Genomics

Correct Answer: The correct answer is B, "Phenotyping." Natural language processing (NLP) currently plays a major role in most techniques used to attribute phenotypes to patients because it allows free text to be searched in addition to structured data.

Patient preferences are typically stored in structured data in the electronic health record to allow easy retrieval. Additionally, there is not much research in this area to use NLP for retrieving this information from free text.

Data standards creation and data communication is, somewhat by definition, not intended to use NLP, and is therefore incorrect. While it might be possible to store and retrieve information in this manner for data communication purposes, the use of standards should remove the need for the use of NLP.

Finally, genomics is currently not often stored in the EHR through structured data or free text. Additionally, methods other than NLP are typically used for retrieval of structured information from the genetic data.

Conflict of Interest

None.

References

1 Murphy S, Wilcox A. Mission and sustainability of informatics for integrating biology and the bedside (i2b2). EGEMS (Wash DC) 2014;2(02):1074

2 Sitapati A, Kim H, Berkovich B, et al. Integrated precision medicine: the role of electronic health records in delivering personalized treatment. Wiley Interdiscip Rev Syst Biol Med 2017;9(03). Doi: 10.1002/wsbm.1378

3 Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 2013;51(08, Suppl 3):S30–S37

4 Hersh WR, Cimino J, Payne PR, et al. Recommendations for the use of operational electronic health record data in comparative effectiveness research. EGEMS (Wash DC) 2013;1(01):1018

5 Zulman DM, Shah NH, Verghese A. Evolutionary pressures on the electronic health record: Caring for complexity. JAMA 2016;316 (09):923–924

6 Ajami S, Arab-Chadegani R. Barriers to implement electronic health records (EHRs). Mater Sociomed 2013;25(03):213–215

7 Meigs SL, Solomon M. Electronic health record use a bitter pill for many physicians. Perspect Health Inf Manag 2016;13(Winter):1d

8 Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB. Missing clinical and behavioral health data in a large electronic health record (EHR) system. J Am Med Inform Assoc 2016;23(06): 1143–1149

9 Smith M, Saunders R, Stuckhardt L, McGinnis JM, eds. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, DC: The National Academies Press; 2013

10 Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(07):e1000097

11 Kraemer HC. Extension of the kappa coefficient. Biometrics 1980;36(02):207–216

12 Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378–382

13 Epstein RH, St Jacques P, Stockin M, Rothman B, Ehrenfeld JM, Denny JC. Automated identification of drug and food allergies entered using non-standard terminology. J Am Med Inform Assoc 2013;20(05):962–968

14 Sohn S, Kocher J-PA, Chute CG, Savova GK. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. J Am Med Inform Assoc 2011;18(Suppl 1): i144–i149

15 Plasek JM, Goss FR, Lai KH, et al. Food entries in a large allergy data repository. J Am Med Inform Assoc 2016;23(no. e1): e79–e87

16 Wang G, Jung K, Winnenburg R, Shah NH. A method for systematic discovery of adverse drug events from clinical notes. J Am Med Inform Assoc 2015;22(06):1196–1204

17 Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. J Am Med Inform Assoc 2014;21(02):308–314

18 Rochefort CM, Verma AD, Eguale T, Lee TC, Buckeridge DL. A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data. J Am Med Inform Assoc 2015;22(01):155–165

19 Dekarske BM, Zimmerman CR, Chang R, Grant PJ, Chaffee BW. Increased appropriateness of customized alert acknowledgement reasons for overridden medication alerts in a computerized provider order entry system. Int J Med Inform 2015;84(12):1085–1093

20 Flink M, Bergenbrant Glas S, Airosa F, et al. Patient-centered handovers between hospital and primary health care: an assessment of medical records. Int J Med Inform 2015;84 (05):355–362

21 Ancker JS, Kern LM, Edwards A, et al; HITEC Investigators. How is the electronic health record being used? Use of EHR data to assess physician-level variability in technology use. J Am Med Inform Assoc 2014;21(06):1001–1008

22 Flemming D, Przysucha M, Hübner U. Cognitive maps to visualise clinical cases in handovers. Design, implementation, usability, and attractiveness testing. Methods Inf Med 2015;54(05):412–423

23 Schuster KM, Jenq GY, Thung SF, et al. Electronic handoff instruments: a truly multidisciplinary tool? J Am Med Inform Assoc 2014;21(e2, no. e2):e352–e357

24 Rinner C, Janzek-Hawlat S, Sibinovic S, Duftschmid G. Semantic validation of standard-based electronic health record documents with W3C XML schema. Methods Inf Med 2010;49(03):271–280

25 Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. J Am Med Inform Assoc 2015;22(01):76–85

26 Späth MB, Grimson J. Applying the archetype approach to the database of a biobank information management system. Int J Med Inform 2011;80(03):205–226

27 Duftschmid G, Wrba T, Rinner C. Extraction of standardized archetyped data from electronic health record systems based on the Entity-Attribute-Value Model. Int J Med Inform 2010;79(08): 585–597

28 Sánchez-de-Madariaga R, Muñoz A, Cáceres J, et al. ccML, a new mark-up language to improve ISO/EN 13606-based electronic health record extracts practical edition. J Am Med Inform Assoc 2013;20(02):298–304

29 D'Amore JD, Mandel JC, Kreda DA, et al. Are meaningful use Stage 2 certified EHRs ready for interoperability? Findings from the SMART C-CDA Collaborative. J Am Med Inform Assoc 2014;21 (06):1060–1068

30 Bache R, Taweel A, Miles S, Delaney BC. An eligibility criteria query language for heterogeneous data warehouses. Methods Inf Med 2015;54(01):41–44

31 Warner JL, Maddux SE, Hughes KS, et al. Development, implementation, and initial evaluation of a foundational open interoperability standard for oncology treatment planning and summarization. J Am Med Inform Assoc 2015;22(03):577–586

32 Chen ES, Carter EW, Winden TJ, Sarkar IN, Wang Y, Melton GB. Multi-source development of an integrated model for family health history. J Am Med Inform Assoc 2015;22(no. e1):e67–e80

33 Blobel B, Goossen W, Brochhausen M. Clinical modeling–a critical analysis. Int J Med Inform 2014;83(01):57–69

34 van der Bij S, Khan N, Ten Veen P, de Bakker DH, Verheij RA. Improving the quality of EHR recording in primary care: a data quality feedback tool. J Am Med Inform Assoc 2016;24(01):81–87

35 Ronquillo JG, Li C, Lester WT. Genetic testing behavior and reporting patterns in electronic medical records for physicians trained in a primary care specialty or subspecialty. J Am Med Inform Assoc 2012;19(04):570–574

36 Shoenbill K, Fost N, Tachinardi U, Mendonca EA. Genetic data and electronic health records: a discussion of ethical, logistical and technological considerations. J Am Med Inform Assoc 2014;21 (01):171–180

37 Alterovitz G, Warner J, Zhang P, et al. SMART on FHIR Genomics: facilitating standardized clinico-genomic apps. J Am Med Inform Assoc 2015;22(06):1173–1178

38 Shirts BH, Salama JS, Aronson SJ, et al. CSER and eMERGE: current and potential state of the display of genetic information in the

electronic health record. J Am Med Inform Assoc 2015;22(06): 1231–1242

39 Manzi SF, Fusaro VA, Chadwick L, et al. Creating a scalable clinical pharmacogenomics service with automated interpretation and medical record result integration - experience from a pediatric tertiary care facility. J Am Med Inform Assoc 2016;24(01):74–80

40 Hoffman JM, Dunnenberger HM, Kevin Hicks J, et al. Developing knowledge resources to support precision medicine: principles from the Clinical Pharmacogenetics Implementation Consortium (CPIC). J Am Med Inform Assoc 2016;23(04):796–801

41 Xu H, Jiang M, Oetjens M, et al. Facilitating pharmacogenetic studies using electronic health records and natural-language processing: a case study of warfarin. J Am Med Inform Assoc 2011;18(04):387–391

42 Carrell DS, Cronkite D, Palmer RE, et al. Using natural language processing to identify problem usage of prescription opioids. Int J Med Inform 2015;84(12):1057–1064

43 Bobo WV, Pathak J, Kremers HM, et al. An electronic health record driven algorithm to identify incident antidepressant medication users. J Am Med Inform Assoc 2014;21(05):785–791

44 Hasan S, Duncan GT, Neill DB, Padman R. Automatic detection of omissions in medication lists. J Am Med Inform Assoc 2011;18 (04):449–458

45 Schnipper JL, Liang CL, Hamann C, et al. Development of a tool within the electronic medical record to facilitate medication reconciliation after hospital discharge. J Am Med Inform Assoc 2011;18(03):309–313

46 Heyworth L, Paquin AM, Clark J, et al. Engaging patients in medication reconciliation via a patient portal following hospital discharge. J Am Med Inform Assoc 2014;21(e1, no. e1):e157–e162

47 Spasic I, Sarafraz F, Keane JA, Nenadic G. Medication information extraction with linguistic pattern matching and semantic rules. J Am Med Inform Assoc 2010;17(05):532–535

48 Mork JG, Bodenreider O, Demner-Fushman D, et al. Extracting Rx information from clinical narrative. J Am Med Inform Assoc 2010;17(05):536–539

49 Meystre SM, Thibault J, Shen S, Hurdle JF, South BR. Textractor: a hybrid system for medications and reason for their prescription extraction from clinical text documents. J Am Med Inform Assoc 2010;17(05):559–562

50 Uzuner O, Solti I, Xia F, Cadag E. Community annotation experiment for ground truth generation for the i2b2 medication challenge. J Am Med Inform Assoc 2010;17(05):519–523

51 Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. J Am Med Inform Assoc 2010;17(05):514–518

52 Patrick J, Li M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. J Am Med Inform Assoc 2010;17(05):524–527

53 Hamon T, Grabar N. Linguistic approach for identification of medication names and related information in clinical narratives. J Am Med Inform Assoc 2010;17(05):549–554

54 Doan S, Bastarache L, Klimkowski S, Denny JC, Xu H. Integrating existing natural language processing tools for medication extraction from discharge summaries. J Am Med Inform Assoc 2010;17(05):528–531

55 Wilson CJ, Newman J, Tapper S, et al. Multiple locations of advance care planning documentation in an electronic health record: are they easy to find? J Palliat Med 2013;16(09):1089–1094

56 Turley M, Wang S, Meng D, Kanter MH, Garrido T. An information model for automated assessment of concordance between advance care preferences and care delivered near the end of life. J Am Med Inform Assoc 2016;23(e1):e118–e124

57 Klemets J, Toussaint P. Does revealing contextual knowledge of the patient's intention help nurses' handling of nurse calls? Int J Med Inform 2016;86:1–9

58 Woods SS, Evans NC, Frisbee KL. Integrating patient voices into health information for self-care and patient-clinician partnerships: Veterans Affairs design recommendations for patient-

generated data applications. J Am Med Inform Assoc 2016;23 (03):491–495

59 Adler NE, Stead WW. Patients in context–EHR capture of social and behavioral determinants of health. N Engl J Med 2015;372 (08):698–701

60 Hripcsak G, Forrest CB, Brennan PF, Stead WW. Informatics to support the IOM social and behavioral domains and measures. J Am Med Inform Assoc 2015;22(04):921–924

61 Kumar RB, Goren ND, Stark DE, Wall DP, Longhurst CA. Automated integration of continuous glucose monitor data in the electronic health record using consumer technology. J Am Med Inform Assoc 2016;23(03):532–537

62 Reeder B, Meyer E, Lazar A, Chaudhuri S, Thompson HJ, Demiris G. Framing the evidence for health smart homes and home-based consumer health technologies as a public health intervention for independent aging: a systematic review. Int J Med Inform 2013;82(07):565–579

63 Reeder B, Chung J, Le T, Thompson H, Demiris G. Assessing older adults' perceptions of sensor data and designing visual displays for ambient environments. An exploratory study. Methods Inf Med 2014;53(03):152–159

64 Knaup P, Schöpe L. Using data from ambient assisted living and smart homes in electronic health records. Methods Inf Med 2014;53(03):149–151

65 Wright A, Pang J, Feblowitz JC, et al. A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record. J Am Med Inform Assoc 2011;18(06):859–867

66 Tian TY, Zlateva I, Anderson DR. Using electronic health records data to identify patients with chronic pain in a primary care setting. J Am Med Inform Assoc 2013;20(e2, no. e2):e275–e280

67 Wei W-Q, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J Am Med Inform Assoc 2016;23(e1):e20–e27

68 Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc 2013;20(e1, no. e1):e147–e154

69 Overby CL, Pathak J, Gottesman O, et al. A collaborative approach to developing an electronic health record phenotyping algorithm for drug-induced liver injury. J Am Med Inform Assoc 2013;20 (e2, no. e2):e243–e252

70 Zhong VW, Obeid JS, Craiq JB, et al. An efficient approach for surveillance of childhood diabetes by type derived from electronic health record data: the SEARCH for Diabetes in Youth Study. J Am Med Inform Assoc 2016;23(06):1060–1067

71 Byrd RJ, Steinhubl SR, Sun J, Ebadollahi S, Stewart WF. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. Int J Med Inform 2014;83(12):983–992

72 Agarwal V, Podchiyska T, Banda JM, et al. Learning statistical models of phenotypes using noisy labeled training data. J Am Med Inform Assoc 2016;23(06):1166–1173

73 Osborne JD, Wyatt M, Westfall AO, Willig J, Bethard S, Gordon G. Efficient identification of nationally mandated reportable cancer cases using natural language processing and machine learning. J Am Med Inform Assoc 2016;23(06):1077–1084

74 Halpern Y, Horng S, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor and learn framework. J Am Med Inform Assoc 2016;23(04):731–740

75 Friedman C, Elhadad N. Natural language processing in health care and biomedicine. In: Cimino JJ, Shortliffe EH, eds. Biomedical Informatics: Computer Application in Health Care and Biomedicine, 4th ed., Vol. 1. London: Springer; 2014:255–284

76 Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med 2010;2(57):57cm29

77 Payne TH, Corley S, Cullen TA, et al. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. J Am Med Inform Assoc 2015;22(05):1102–1110

78 Shortliffe EH, Cimino JJ (eds). Biomedical Informatics—Computer Applications in Health (4th ed.). Springer; 2014

79 Liaw S-T, Chen HY, Maneze D, et al. Health reform: is routinely collected electronic information fit for purpose? Emerg Med Australas 2012;24(01):57–63

80 Dixon BE, Rosenman M, Xia Y, Grannis SJ. A vision for the systematic monitoring and improvement of the quality of electronic health data. Stud Health Technol Inform 2013;192:884–888

81 Cusack CM, Hripcsak G, Bloomrosen M, et al. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. J Am Med Inform Assoc 2013;20(01):134–140

82 Abramson EL, Malhotra S, Osorio SN, et al. A long-term follow-up evaluation of electronic health record prescribing safety. J Am Med Inform Assoc 2013;20(e1, no. e1):e52–e58

83 Ceusters W, Capolupo M, de Moor G, Devlies J, Smith B. An evolutionary approach to realism-based adverse event representations. Methods Inf Med 2011;50(01):62–73

84 Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. J Am Med Inform Assoc 2012;19(01):79–85

85 Declerck G, Hussain S, Daniel C, et al. Bridging data models and terminologies to support adverse drug event reporting using EHR data. Methods Inf Med 2015;54(01):24–31

86 Cheung K-C, van der Veen W, Bouvy ML, Wensing M, van den Bemt PM, de Smet PA. Classification of medication incidents associated with information technology. J Am Med Inform Assoc 2014;21(e1, no. e1):e63–e70

87 Bates J, Fodeh SJ, Brandt CA, Womack JA. Classification of radiology reports for falls in an HIV study cohort. J Am Med Inform Assoc 2016;23(e1, no. e1):e113–e117

88 Harpaz R, Vilar S, Dumouchel W, et al. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. J Am Med Inform Assoc 2013;20(03):413–419

89 Liu M, McPeek Hinz ER, Matheny ME, et al. Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. J Am Med Inform Assoc 2013;20(03):420–426

90 Mei YY, Marquard J, Jacelon C, DeFeo AL. Designing and evaluating an electronic patient falls reporting system: perspectives for the implementation of health information technology in long-term residential care facilities. Int J Med Inform 2013;82(11):e294–e306

91 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(05):947–953

92 McCart JA, Berndt DJ, Jarman J, Finch DK, Luther SL. Finding falls in ambulatory care clinical documents using statistical text mining. J Am Med Inform Assoc 2013;20(05):906–914

93 Koutkias VG, McNair P, Kilintzis V, et al. From adverse drug event detection to prevention. A novel clinical decision support framework for medication safety. Methods Inf Med 2014;53(06):482–492

94 Missiakos O, Baysari MT, Day RO. Identifying effective computerized strategies to prevent drug-drug interactions in hospital: a user-centered approach. Int J Med Inform 2015;84(08):595–600

95 Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. J Am Med Inform Assoc 2014;21(02):353–362

96 Li Q, Melton K, Lingren T, et al. Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care. J Am Med Inform Assoc 2014;21(05):776–784

97 Topaz M, Seger DL, Goss F, et al. Standard information models for representing adverse sensitivity information in clinical documents. Methods Inf Med 2016;55(02):151–157

98 Ban VS, Madden CJ, Browning T, O'Connell E, Marple BF, Moran B. A novel use of the discrete templated notes within an electronic health record software to monitor resident supervision. J Am Med Inform Assoc 2016;24(e1):e2–e8

99 Flemming D, Hübner U. How to improve change of shift handovers and collaborative grounding and what role does the electronic patient record system play? Results of a systematic literature review. Int J Med Inform 2013;82(07):580–592

100 Krauss JC, Boonstra PS, Vantsevich AV, Friedman CP. Is the problem list in the eye of the beholder? An exploration of consistency across physicians. J Am Med Inform Assoc 2016;23(05):859–865

101 Benham-Hutchins MM, Effken JA. Multi-professional patterns and methods of communication during patient handoffs. Int J Med Inform 2010;79(04):252–267

102 Balka E, Tolar M, Coates S, Whitehouse S. Socio-technical issues and challenges in implementing safe patient handovers: insights from ethnographic case studies. Int J Med Inform 2013;82(12):e345–e357

103 Jang J, Yu SH, Kim C-B, Moon Y, Kim S. The effects of an electronic medical record on the completeness of documentation in the anesthesia record. Int J Med Inform 2013;82(08):702–707

104 Saleem JJ, Flanagan ME, Wilck NR, Demetriades J, Doebbeling BN. The next-generation electronic health record: perspectives of key leaders from the US Department of Veterans Affairs. J Am Med Inform Assoc 2013;20(e1, no. e1):e175–e177

105 Braga RD; Panel of Specialists in Health. A multiprofessional information model for Brazilian primary care: Defining a consensus model towards an interoperable electronic health record. Int J Med Inform 2016;90:48–57

106 Berges I, Bermudez J, Illarramendi A. Binding SNOMED CT terms to archetype elements. Establishing a baseline of results. Methods Inf Med 2015;54(01):45–49

107 Mo H, Thompson WK, Rasmussen LV, et al. Desiderata for computable representations of electronic health records-driven phenotype algorithms. J Am Med Inform Assoc 2015;22(06):1220–1230

108 Milberg JA. Development, use, and integration of a nationally-distributed HIV/AIDS electronic health information system. J Am Med Inform Assoc 2016;23(06):1190–1194

109 Kush R, Goldman M. Fostering responsible data sharing through standards. N Engl J Med 2014;370(23):2163–2165

110 Liu D, Wang X, Pan F, et al. Harmonization of health data at national level: a pilot study in China. Int J Med Inform 2010;79(06):450–458

111 Bauer CR, Ganslandt T, Baum B, et al. Integrated data repository toolkit (IDRT). A suite of programs to facilitate health analytics on heterogeneous medical data. Methods Inf Med 2016;55(02):125–135

112 Chen C, Haddad D, Selsky J, et al. Making sense of mobile health data: an open architecture to improve individual- and population-level health. J Med Internet Res 2012;14(04):e112

113 Nogueira JR, Cook TW, Cavalini LT. Mapping a nursing terminology subset to openEHR archetypes. A case study of the international classification for nursing practice. Methods Inf Med 2015;54(03):271–275

114 Hägglund M, Chen R, Koch S. Modeling shared care plans using CONTsys and openEHR to support shared homecare of the elderly. J Am Med Inform Assoc 2011;18(01):66–69

115 Bettencourt-Silva J, De La Iglesia B, Donell S, Rayward-Smith V. On creating a patient-centric database from multiple hospital information systems. Methods Inf Med 2012;51(03):210–220

116 Moreno-Conde A, Jódar-Sánchez F, Kalra D. Requirements for clinical information modelling tools. Int J Med Inform 2015;84(07):524–536

117 Mandl KD, Kohane IS, McFadden D, et al. Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): architecture. J Am Med Inform Assoc 2014;21(04):615–620

118 Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for

electronic health records. J Am Med Inform Assoc 2016;23(05):899–908

119 Marcos C, González-Ferrer A, Peleg M, Cavero C. Solving the interoperability challenge of a distributed complex patient guidance system: a data integrator based on HL7's Virtual Medical Record standard. J Am Med Inform Assoc 2015;22(03):587–599

120 Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015;17(06):e137

121 Mandl KD, Mandel JC, Murphy SN, et al. The SMART Platform: early experience enabling substitutable applications for electronic health records. J Am Med Inform Assoc 2012;19(04):597–603

122 Legaz-García MC, Menárguez-Tortosa M, Fernández-Breis JT, Chute CG, Tao C. Transformation of standardized clinical models based on OWL technologies: from CEM to OpenEHR archetypes. J Am Med Inform Assoc 2015;22(03):536–544

123 Goldspiel BR, Flegel WA, DiPatrizio G, et al. Integrating pharmacogenetic information and clinical decision support into the electronic health record. J Am Med Inform Assoc 2014;21(03):522–528

124 Kullo IJ, Fan J, Pathak J, Savova GK, Ali Z, Chute CG. Leveraging informatics for genetic studies: use of the electronic medical record to enable a genome-wide association study of peripheral arterial disease. J Am Med Inform Assoc 2010;17(05):568–574

125 Nishimura AA, Shirts BH, Salama J, Smith JW, Devine B, Tarczy-Hornoch P. Physician perspectives of CYP2C19 and clopidogrel drug-gene interaction active clinical decision support alerts. Int J Med Inform 2016;86:117–125

126 Warner JL, Rioth MJ, Mandl KD, et al. SMART precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. J Am Med Inform Assoc 2016;23(04):701–710

127 Anand V, Rosenman MB, Downs SM. Translating genome wide association study results to associations among common diseases: in silico study with an electronic medical record. Int J Med Inform 2013;82(09):864–874

128 Yang H. Automatic extraction of medication information from medical discharge summaries. J Am Med Inform Assoc 2010;17 (05):545–548

129 Turley M, Wang S, Meng D, Kanter MH, Garrido T. An information model for automated assessment of concordance between advance care preferences and care delivered near the end of life. J Am Med Inform Assoc 2016;23(e1, no. e1):e118–e124

130 Cascade E, Marr P, Winslow M, Burgess A, Nixon M. Conducting research on the Internet: medical record data integration with patient-reported outcomes. J Med Internet Res 2012;14(05):e137

131 Estabrooks PA, Boyle M, Emmons KM, et al. Harmonized patient-reported data elements in the electronic health record: supporting meaningful use by primary care action on health behaviors and key psychosocial factors. J Am Med Inform Assoc 2012;19(04):575–582

132 Chung AE, Basch EM. Incorporating the patient's voice into electronic health records through patient-reported outcomes as the "review of systems". J Am Med Inform Assoc 2015;22(04):914–916

133 Ohno-Machado L. Informatics 2.0: implications of social media, mobile health, and patient-reported outcomes for healthcare and individual privacy. J Am Med Inform Assoc 2012;19(05):683–683

134 Harle CA, Listhaus A, Covarrubias CM, et al. Overcoming barriers to implementing patient-reported outcomes in an electronic health record: a case report. J Am Med Inform Assoc 2016;23 (01):74–79

135 White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. J Am Med Inform Assoc 2013;20(03):404–408

136 Richesson RL, Rusincovitch SA, Wixted D, et al. A comparison of phenotype definitions for diabetes mellitus. J Am Med Inform Assoc 2013;20(e2, no. e2):e319–e326

137 Chen Y, Carroll RJ, Hinz ER, et al. Applying active learning to high-throughput phenotyping algorithms for electronic health records data. J Am Med Inform Assoc 2013;20(e2, no. e2):e253–e259

138 Shivade C, Raghavan P, Fosler-Lussier E, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc 2014;21(02):221–230

139 Falck S, Adimadhyam S, Meltzer DO, Walton SM, Galanter WL. A trial of indication based prescribing of antihypertensive medications during computerized order entry to improve problem list documentation. Int J Med Inform 2013;82(10):996–1003

140 Ni Y, Kennebeck S, Dexheimer JW, et al. Automated clinical trial eligibility prescreening: increasing the efficiency of patient identification for clinical trials in the emergency department. J Am Med Inform Assoc 2015;22(01):166–178

141 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;20(e2, no. e2):e334–e340

142 Bellows BK, LaFleur J, Kamauu AW, et al. Automated identification of patients with a diagnosis of binge eating disorder from narrative electronic health records. J Am Med Inform Assoc 2014;21(e1, no. e1):e163–e168

143 Lin C, Karlson EW, Dligach D, et al. Automatic identification of methotrexate-induced liver toxicity in patients with rheumatoid arthritis from the electronic medical record. J Am Med Inform Assoc 2015;22(e1, no. e1):e151–e161

144 Fan J, Arruda-Olson AM, Leibson CL, et al. Billing code algorithms to identify cases of peripheral artery disease from administrative data. J Am Med Inform Assoc 2013;20(e2, no. e2):e349–e354

145 Miotto R, Weng C. Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. J Am Med Inform Assoc 2015;22(e1, no. e1):e141–e150

146 Wei W-Q, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J Am Med Inform Assoc 2016;23(e1, no. e1):e20–e27

147 Abhyankar S, Demner-Fushman D, Callaghan FM, McDonald CJ. Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. J Am Med Inform Assoc 2014;21(05):801–807

148 Walker AM, Zhou X, Ananthakrishnan AN, et al. Computer-assisted expert case definition in electronic health records. Int J Med Inform 2016;86:62–70

149 Rosenman M, He J, Martin J, et al. Database queries for hospitalizations for acute congestive heart failure: flexible methods and validation based on set theory. J Am Med Inform Assoc 2014;21 (02):345–352

150 Nelson HD, Weerasinghe R, Martel M, et al. Development of an electronic breast pathology database in a community health system. J Pathol Inform 2014;5(01):26

151 Richesson RL, Hammond WE, Nahm M, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. J Am Med Inform Assoc 2013;20(e2, no. e2):e226–e231

152 Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. J Am Med Inform Assoc 2013;20(e2, no. e2):e206–e211

153 Pradhan S, Elhadad N, South BR, et al. Evaluating the state of the art in disorder recognition and normalization of the clinical narrative. J Am Med Inform Assoc 2015;22(01):143–154

154 D'Avolio LW, Nguyen TM, Farwell WR, et al. Evaluation of a generalizable approach to clinical information retrieval using the automated retrieval console (ARC). J Am Med Inform Assoc 2010;17(04):375–382

155 Tate AR, Beloff N, Al-Radwan B, et al. Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface. J Am Med Inform Assoc 2014;21(02):292–298

156 Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc 2016;23(05):1007–1015

157 Hota B, Lin M, Doherty JA, et al; CDC Prevention Epicenter Program. Formulation of a model for automating infection surveillance: algorithmic detection of central-line associated bloodstream infection. J Am Med Inform Assoc 2010;17(01):42–48

158 Gultepe E, Green JP, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. J Am Med Inform Assoc 2014;21(02):315–325

159 Wiley LK, Shah A, Xu H, Bush WS. ICD-9 tobacco use codes are effective identifiers of smoking status. J Am Med Inform Assoc 2013;20(04):652–658

160 Lyalina S, Percha B, LePendu P, Iyer SV, Altman RB, Shah NH. Identifying phenotypic signatures of neuropsychiatric disorders from electronic medical records. J Am Med Inform Assoc 2013;20 (e2, no. e2):e297–e305

161 Strauss JA, Chao CR, Kwan ML, Ahmed SA, Schottinger JE, Quinn VP. Identifying primary and recurrent cancers using a SAS-based natural language processing algorithm. J Am Med Inform Assoc 2013;20(02):349–355

162 Wei W-Q, Leibson CL, Ransom JE, et al. Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. J Am Med Inform Assoc 2012; 19(02):219–224

163 Ye Y, Tsui FR, Wagner M, Espino JU, Li Q. Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. J Am Med Inform Assoc 2014;21(05):815–823

164 Cartagena FP, Schaeffer M, Rifai D, Doroshenko V, Goldberg HS. Leveraging the NLM map from SNOMED CT to ICD-10-CM to facilitate adoption of ICD-10-CM. J Am Med Inform Assoc 2015; 22(03):659–670

165 Heintzelman NH, Taylor RJ, Simonsen L, et al. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text. J Am Med Inform Assoc 2013;20(05):898–905

166 Pathak J, Wang J, Kashyap S, et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. J Am Med Inform Assoc 2011;18(04):376–386

167 Mani S, Ozdas A, Aliferis C, et al. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. J Am Med Inform Assoc 2014;21(02):326–336

168 Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc 2013;20(01):117–121

169 Marafino BJ, Davies JM, Bardach NS, Dean ML, Dudley RA. N-gram support vector machines for scalable procedure and diagnosis classification, with applications to clinical free text data from the intensive care unit. J Am Med Inform Assoc 2014;21(05):871–875

170 Pathak J, Bailey KR, Beebe CE, et al. Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium. J Am Med Inform Assoc 2013; 20(e2, no. e2):e341–e348

171 Kirby JC, Speltz P, Rasmussen LV, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc 2016;23(06):1046–1052

172 Bejan CA, Xia F, Vanderwende L, Wurfel MM, Yetisgen-Yildiz M. Pneumonia identification using statistical feature selection. J Am Med Inform Assoc 2012;19(05):817–823

173 Carroll RJ, Thompson WK, Eyler AE, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. J Am Med Inform Assoc 2012;19(e1, no. e1):e162–e169

174 Syed-Abdul S, Moldovan M, Nguyen PA, et al. Profiling phenome-wide associations: a population-based observational study. J Am Med Inform Assoc 2015;22(04):896–899

175 Xu J, Rasmussen LV, Shaw PL, et al. Review and evaluation of electronic health records-driven phenotype algorithm authoring tools for clinical and translational research. J Am Med Inform Assoc 2015;22(06):1251–1260

176 Warner JL, Denny JC, Kreda DA, Alterovitz G. Seeing the forest through the trees: uncovering phenomic complexity through interactive network visualization. J Am Med Inform Assoc 2015; 22(02):324–329

177 Warner JL, Zollanvari A, Ding Q, Zhang P, Snyder GM, Alterovitz G. Temporal phenome analysis of a large electronic health record cohort enables identification of hospital-acquired complications. J Am Med Inform Assoc 2013;20(e2, no. e2):e281–e287

178 Wei W-Q, Leibson CL, Ransom JE, Kho AN, Chute CG. The absence of longitudinal data limits the accuracy of high-throughput clinical phenotyping for identifying type 2 diabetes mellitus subjects. Int J Med Inform 2013;82(04):239–247

179 Buckley JM, Coopey SB, Sharko J, et al. The feasibility of using natural language processing to extract clinical information from breast pathology reports. J Pathol Inform 2012;3:23

180 Yu S, Liao KP, Shaw SY, et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. J Am Med Inform Assoc 2015;22(05):993–1000

181 Huang SH, LePendu P, Iyer SV, Tai-Seale M, Carrell D, Shah NH. Toward personalizing treatment for depression: predicting diagnosis and severity. J Am Med Inform Assoc 2014;21(06): 1069–1075

182 Chase HS, Radhakrishnan J, Shirazian S, Rao MK, Vawdrey DK. Under-documentation of chronic kidney disease in the electronic health record in outpatients. J Am Med Inform Assoc 2010;17 (05):588–594

183 Kho AN, Hayes MG, Rasmussen-Torvik L, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. J Am Med Inform Assoc 2012;19(02):212–218

184 Mahajan R, Moorman AC, Liu SJ, Rupp L, Klevens RM; Chronic Hepatitis Cohort Study (CHeCS) investigators*. Use of the International Classification of Diseases, 9th revision, coding in identifying chronic hepatitis B virus infection in health system data: implications for national surveillance. J Am Med Inform Assoc 2013;20(03):441–445

185 Savard N, Bédard L, Allard R, Buckeridge DL. Using age, triage score, and disposition data from emergency department electronic records to improve Influenza-like illness surveillance. J Am Med Inform Assoc 2015;22(03):688–696

186 Trinh N-HT, Youn SJ, Sousa J, et al. Using electronic medical records to determine the diagnosis of clinical depression. Int J Med Inform 2011;80(07):533–540

187 Rahimi A, Liaw S-T, Taggart J, Ray P, Yu H. Validating an ontology-based algorithm to identify patients with type 2 diabetes mellitus in electronic health records. Int J Med Inform 2014;83(10): 768–778