

Original Paper

# Analyzing Patient Secure Messages Using a Fast Health Care Interoperability Resources (FHIR)–Based Data Model: Development and Topic Modeling Study

Amrita De<sup>1\*</sup>, MA; Ming Huang<sup>1\*</sup>, PhD; Tinghao Feng<sup>2</sup>, MSc; Xiaomeng Yue<sup>3</sup>, PhD; Lixia Yao<sup>1</sup>, PhD

<sup>1</sup>Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, United States

<sup>2</sup>Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC, United States

<sup>3</sup>Division of Pharmacy Practice and Administrative Sciences, James L. Winkle College of Pharmacy, University of Cincinnati, Cincinnati, OH, United States

\*these authors contributed equally

**Corresponding Author:**

Lixia Yao, PhD

Department of Artificial Intelligence and Informatics

Mayo Clinic

200 1st St SW

Rochester, MN

United States

Phone: 1 5072842511

Email: [lixia.cn.yao@gmail.com](mailto:lixia.cn.yao@gmail.com)

## Abstract

**Background:** Patient portals tethered to electronic health records systems have become attractive web platforms since the enacting of the Medicare Access and Children's Health Insurance Program Reauthorization Act and the introduction of the *Meaningful Use* program in the United States. Patients can conveniently access their health records and seek consultation from providers through secure web portals. With increasing adoption and patient engagement, the volume of patient secure messages has risen substantially, which opens up new research and development opportunities for patient-centered care.

**Objective:** This study aims to develop a data model for patient secure messages based on the Fast Healthcare Interoperability Resources (FHIR) standard to identify and extract significant information.

**Methods:** We initiated the first draft of the data model by analyzing FHIR and manually reviewing 100 sentences randomly sampled from more than 2 million patient-generated secure messages obtained from the online patient portal at the Mayo Clinic Rochester between February 18, 2010, and December 31, 2017. We then annotated additional sets of 100 randomly selected sentences using the Multi-purpose Annotation Environment tool and updated the data model and annotation guideline iteratively until the interannotator agreement was satisfactory. We then created a larger corpus by annotating 1200 randomly selected sentences and calculated the frequency of the identified medical concepts in these sentences. Finally, we performed topic modeling analysis to learn the hidden topics of patient secure messages related to 3 highly mentioned microconcepts, namely, fatigue, prednisone, and patient visit, and to evaluate the proposed data model independently.

**Results:** The proposed data model has a 3-level hierarchical structure of health system concepts, including 3 macroconcepts, 28 mesoconcepts, and 85 microconcepts. Foundation and base macroconcepts comprise 33.99% (841/2474), clinical macroconcepts comprise 64.38% (1593/2474), and financial macroconcepts comprise 1.61% (40/2474) of the annotated corpus. The top 3 mesoconcepts among the 28 mesoconcepts are condition (505/2474, 20.41%), medication (424/2474, 17.13%), and practitioner (243/2474, 9.82%). Topic modeling identified hidden topics of patient secure messages related to fatigue, prednisone, and patient visit. A total of 89.2% (107/120) of the top-ranked topic keywords are actually the health concepts of the data model.

**Conclusions:** Our data model and annotated corpus enable us to identify and understand important medical concepts in patient secure messages and prepare us for further natural language processing analysis of such free texts. The data model could be potentially used to automatically identify other types of patient narratives, such as those in various social media and patient forums. In the future, we plan to develop a machine learning and natural language processing solution to enable automatic triaging solutions to reduce the workload of clinicians and perform more granular content analysis to understand patients' needs and improve patient-centered care.

**KEYWORDS**

patient secure messages; patient portal; data model; FHIR; annotated corpus; topic modeling

## Introduction

### Background

In the United States, the Medicare Access and CHIP (Children's Health Insurance Program) Reauthorization Act [1] and *Meaningful Use* program [2] have incentivized the growing adoption of electronic health records (EHRs) and patient health records with the goal of improving the quality of health care delivery systems. Consequently, many health care delivery systems now offer patient portals, tethered to their EHR systems, that allow patients to access their medical records and communicate with their clinicians through secure messages [3]. Patient portals encourage patients to become equal partners in their care and health management and be more engaged and participatory in shared decision making [4]. After the Health Information Technology for Economic and Clinical Health Act was enacted in 2009, patient portals have gained widespread adoption by health care delivery systems in the United States [5,6]. In 2017, more than 90% of the health care delivery systems, including the Veterans Health Administration, Mass General Brigham, Kaiser Permanente, and the Mayo Clinic, offered patient portal access to their patients [7]. Currently, patients send secure web-based messages to request medical appointments and prescription refills [8,9]. Clinicians send patients appointment reminders and promote timely preventive care [10,11]. Patients and clinicians can communicate back and forth easily and in a timely manner about complex situations such as new symptoms, follow-up visits, medication concerns, and medical questions.

With the increase in the number of patients signing up for these portals, the number of secure messages has risen substantially [12-15]. Unfortunately, the content of the large number of patient secure messages in free-text format has not been processed and analyzed systematically and incorporated into the present EHR systems organically to unfold its potential for improving patient-centered care because of technical hurdles. For instance, existing annotated corpora have been mainly developed for sublanguages such as scientific literature in biomedicine and clinical notes; no annotated corpus is available for developing natural language processing (NLP) capabilities in patient-generated formal language. In this study, we propose to develop a data model of health concepts for patient secure messages based on the Fast Healthcare Interoperability Resources (FHIR) standard [16].

A data model is usually made up of entities that represent important items in the domain and relationship assertions among the entities. In our case, a data model will illustrate the key concepts occurring in patient secure messages and the relationships among them. The data model is critical to the development of any information system (eg, a health information exchange system or NLP-based semantic representation system for a patient portal) by providing the definition of the concepts and format of data. Building the factual and useful data model

requires a deep understanding of the underlying process and data. Therefore, we also create a large annotated corpus for analyzing the contents of sampled patient secure messages to better understand patients' concerns. Once complete, we further apply topic modeling techniques independently to investigate whether the patients' focuses and concerns in 3 common medical conditions align with the developed data model. We build the annotated corpus primarily to build the data model, and topic modeling can serve as an independent and primitive validation of the data model. We choose topic modeling instead of information extraction because we build the annotated corpus primarily to build the data model. Topic modeling, as an unsupervised method, generates results independently of the corpus and thus can serve as an independent validation of the data model. We expect a much more rigorous evaluation and validation by building collaborations and partnerships with domestic and international researchers in the field.

With all the necessary preparation, our ultimate objective is to develop an NLP system that will automatically identify and extract significant information from unstructured patient secure messages for the purpose of automatically triaging patient secure messages, reducing the workload of clinicians by chatbot, and performing more granular and sophisticated content analysis to understand patients' needs and improve shared decision making and patient-centered care.

### Related Work

As patient secure messages are relatively new, very little research has focused on automatically identifying and standardizing their content despite their important implications. North et al [17] analyzed the content of 6430 secure messages to assess the overall risk associated with the messages and to determine whether patients were using portal messages for symptoms requiring urgent evaluation. Their study showed that patients used portal messages 3.5% of the time for potentially high-risk symptoms of chest pain, breathing concerns, abdominal pain, palpitations, lightheadedness, and vomiting. Sulieman et al [18] also developed machine learning models on patient portal secure messages regarding surgical issues to identify message threads that involve medical decision making from a health care provider and to classify the complexity of the decision. Cronin et al [19] built patient portal message classifiers using rule-based and NLP techniques such as the bag-of-words model. They curated a gold standard data set of 3253 portal messages annotated by communication types such as informational, medical, logistical, and social. This study also focuses on developing a data model—a standard framework—to address the issues of content analysis, information extraction, and integration of significant information from patient secure messages leveraging Health Level-7 (HL7) FHIR.

HL7 is a nonprofit standard development organization accredited by the American National Standards Institute, and it is dedicated to providing a comprehensive framework and related standards

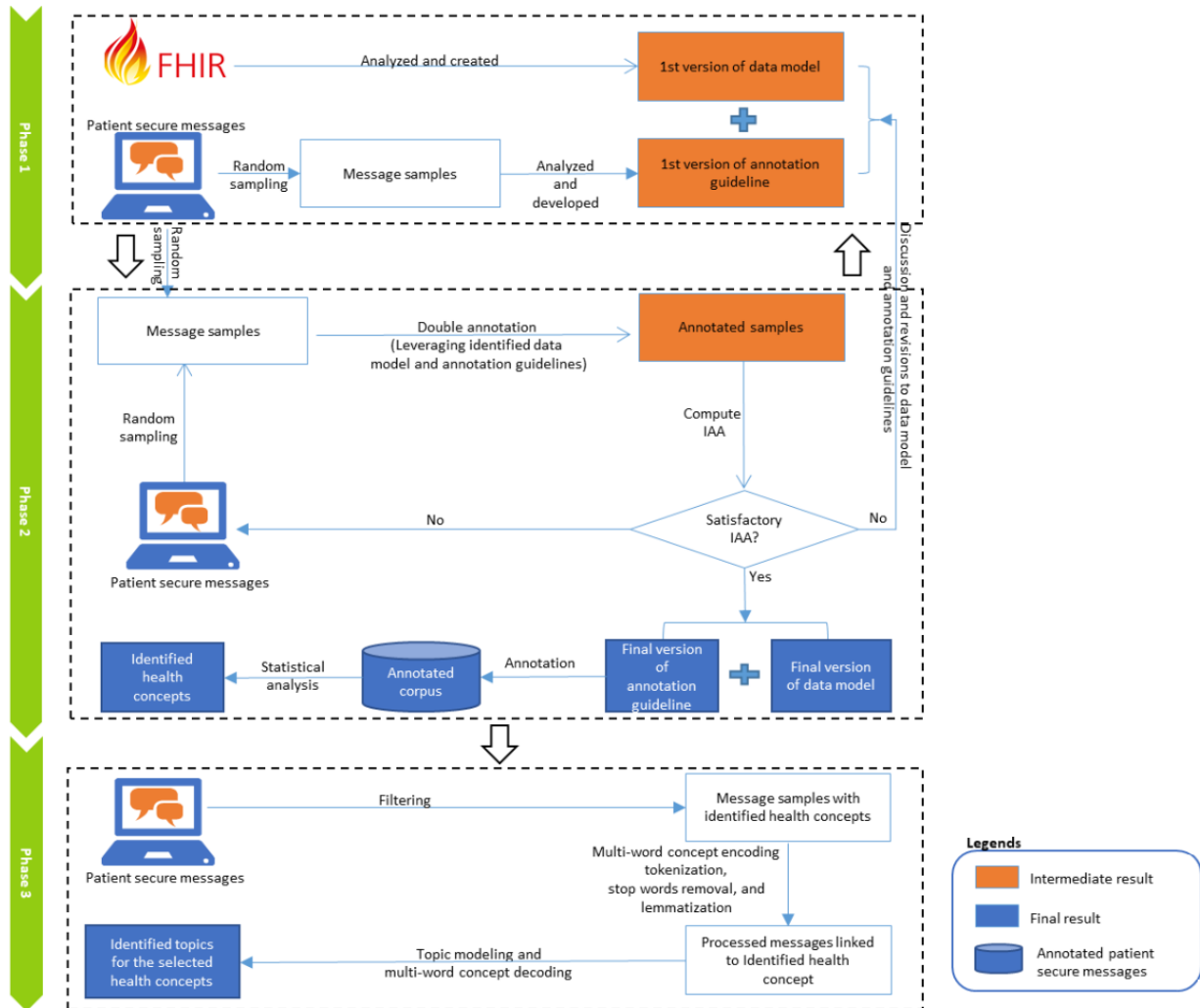
for the data exchange, integration, sharing, and retrieval of electronic health information [20]. FHIR is an improved health data exchange standard that comprehensively defines how information can be exchanged among different systems regardless of how it has been stored and allows health care information to be accessible to those who need it for the benefit of health care quickly and easily [21]. Instead of traditional document-centric approaches, HL7 FHIR takes a modular approach and represents atomic or granular health care data (eg, heart rate, procedure, medication, and allergies) as independent modular entities, concepts, and actions involved in health care information analysis, exchange, and integration as resources [22]. Existing studies on developing data standards have mostly focused on analyzing, extracting, and integrating structured data from EHRs, mobile-based patient health records, and medical apps. In 2018, Hong et al [23] for the first time introduced a scalable and standard-based framework for analyzing and integrating both structured and unstructured EHR data by leveraging the FHIR specification [23]. The scope and use of the FHIR framework do not completely meet the requirements of our study. We aim to develop an HL7 FHIR-based data model that precisely analyzes and extracts patient secure messages.

## Methods

### Overview

A 3-phase workflow for developing the data model and annotated corpus is shown in [Figure 1](#). We collected more than 2 million patient-generated secure messages from Mayo Clinic Patient Online Services [24]. We developed the first draft of the data model and annotation guideline by analyzing FHIR and manually reviewing 100 sentences from the sampled secure messages. We then randomly selected, annotated, and examined additional sets of 100 sentences from the secure messages to iteratively update the data model and annotation guideline until interannotator agreement (IAA) was achieved. Subsequently, we created the annotated corpus by annotating 1200 sentences from the randomly selected 2100 sentences. Finally, we calculated the frequency of the identified health concepts in the annotated corpus and performed topic modeling to extract hidden topics of all the patient secure messages linked to frequently mentioned health concepts. In the following sections, we will discuss data collection and preprocessing, design of the data model, development of the annotation guideline, creation and analysis of the annotated corpus, and topic modeling in more detail. The entire data model can be found in [Multimedia Appendix 1](#). The details of data set collection and processing and annotation and topic modeling and annotation guideline can be found in [Multimedia Appendix 2](#) [16,24-33] and [Multimedia Appendix 3](#) [1-3,16,24].

**Figure 1.** The workflow for developing the data model for patient secure messages, annotated corpus, and topic modeling analysis. FHIR: Fast Healthcare Interoperability Resources; IAA: interannotator agreement.



### Ethics Approval

No patient was exposed to any intervention. We used data from the Mayo Clinic Unified Data Platform to develop the annotated corpus and for the analysis. The study was approved by the Mayo Clinic Institutional Review Board (19-002211).

### Data Set Collection and Preprocessing

The Mayo Clinic Rochester started the patient portal (Patient Online Services) in 2010 for primary care practice and later extended it to specialty practice in 2013 [24]. We collected more than 2 million patient-generated secure messages from the online patient portal between February 18, 2010, and December 31, 2017. We removed messages with empty message bodies. Each message has a unique message ID, previous message ID, initial message ID, sender ID, recipient ID, date and time of the message creation, message subject, and message body. As we have mentioned earlier, the details are in Multimedia Appendix 2.

### Design of a Data Model for Patient Secure Messages

Creating a data model for unstructured patient narratives is a very challenging task. After the literature review and initial analysis of the sampled secure messages, we decided to use the

FHIR standard to develop the data model (Multimedia Appendix 1) because it comprehensively represents the modular entities, concepts, and actions involved in health care information exchange.

FHIR defines a hierarchical set of core and infrastructure resources for handling health concepts in an EHR [34]. After analyzing version 4 of the FHIR standard [16] and the sampled secure messages, we generated the first draft of the data model with 3 hierarchical levels—macroconcepts, mesoconcepts, and microconcepts—to extract information from the patient secure messages. We merged and revised some concepts from FHIR after analyzing the messages. For example, we merged 2 similar but separately defined top-level concepts (foundation concept and base concept) in FHIR into 1 macroconcept (foundation and base concept) in the data model. We also introduced a microconcept—*unspecified*—as an attribute under all mesoconcepts. *Unspecified* refers to the general terms under most of the mesoconcepts that cannot be categorized into any specific microconcepts. We deleted some mesoconcepts under clinical macroconcepts such as *clinical impression*, *detected issue*, *medication knowledge*, *molecular sequence*, and *care team* because they are not relevant to patient secure messages. The data model also underwent rounds of revisions during the annotation process to handle inconsistencies, and disagreement



occurred between the annotators. For instance, in the first draft of the data model, the mesoconcept *patient* has 7 microconcepts following FHIR, such as *identifier*, *name*, *telecom*, *gender*, *birthdate*, *address*, and *marital status*. In the final data model, the mesoconcept *patient* has 4 microconcepts, such as *unspecified*, *privacy*, *lifestyle*, and *diet*, to better understand patients' medical records and history. All personal information related to the patient is kept under *privacy* to maintain data privacy.

### Development of an Annotation Guideline

For any annotation task, it is important that all annotators follow the same standard (annotation guideline) to minimize annotation confusion and errors. We manually reviewed several sets of 100 randomly selected message sentences to develop an annotation guideline. More specifically, we created the first version of our annotation guideline by analyzing the first set of 100 message sentences. Subsequently, our annotators independently annotated another set of 100 message sentences to examine the effectiveness of the annotation guideline until the annotators reached a considerable amount of agreement.

The final annotation guideline consists of 2 sections: (1) the first section discusses the general annotation rules; (2) the second section describes the health concepts and associative rules for the identification and extraction of these concepts together with specific examples. For instance, it is challenging to differentiate 2 microconcepts (*name* and *symptoms*) under the mesoconcept *condition*. The *condition-name* microconcept refers to the name of a disease and/or medical condition (eg, *rheumatoid arthritis*, *diabetes mellitus*, and *influenza*). The *condition-symptom* microconcept denotes a physical or mental feature or symptom (eg, *sore bottom* and *numb arm*) of a disease and/or medical condition. As per general annotation rules, all modifiers (eg, adjectives and adverbs) and possessives are removed for annotation to make text spans consistent. For the cases *special calcium pill*, *this medicine*, and *my rheumatoid arthritis*, the modifiers and possessives *special*, *this*, and *my* are not considered for annotation. All patient private information (eg, name, identity number, and clinic number) is not disclosed to maintain data privacy. The guideline clearly defines the concepts and rules to lessen the scope of ambiguity and error in the annotation and increase the possibility of agreement among the annotators to develop a quality corpus. The complete guideline for annotating patient secure messages is listed in [Multimedia Appendix 3](#).

### Development and Analysis of an Annotated Corpus of Patient Secure Messages

We chose the Multi-purpose Annotation Environment tool to annotate patient secure messages because of its ease of use and ability to fix discrepancies [25]. The Multi-purpose Annotation Environment tool requires a document type definition file with concept tags and attributes. We chose 2 professional annotators—a clinically trained linguist and a student pharmacist—to create a standard error-free corpus. They initially analyzed and revised several sets of 100 randomly sampled message sentences to iteratively improve the data model and annotation guideline. To check the consistency between the annotators, we calculated the F1 score of 2 separate annotation

sets as the IAA score using General Architecture for Text Engineering software [26]. The IAA scores were computed using the F1 score as a criterion at the level of entities. This helped us to understand the span of the concepts on which the annotators agreed, disagreed, and partially agreed. We decided to follow the lenient parameter for measuring the IAA. With the lenient approach, the annotations that overlap are counted as a partial match, in contrast with the strict approach in which the annotations have to match with one another completely. We can consider this sentence as an example: “My mother has severe sinus headaches for several months.” Now, if one annotator annotates “sinus headache” and another annotates “severe sinus headache,” then the strict IAA approach will give us no match, but the lenient approach will consider this as an overlap. The F1 scores of the first set of annotations were 0.42 (macro mean) and 0.67 (micro mean). After discussing and resolving the disagreements, the annotators annotated another set of the same sentences, and the F1 scores were quite satisfactory: 0.62 (macro mean) and 0.74 (micro mean). Our annotators finally developed a quality corpus of 1200 randomly selected sentences.

After annotation, we performed summary statistics to calculate the frequencies of the identified health concepts (ie, macroconcepts, mesoconcepts, and microconcepts) in the annotated corpus of patient secure messages. The distribution of these health concepts helps us to understand which concepts patients were mostly concerned about and communicated to their health care providers. The details are provided in [Multimedia Appendix 2](#).

### Topic Modeling

After analyzing the annotated corpus, we selected 3 health microconcepts (ie, fatigue, prednisone, and patient visit) as representative cases for topic modeling analysis. The chosen microconcepts and the corresponding meso concepts and macroconcepts were frequently discussed in the patient portal messages (refer to the *Results* section for more details). Fatigue is an instance of a top-mentioned microconcept, symptom, under the condition mesoconcept and clinical macroconcept. Prednisone is a case of a microconcept, name, under the medication mesoconcept and clinical macroconcept, about which patients have expressed most concern in the patient secure messages. Patient visit is an example of a largely discussed microconcept, type, under the appointment mesoconcept and the foundation and base macroconcept.

After multi-word concept encoding and health concept recognition using MetaMap [27], we collected 41,490, 27,743, and 95,533 patient secure messages that mentioned the health microconcepts fatigue, prednisone, and patient visit, respectively, to examine the focus of those messages. MetaMap is a highly configurable program. It has been developed by Dr Alan Aronson at the National Library of Medicine to map biomedical texts to the Unified Medical Language System.

Topic modeling automatically identifies topics or themes in a large collection of documents in terms of a set of keywords that occur together and most frequently [35,36]. We used latent Dirichlet allocation (LDA) [28], a state-of-the-art unsupervised topic modeling method as implemented in Machine Learning

for Language Toolkit [29], to learn the hidden topics of patient secure messages related to each of the 3 health microconcepts. After tokenization, stop word removal, and lemmatization, each patient secure message was converted into a vocabulary vector where the elements were the frequency of each lemma (including Concept Unique Identifier encoded by MetaMap) without considering the order of the lemma.

We quantitatively calculated topic coherence [30] and asked domain experts to qualitatively evaluate the learned topics. More specifically, we evaluated the topic coherence at different topic numbers (ie, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30) to determine the optimal topic number for the 3 selected health microconcepts. We found that the optimal topic number was 8 for fatigue-related messages, 10 for prednisone-related messages, and 12 for patient visit-related messages. We also investigated the hyperparameters of and in the LDA. controls the topic distributions over a document. A smaller results in fewer topics that are statistically associated with a document. determines the word distributions over a topic. A smaller leads to fewer words that are statistically linked to a topic. We set the automatic optimization of the hyperparameters of and in Machine Learning for Language Toolkit (Mallet) during topic modeling.

## Results

### Data Model for Patient Secure Messages

The data model for patient secure messages has a 3-level hierarchical structure consisting of 3 macroconcepts, 28

mesoconcepts, and 85 microconcepts. [Textbox 1](#) provides a partial illustration of the data model (refer to [Multimedia Appendix 1](#) for the full data model). The 3 macroconcepts in the data model are the foundation and base macroconcept, clinical macroconcept, and financial macroconcept. Foundation and base concepts are the basic infrastructure of the health care system concepts on which the rest of the specifications are built. The mesoconcepts of the foundation and base macroconcept include patient, practitioner, related person, organization, health care service, device, appointment, encounter, and document reference. Clinical macroconcepts refer to core clinical components, including allergy intolerance, adverse event, body structure, specimen, condition, procedure, family member history, observation, laboratory test, imaging, medication, immunization, care plan, care team, referral, and risk. The financial macroconcepts cover all finance-related issues such as coverage eligibility, claim payment, account, and explanation of benefits. All the mesoconcepts have been further categorized into microconcepts as attributes, and each mesoconcept has an *unspecified* microconcept. For example, as shown in [Textbox 1](#), the mesoconcept *patient* is categorized as *unspecified*, *privacy*, *lifestyle*, and *diet*. The mesoconcept *coverage-eligibility* has 5 microconcepts, including *unspecified*, *percentage*, *insurance ID*, *benefit category*, *insurer*, and *insurance*.

**Textbox 1.** A partial illustration of the data model for patient secure messages.

Health Care System Concepts and Their Definitions
<p>Foundation and base concepts</p> <ul style="list-style-type: none"> <li>● Patient—Demographic and other administrative information about an individual receiving health care services <ul style="list-style-type: none"> <li>● Unspecified</li> <li>● Privacy</li> <li>● Lifestyle</li> <li>● Diet</li> </ul> </li> <li>● Appointment—A booking of a health care event between patients and practitioners (or other related persons or devices) on a specific date at a specific time <ul style="list-style-type: none"> <li>● Unspecified</li> <li>● Status</li> <li>● Type</li> <li>● Reason</li> </ul> </li> </ul>
<p>Clinical concepts</p> <ul style="list-style-type: none"> <li>● Specimen—A sample taken from a biological entity for laboratory analysis <ul style="list-style-type: none"> <li>● Unspecified</li> <li>● Name</li> </ul> </li> <li>● Laboratory test—Tests (eg, clinical, hematological, or microbiology tests) performed on patients and groups of patients and the results derived from the tests <ul style="list-style-type: none"> <li>● Unspecified</li> <li>● Name</li> <li>● Result</li> </ul> </li> </ul>
<p>Financial concepts</p> <ul style="list-style-type: none"> <li>● Coverage eligibility—Information on patients, insurers, insurance, coverage, plan details, reimbursement, and payment for health care services <ul style="list-style-type: none"> <li>● Unspecified</li> <li>● Percentage</li> <li>● Insurance ID</li> <li>● Benefit category</li> <li>● Insurer</li> <li>● Insurance</li> </ul> </li> <li>● Explanation of benefits—Information on claim details and adjudication details from the processing of claims <ul style="list-style-type: none"> <li>● Unspecified</li> </ul> </li> </ul>

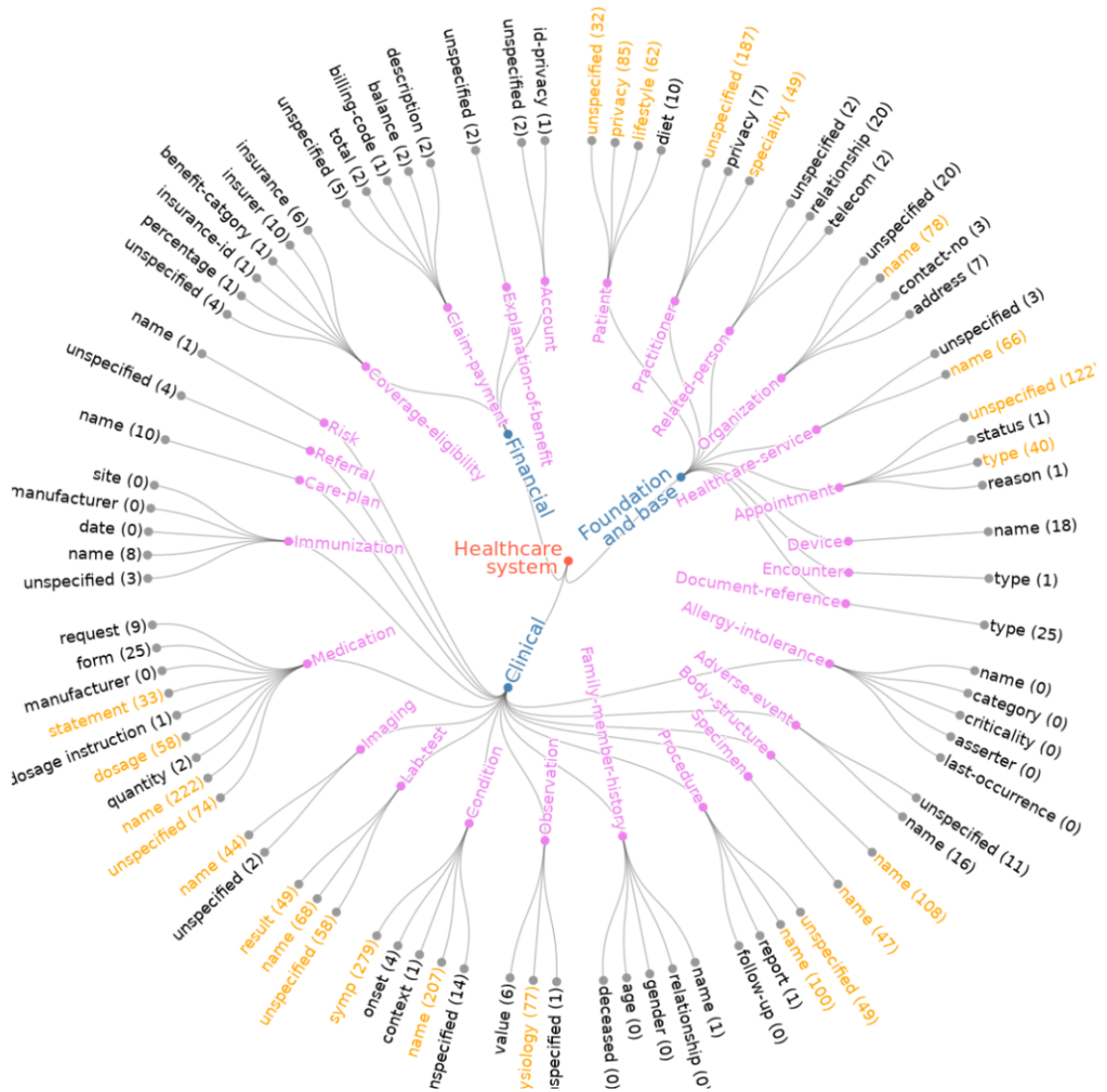
### Annotated Corpus of Patient Secure Messages

We annotated 1200 sentences of patient secure messages based on the data model. [Figure 2](#) illustrates the frequency of hierarchical health concepts in the annotated corpus. The concepts in blue, purple, and black represent macroconcepts, mesoconcepts, and microconcepts, respectively. The concepts in orange illustrate the most frequent microconcepts along with their occurrences in the annotated text. Foundation and base macroconcepts make up 33.99% (841/2474) of the annotated

corpus, clinical macroconcepts make up 64.38% (1593/2474) of the annotated corpus, and financial macroconcepts make up 1.61% (40/2474) of the annotated corpus, respectively. Patients shared some information about insurance, coverage, and payments in the secure messages. Among the 28 mesoconcepts, the most discussed were condition (505/2474, 20.41%), medication (424/2474, 17.13%), practitioner (243/2474, 9.82%), patient (189/2474, 7.63%), laboratory test (175/2474, 7.07%), appointment (164/2474, 6.62%), procedure (150/2474, 6.06%), and organization (108/2474, 4.36%). The most frequently used

microconcepts were various condition names (eg, fatigue), patient visit), medication names (eg, prednisone), and appointment types (eg,

**Figure 2.** Radial tree to illustrate hierarchical health concepts in the annotated corpus.



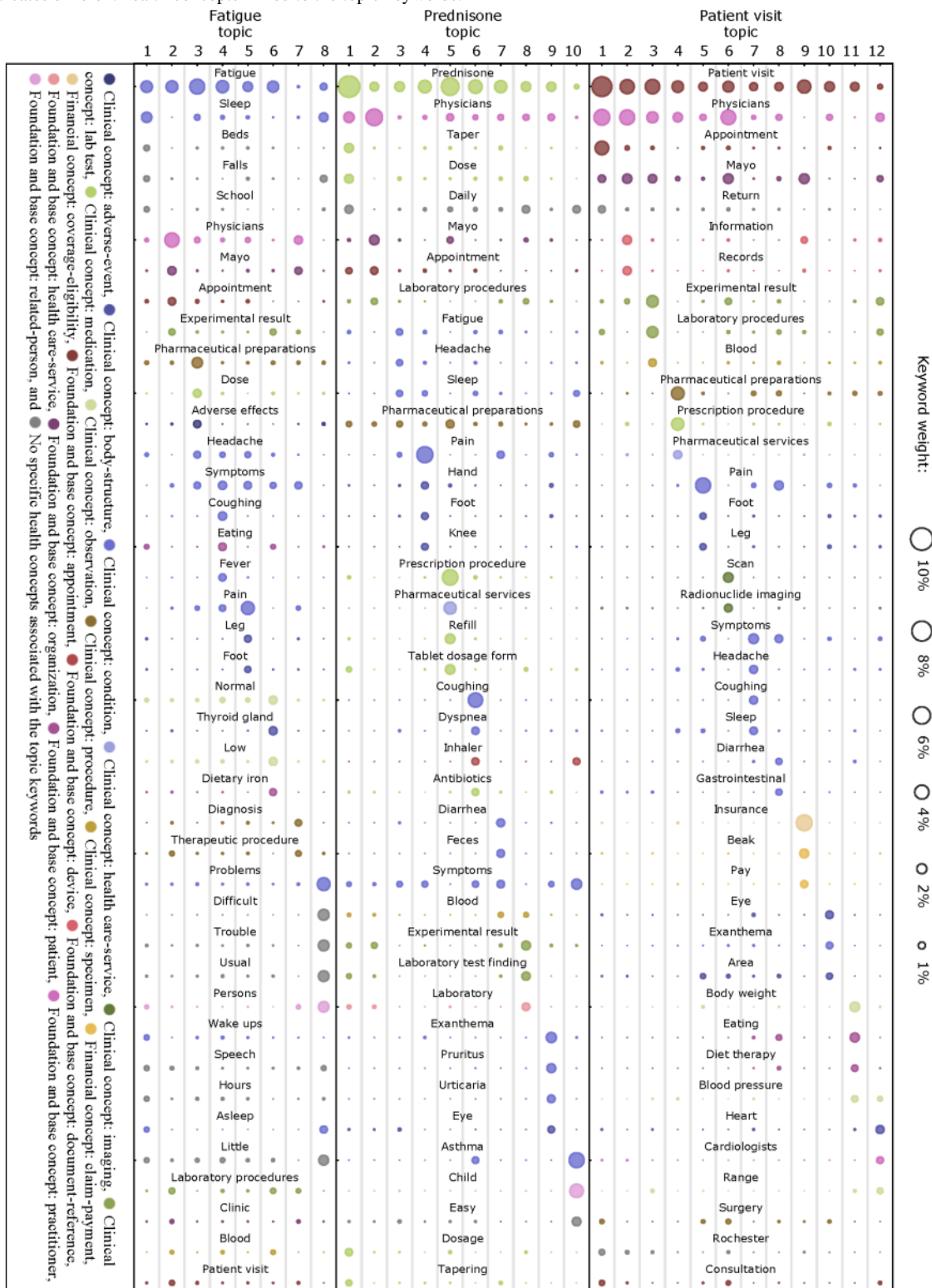
**Topics of Patient Secure Messages**

After identifying the frequently discussed health concepts, we used a topic modeling technique to discover the latent topics in terms of a group of keywords for patient secure messages mentioning fatigue, prednisone, and patient visit. As shown in Figure 3, we highlight 8, 10, and 12 meaningful topics of patient secure messages related to fatigue, prednisone, and patient visit,

respectively. We use a color scheme to represent the top health microconcepts in the topics of patient secure messages associated with fatigue, prednisone, and patient visit. More topic keywords and details can be found in Multimedia Appendix 4. Among the 120 top keywords, 107 (89.2%) were found to be health concepts in the data model. The concepts that were not mapped include general English words such as *hours*, *trouble*, and *difficult*.



**Figure 3.** The distribution of 40 keywords representing different topic themes (topic 1-12) for the microconcepts fatigue, prednisone, and patient visit. The color indicates different health concepts linked to the topic keywords.



**Fatigue**

Fatigue is a common clinical condition discussed by patients in the patient secure messages. Topic modeling on patient secure messages targeting fatigue showed that most of the patient discussions were centered around sleepiness, adverse drug effects, and relevant symptoms and conditions. More specifically, the keywords *sleep*, *beds*, *fall*, *asleep*, *little*, and

*wake ups* frequently appeared in these patient secure messages, which highlighted the strong association between fatigue and sleepiness. The topic words *pharmaceutical preparations*, *dose*, *adverse effects*, and *headache* revealed that fatigue might be an adverse effect related to some drugs. Other relevant symptoms and conditions such as *headache*, *coughing*, *fever*, and *pain* were also found in patient secure messages discussing fatigue. After tracing back to the original patient secure messages, we

also found mentions of *low*, *thyroid gland*, and *dietary iron*. This finding is consistent with studies reporting that fatigue is one of the most common signs of an underactive thyroid (hypothyroidism) [37] and iron deficiency [38].

### **Prednisone**

Prednisone is a corticosteroid drug that controls inflammation by suppressing the human immune system. Prednisone appeared frequently in the patient secure messages. The focus of patient secure messages related to prednisone was primarily on drug use, side effects, and disease treatment. The topic words related to prednisone use and dose such as *taper*, *dose*, *daily*, *dosage*, and *tapering* were mentioned most often. Prednisone has been used to treat a variety of medical conditions, as shown by the keywords such as *cough*, *dyspnea*, *asthma*, *pruritus*, and *urticaria* [39-42]. In addition, the patients reported side effects such as *fatigue*, *headache*, *sleep* and *pain* after taking prednisone [43].

### **Patient Visit**

Patient visit is a type (attribute) of patient *appointment* under the *foundation and base* concept. The topics of patient secure messages related to patient visits revealed various potential reasons for patient visits by making appointments. For example, many patients requested appointments to visit the Mayo Clinic by mentioning their health information and medical records (*mayo*, *information*, and *records*), laboratory tests and test results (*experimental result*, *laboratory procedure*, *blood*, *scan*, and *radionuclide imaging*), and medications (*pharmaceutical preparation*, *prescription procedure*, and *pharmaceutical services*). The keywords *pain*, *symptoms*, *headache*, *coughing*, *diarrhea*, *exanthema*, *heart*, and *cardiologists* suggest the purpose of their visit to the Mayo Clinic. In addition, some patients made appointments for financial issues indicated by the topic words *insurance* and *pay*.

## **Discussion**

### **Principal Findings**

Patient secure messages contain valuable information about the quality of health care delivery systems, drug efficacy and safety signals, and other pain points of patient health management. The data model and annotated corpus enable us to extract information from patient secure messages for systematic content analysis and to resolve the challenges of data analysis and integration by standardizing the unstructured data with the structured data system. In future, this data model can be used and extended to model patient narratives from social media platforms [44] to analyze their content. Our future aim is to leverage this data model and annotated corpus for developing machine learning models and NLP-enabled parsing tools for automatically parsing patient narratives to advance clinical research and practice [45,46].

During annotation, we faced some challenges because of the heterogeneous and superficial nature of the language used in the messages, which often deviated from formal English grammar, spelling, and punctuation rules. These challenges can often lower the quality of the data and make them less accessible to automated processing by a system. We discussed and

designed some solutions to overcome these challenges. The challenges and their solutions are discussed in detail in [Multimedia Appendix 2](#). During the comparison between the 2 sets of the first set of annotation, the level of agreement between the annotators was consistent for a few subconcepts, such as *appointment*, *diagnostic report*, *immunization*, *medication*, and *practitioner*. In contrast, it was very unsatisfactory for subconcepts such as *specimen*, *related-person*, *document-reference*, *eligibility*, and *health care-service*. Although the 2 annotators achieved a satisfactory IAA score later, annotation bias likely exists in the annotated corpus.

After analyzing the annotated corpus and identifying important health concepts in patient secure messages, we further used topic modeling to automatically uncover hidden topics of patient secure messages mentioning health concepts [47]. In this way, we were able to evaluate the data model and understand the focus of patient discussion and concerns in the patient secure messages. For example, prednisone, because it is a corticosteroid drug that controls inflammation, was a highly discussed topic in the patient secure messages. The patient-provider discussion on prednisone was primarily centered around medication use, side effects, and disease treatment [39-43]. These findings offer useful information for shared clinical decision making and patient-centered care. LDA exploits statistical inference to identify latent topics using a bag-of-words model and term frequency. Therefore, LDA mainly discovers topics with high frequency and dominant terms and pays little attention to rare, yet meaningful, topics from patient secure messages. In addition, hyperparameter tuning in LDA can be more art than science.

Our study is the first to develop a data model based on HL7 FHIR to understand and analyze the content of patient secure messages. This study has several limitations. For annotation, the F1 scores were 0.62 (macro mean) and 0.74 (micro mean). This indicates that the task of annotation is a difficult one because we need to assign 3-level hierarchical health concepts to an identified health entity, and there are 85 microconcepts to be selected in the data model. Our annotators not only need to assign a category to these potential entities but also to identify their boundaries. If the boundaries are defined differently by different annotators, it is considered *notconsistent*. The language used in patient secure messages to describe medical concepts is casual, colloquial, and ambiguous. We revised our annotation guideline 5 times and trained our annotators 4 times. We believe that the F1 score could be improved, given more resources for annotation.

All the messages analyzed in this study were sent by patients. We did not analyze the messages generated by the clinicians who read and replied to these patient secure messages. The Mayo Clinic at Rochester, Minnesota, is a large nonprofit academic medical center that provides comprehensive patient care in the United States. We acknowledge that the patients and medical practices at the Mayo Clinic are not necessarily a representative cross-section of all patients and medical practices in the country. It is also unknown how the data model can be applied to other hospitals in other countries, given different models of care and patient-clinician relationships. For example, we are aware that in China, there are no patient portals, and patients communicate with their providers through other means.

Thus, we share our model, annotation guideline, and other materials with the broader scientific community and welcome all sorts of collaboration or partnership.

Finally, we acknowledge that evaluation and validation are key and challenging in this case. Strictly speaking, the data model can only be validated for the claimed utilities in real-world implementations. The topic modeling analysis we conducted was only an initial and weak evaluation. We expect to conduct a much more rigorous evaluation and validation by building collaborations and partnerships with domestic and international researchers in the field and by moving forward with building the NLP systems.

In the next stage of our study, we aim to curate a larger annotated corpus and use it for training and testing machine learning models for automated triaging. The manually annotated corpus is reusable for future NLP research to save manual effort and cost, although there might be some challenges related to its reuse because of data privacy and confidentiality challenges. This corpus is also generated based on patient secure messages; therefore, there always will remain a question of its usability in different domains of medical research.

## Conclusions

A patient portal as a tethered EHR system enables patients to access their medical records, seek support, and share their opinions with their caregivers through secure messaging between their clinical visits. The large volume of secure messages opens new opportunities and challenges for understanding patient concerns and information integration into EHRs to improve patient-centered care. This study is a novel attempt to identify the content of patient secure messages based on the foundation and base, clinical, and financial concepts of HL7 FHIR standards. The data model and annotated corpus enable us to meet the challenges of analyzing and understanding unstructured health information from patient secure messages along with the topic modeling technique to discover the hidden topics on interesting health concepts in patient secure messages.

The data model could be potentially used for automatically identifying and analyzing other types of patient narratives such as those in various social media and patient forums. In the future, we plan to develop a machine learning and NLP solution to enable automatic triaging solutions to reduce the workload of clinicians and perform more granular content analysis to understand patients' needs and improve patient-centered care.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Data model for patient secure messages based on Health Level-7 Fast Healthcare Interoperability Resources.  
[\[XLSX File \(Microsoft Excel File\), 13 KB-Multimedia Appendix 1\]](#)

---

## Multimedia Appendix 2

Description of methodology in detail.  
[\[PDF File \(Adobe PDF File\), 285 KB-Multimedia Appendix 2\]](#)

---

## Multimedia Appendix 3

Annotation guideline.  
[\[PDF File \(Adobe PDF File\), 343 KB-Multimedia Appendix 3\]](#)

---

## Multimedia Appendix 4

Results of topic modeling.  
[\[XLSX File \(Microsoft Excel File\), 33 KB-Multimedia Appendix 4\]](#)

---

## References

1. Casalino LP. The medicare access and CHIP reauthorization act and the corporate transformation of American medicine. *Health Aff* 2017 May 1;36(5):865-869. [doi: [10.1377/hlthaff.2016.1536](https://doi.org/10.1377/hlthaff.2016.1536)] [Medline: [28461353](https://pubmed.ncbi.nlm.nih.gov/28461353/)]
2. Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. *N Engl J Med* 2010 Aug 5;363(6):501-504. [doi: [10.1056/NEJMp1006114](https://doi.org/10.1056/NEJMp1006114)] [Medline: [20647183](https://pubmed.ncbi.nlm.nih.gov/20647183/)]
3. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tullidge-Scheitel SM, et al. Impact of patient portal secure messages and electronic visits on adult primary care office visits. *Telemed J E Health* 2014 Mar;20(3):192-198 [[FREE Full text](#)] [doi: [10.1089/tmj.2013.0097](https://doi.org/10.1089/tmj.2013.0097)] [Medline: [24350803](https://pubmed.ncbi.nlm.nih.gov/24350803/)]
4. Plastiras P, O'Sullivan D, Weller P. An Ontology-driven Information Model for Interoperability of Personal and Electronic Health Records. ResearchGate. 2014 Apr. URL: [https://www.researchgate.net/publication/313512132\\_An\\_Ontology-Driven\\_Information\\_Model\\_for\\_Interoperability\\_of\\_Personal\\_and\\_Electronic\\_Health\\_Records](https://www.researchgate.net/publication/313512132_An_Ontology-Driven_Information_Model_for_Interoperability_of_Personal_and_Electronic_Health_Records) [accessed 2019-01-20]

5. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. *J Med Internet Res* 2015 Dec 3;17(12):e275 [FREE Full text] [doi: [10.2196/jmir.4847](https://doi.org/10.2196/jmir.4847)] [Medline: [26681155](https://pubmed.ncbi.nlm.nih.gov/26681155/)]
6. Medicare and Medicaid Programs; Electronic Health Record Incentive Program-stage 2. Federal Register. 2012. URL: <https://www.federalregister.gov/documents/2012/09/04/2012-21050/medicare-and-medicaid-programs-electronic-health-record-incentive-program-stage-2> [accessed 2020-02-13]
7. Henry J, Barker W, Kachay L. Electronic Capabilities for Patient Engagement among U.S. Non-Federal Acute Care Hospitals: 2013-2017. Office of the National Coordinator for Health Information. 2019 Apr. URL: <https://www.healthit.gov/sites/default/files/page/2019-04/AHApatientengagement.pdf> [accessed 2020-02-13]
8. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. *J Am Med Inform Assoc* 2013 May 1;20(3):519-525 [FREE Full text] [doi: [10.1136/amiajnl-2012-001253](https://doi.org/10.1136/amiajnl-2012-001253)] [Medline: [23242764](https://pubmed.ncbi.nlm.nih.gov/23242764/)]
9. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. *J Med Internet Res* 2013 Jul 3;15(7):e133 [FREE Full text] [doi: [10.2196/jmir.2589](https://doi.org/10.2196/jmir.2589)] [Medline: [23823974](https://pubmed.ncbi.nlm.nih.gov/23823974/)]
10. Horvath M, Levy J, L'Engle P, Carlson B, Ahmad A, Ferranti J. Impact of health portal enrollment with email reminders on adherence to clinic appointments: a pilot study. *J Med Internet Res* 2011 May 26;13(2):e41 [FREE Full text] [doi: [10.2196/jmir.1702](https://doi.org/10.2196/jmir.1702)] [Medline: [21616784](https://pubmed.ncbi.nlm.nih.gov/21616784/)]
11. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. *Health Serv Res* 2012 Apr;47(2):614-632 [FREE Full text] [doi: [10.1111/j.1475-6773.2011.01342.x](https://doi.org/10.1111/j.1475-6773.2011.01342.x)] [Medline: [22091980](https://pubmed.ncbi.nlm.nih.gov/22091980/)]
12. Crotty BH, Tamrat Y, Mostaghimi A, Safran C, Landon BE. Patient-to-physician messaging: volume nearly tripled as more patients joined system, but per capita rate plateaued. *Health Aff* 2014 Oct;33(10):1817-1822 [FREE Full text] [doi: [10.1377/hlthaff.2013.1145](https://doi.org/10.1377/hlthaff.2013.1145)] [Medline: [25288428](https://pubmed.ncbi.nlm.nih.gov/25288428/)]
13. Cronin R, Davis S, Shenson J, Chen Q, Rosenbloom S, Jackson G. Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. *Appl Clin Inform* 2015;6(2):288-304 [FREE Full text] [doi: [10.4338/ACI-2014-12-RA-0117](https://doi.org/10.4338/ACI-2014-12-RA-0117)] [Medline: [26171076](https://pubmed.ncbi.nlm.nih.gov/26171076/)]
14. Shenson JA, Cronin RM, Davis SE, Chen Q, Jackson GP. Rapid growth in surgeons' use of secure messaging in a patient portal. *Surg Endosc* 2016 Apr;30(4):1432-1440 [FREE Full text] [doi: [10.1007/s00464-015-4347-y](https://doi.org/10.1007/s00464-015-4347-y)] [Medline: [26123340](https://pubmed.ncbi.nlm.nih.gov/26123340/)]
15. Masterman M, Cronin R, Davis S, Shenson J, Jackson G. Adoption of Secure Messaging in a Patient Portal across Pediatric Specialties. *AMIA Annu Symp Proc* 2016;2016:1930-1939 [FREE Full text] [Medline: [28269952](https://pubmed.ncbi.nlm.nih.gov/28269952/)]
16. HL7 FHIR Release 4. Health Level Seven International. 2019. URL: <https://www.hl7.org/fhir/> [accessed 2019-11-04]
17. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tullidge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk? *J Am Med Inform Assoc* 2013;20(6):1143-1149 [FREE Full text] [doi: [10.1136/amiajnl-2012-001208](https://doi.org/10.1136/amiajnl-2012-001208)] [Medline: [23703826](https://pubmed.ncbi.nlm.nih.gov/23703826/)]
18. Sulieman L, Robinson JR, Jackson GP. Automating the classification of complexity of medical decision-making in patient-provider messaging in a patient portal. *J Surg Res* 2020 Nov;255:224-232 [FREE Full text] [doi: [10.1016/j.jss.2020.05.039](https://doi.org/10.1016/j.jss.2020.05.039)] [Medline: [32570124](https://pubmed.ncbi.nlm.nih.gov/32570124/)]
19. Cronin RM, Fabbri D, Denny JC, Rosenbloom ST, Jackson GP. A comparison of rule-based and machine learning approaches for classifying patient portal messages. *Int J Med Inform* 2017 Sep;105:110-120 [FREE Full text] [doi: [10.1016/j.ijmedinf.2017.06.004](https://doi.org/10.1016/j.ijmedinf.2017.06.004)] [Medline: [28750904](https://pubmed.ncbi.nlm.nih.gov/28750904/)]
20. About HL7. Health Level Seven International. 2019. URL: <https://www.hl7.org/about/> [accessed 2020-01-20]
21. What Is FHIR? Office of the National Coordinator for Health Information Technology. 2019. URL: <https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf> [accessed 2020-01-20]
22. Saripalle R. Fast health interoperability resources (FHIR): current status in the healthcare system. *Int J E-Health Med Commun* 2019 Jan;10(1):76-93. [doi: [10.4018/IJEHMC.2019010105](https://doi.org/10.4018/IJEHMC.2019010105)]
23. Hong N, Wen A, Shen F, Sohn S, Liu S, Liu H, et al. Integrating structured and unstructured EHR data using an FHIR-based type system: a case study with medication data. *AMIA Jt Summits Transl Sci Proc* 2018;2017:74-83 [FREE Full text] [Medline: [29888045](https://pubmed.ncbi.nlm.nih.gov/29888045/)]
24. Patient Online Services. Mayo Clinic. 2010. URL: <https://onlineservices.mayoclinic.org/content/staticpatient/showpage/patientonline> [accessed 2018-10-02]
25. Stubbs A. MAE and MAI: Lightweight Annotation and Adjudication Tools. In: Proceedings of the 5th Linguistic Annotation Workshop. 2011 Jun Presented at: LAW V'11; June 23-24, 2011; Portland, Oregon, USA p. 129-133.
26. Cunningham H, Maynard D, Bontcheva K, Tablan V. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002 Jul 6 Presented at: ACL'02; July 6-12, 2002; Philadelphia, PA, USA p. 168-175.
27. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc* 2010;17(3):229-236 [FREE Full text] [doi: [10.1136/jamia.2009.002733](https://doi.org/10.1136/jamia.2009.002733)] [Medline: [20442139](https://pubmed.ncbi.nlm.nih.gov/20442139/)]
28. Blei D, Ng A, Jordan M. Latent dirichlet allocation. *J Mach Learn Res* 2003 Mar 1;3:993-1022.



29. McCallum AK. Mallet: A machine learning for language toolkit. Mallet. 2002. URL: <http://mallet.cs.umass.edu> [accessed 2019-05-15]
30. Röder M, Both A, Hinneburg A. Exploring the Space of Topic Coherence Measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 Feb Presented at: WSDM'15; January 31-February 6, 2015; Shanghai, China p. 399-408. [doi: [10.1145/2684822.2685324](https://doi.org/10.1145/2684822.2685324)]
31. Irizarry T, Dabbs AD, 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](https://doi.org/10.2196/jmir.4255)] [Medline: [26104044](https://pubmed.ncbi.nlm.nih.gov/26104044/)]
32. Gurulingappa H, Rajput AM, Roberts A, Fluck J, Hofmann-Apitius M, Toldo L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *J Biomed Inform* 2012 Oct;45(5):885-892 [FREE Full text] [doi: [10.1016/j.jbi.2012.04.008](https://doi.org/10.1016/j.jbi.2012.04.008)] [Medline: [22554702](https://pubmed.ncbi.nlm.nih.gov/22554702/)]
33. Blackman NJ, Koval JJ. Interval estimation for Cohen's kappa as a measure of agreement. *Statist Med* 2000 Mar 15;19(5):723-741. [doi: [10.1002/\(sici\)1097-0258\(20000315\)19:5<723::aid-sim379>3.0.co;2-a](https://doi.org/10.1002/(sici)1097-0258(20000315)19:5<723::aid-sim379>3.0.co;2-a)] [Medline: [10700742](https://pubmed.ncbi.nlm.nih.gov/10700742/)]
34. How FHIR Fits into an EHR. Health Level Seven International. 2019. URL: <https://www.hl7.org/fhir/ehr-fm.html> [accessed 2019-12-10]
35. Huang M, ElTayeby O, Zolnoori M, Yao L. Public opinions toward diseases: infodemiological study on news media data. *J Med Internet Res* 2018 May 8;20(5):e10047 [FREE Full text] [doi: [10.2196/10047](https://doi.org/10.2196/10047)] [Medline: [29739741](https://pubmed.ncbi.nlm.nih.gov/29739741/)]
36. Huang M, Zolnoori M, Balls-Berry JE, Brockman TA, Patten CA, Yao L. Technological innovations in disease management: text mining US patent data from 1995 to 2017. *J Med Internet Res* 2019 Apr 30;21(4):e13316 [FREE Full text] [doi: [10.2196/13316](https://doi.org/10.2196/13316)] [Medline: [31038462](https://pubmed.ncbi.nlm.nih.gov/31038462/)]
37. Chaker L, Bianco AC, Jonklaas J, Peeters RP. Hypothyroidism. *Lancet* 2017 Sep 23;390(10101):1550-1562 [FREE Full text] [doi: [10.1016/S0140-6736\(17\)30703-1](https://doi.org/10.1016/S0140-6736(17)30703-1)] [Medline: [28336049](https://pubmed.ncbi.nlm.nih.gov/28336049/)]
38. Camaschella C. Iron-deficiency anemia. *N Engl J Med* 2015 May 7;372(19):1832-1843. [doi: [10.1056/NEJMra1401038](https://doi.org/10.1056/NEJMra1401038)] [Medline: [25946282](https://pubmed.ncbi.nlm.nih.gov/25946282/)]
39. Thompson WH, Nielson CP, Carvalho P, Charan NB, Crowley JJ. Controlled trial of oral prednisone in outpatients with acute COPD exacerbation. *Am J Respir Crit Care Med* 1996 Aug;154(2):407-412. [doi: [10.1164/ajrccm.154.2.8756814](https://doi.org/10.1164/ajrccm.154.2.8756814)] [Medline: [8756814](https://pubmed.ncbi.nlm.nih.gov/8756814/)]
40. Schuh S, Reisman J, Alshehri M, Dupuis A, Corey M, Arseneault R, et al. A comparison of inhaled fluticasone and oral prednisone for children with severe acute asthma. *N Engl J Med* 2000 Sep 7;343(10):689-694. [doi: [10.1056/NEJM200009073431003](https://doi.org/10.1056/NEJM200009073431003)] [Medline: [10974132](https://pubmed.ncbi.nlm.nih.gov/10974132/)]
41. Taranta A, Mark H, Haas RC, Cooper NS. Nodular panniculitis after massive prednisone therapy. *Am J Med* 1958 Jul;25(1):52-61. [doi: [10.1016/0002-9343\(58\)90198-0](https://doi.org/10.1016/0002-9343(58)90198-0)]
42. Pollack CV, Romano TJ. Outpatient management of acute urticaria: the role of prednisone. *Ann Emerg Med* 1995 Nov;26(5):547-551. [doi: [10.1016/s0196-0644\(95\)70002-1](https://doi.org/10.1016/s0196-0644(95)70002-1)]
43. Prednisone. Mayo Clinic. 2020. URL: <https://www.mayoclinic.org/drugs-supplements/prednisone-oral-route/side-effects/drg-20075269?p=1> [accessed 2019-08-10]
44. Ru B, Harris K, Yao L. A Content Analysis of Patient-Reported Medication Outcomes on Social Media. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW). 2015 Nov Presented at: ICDMW'15; November 14-17, 2015; Atlantic City, NJ, USA. [doi: [10.1109/icdmw.2015.150](https://doi.org/10.1109/icdmw.2015.150)]
45. Soni S, Gudala M, Wang D, Roberts K. Using FHIR to construct a corpus of clinical questions annotated with logical forms and answers. *AMIA Annu Symp Proc* 2019;2019:1207-1215 [FREE Full text] [Medline: [32308918](https://pubmed.ncbi.nlm.nih.gov/32308918/)]
46. Zolnoori M, Fung KW, Patrick TB, Fontelo P, Kharrazi H, Faiola A, et al. A systematic approach for developing a corpus of patient reported adverse drug events: a case study for SSRI and SNRI medications. *J Biomed Inform* 2019 Feb;90:103091 [FREE Full text] [doi: [10.1016/j.jbi.2018.12.005](https://doi.org/10.1016/j.jbi.2018.12.005)] [Medline: [30611893](https://pubmed.ncbi.nlm.nih.gov/30611893/)]
47. Zolnoori M, Huang M, Patten CA, Balls-Berry JE, Goudarzvand S, Brockman TA, et al. Mining news media for understanding public health concerns. *J Clin Transl Sci* 2019 Oct 23;5(1):e1 [FREE Full text] [doi: [10.1017/cts.2019.434](https://doi.org/10.1017/cts.2019.434)] [Medline: [33948233](https://pubmed.ncbi.nlm.nih.gov/33948233/)]

## Abbreviations

- EHR:** electronic health record
- FHIR:** Fast Healthcare Interoperability Resources
- HL7:** Health Level-7
- IAA:** interannotator agreement
- LDA:** latent Dirichlet allocation
- NLP:** natural language processing



*Edited by R Kukafka; submitted 06.01.21; peer-reviewed by X Yang, M Rigby; comments to author 16.03.21; revised version received 10.05.21; accepted 24.05.21; published 30.07.21*

*Please cite as:*

*De A, Huang M, Feng T, Yue X, Yao L*

*Analyzing Patient Secure Messages Using a Fast Health Care Interoperability Resources (FIHR)-Based Data Model: Development and Topic Modeling Study*

*J Med Internet Res 2021;23(7):e26770*

*URL: <https://www.jmir.org/2021/7/e26770>*

*doi: [10.2196/26770](https://doi.org/10.2196/26770)*

*PMID: [34328444](https://pubmed.ncbi.nlm.nih.gov/34328444/)*

©Amrita De, Ming Huang, Tinghao Feng, Xiaomeng Yue, Lixia Yao. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.