

Using Temporal Features to Provide Data-Driven Clinical Early Warnings for Chronic Obstructive Pulmonary Disease and Asthma Care Management: Protocol for a Secondary Analysis

Gang Luo¹, PhD; Bryan L Stone², MD, MS; Corinna Koebnick³, PhD; Shan He⁴, PhD; David H Au^{5,6}, MD, MS; Xiaoming Sheng⁷, PhD; Maureen A Murtaugh⁸, PhD, RDN; Katherine A Sward⁷, PhD, RN; Michael Schatz^{3,9}, MD, MS; Robert S Zeiger^{3,9}, MD, PhD; Giana H Davidson¹⁰, MD, MPH; Flory L Nkoy², MD, MS, MPH

¹Department of Biomedical Informatics and Medical Education, University of Washington, UW Medicine South Lake Union, 850 Republican Street, Building C, Box 358047, Seattle, WA 98195, USA

²Department of Pediatrics, University of Utah, 100 N Mario Capecchi Drive, Salt Lake City, UT 84113, USA

³Department of Research & Evaluation, Kaiser Permanente Southern California, 100 S. Los Robles, 2nd Floor, Pasadena, CA 91101, USA

⁴Care Transformation, Intermountain Healthcare, World Trade Center, 16th floor, 60 East South Temple Street, Salt Lake City, UT 84111, USA

⁵Center of Innovation for Veteran-Centered & Value-Driven Care, VA Puget Sound Health Care System, 1660 S. Columbian Way, Seattle, WA 98108, USA

⁶Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of Washington, 1959 N.E. Pacific, Seattle, WA 98195, USA

⁷College of Nursing, University of Utah, 10 South 2000 East, Salt Lake City, UT 84112, USA

⁸Department of Family and Preventive Medicine, University of Utah, 375 Chipeta Way, Suite A, Salt Lake City, UT 84108, USA

⁹Department of Allergy, Kaiser Permanente Southern California, 7060 Clairemont Mesa Blvd, San Diego, CA 92111, USA

¹⁰Department of Surgery, University of Washington, 1959 N.E. Pacific, Box 356410, Seattle, WA 98195, USA

luogang@uw.edu, bryan.stone@hsc.utah.edu, corinna.koebnick@kp.org, shan.he@imail.org, dau@uw.edu, xiaoming.sheng@utah.edu, maureen.murtaugh@hsc.utah.edu, kathy.sward@nurs.utah.edu, michael.x.schatz@kp.org, robert.s.zeiger@kp.org, ghd@uw.edu, flory.nkoy@hsc.utah.edu

Corresponding author:

Gang Luo, PhD

Department of Biomedical Informatics and Medical Education, University of Washington, UW Medicine South Lake Union, 850 Republican Street, Building C, Box 358047, Seattle, WA 98195, USA

Phone: 1-206-221-4596

Fax: 1-206-221-2671

Email: luogang@uw.edu

Abstract

Background: Both chronic obstructive pulmonary disease (COPD) and asthma incur heavy healthcare burdens. To support tailored preventive care for these two diseases, predictive modeling is widely used to give warnings and to identify patients for care management. However, three gaps exist in current modeling methods owing to rarely factoring in temporal aspects showing trends and early health change:

- 1) Existing models seldom use temporal features and often give late warnings, making care reactive. A health risk is often found at a relatively late stage of declining health, when the risk of a poor outcome is high and resolving the issue is difficult and costly. A typical model predicts patient outcomes in the next, say, 12 months. This often does not warn early enough. If a patient will actually be hospitalized for COPD next week, intervening now could be too late to avoid the hospitalization. If temporal features were used, this patient could potentially be identified a few weeks earlier to institute preventive therapy.
- 2) Existing models often miss many temporal features with high predictive power and have low accuracy. This makes care management enroll many patients not needing it and overlook over half of the patients needing it the most.
- 3) Existing models often give no information on why a patient is at high risk, nor about possible interventions to mitigate risk, causing busy care managers to spend more time reviewing charts and to miss suited interventions. Typical automatic explanation methods cannot handle longitudinal attributes and fully address these issues.

Objective: To fill these gaps so that more COPD and asthma patients will receive more appropriate and timely care, we will develop comprehensible data-driven methods to provide accurate early warnings of poor outcomes and to suggest tailored interventions, making care more proactive, efficient, and effective.

Methods: By conducting secondary data analysis and surveys, the study will: a) use temporal features to provide accurate early warnings of poor outcomes, and assess the potential impact on prediction accuracy, risk warning timeliness, and outcomes; b)

automatically identify actionable temporal risk factors for each patient at high risk for future hospital use and assess the impact on prediction accuracy and outcomes; c) assess actionable information's impact on clinicians' acceptance of early warnings and on perceived care plan quality.

Results: We are obtaining clinical and administrative data sets from three leading healthcare systems' enterprise data warehouses. We plan to start data analysis in 2020 and finish our study in 2025.

Conclusions: Techniques to be developed in this study can boost risk warning timeliness, model accuracy, and generalizability; improve patient finding for preventive care; help form tailored care plans; advance machine learning for many clinical applications; and generalize to many other chronic diseases.

Keywords: decision support techniques; forecasting; machine learning; patient care management

1. Introduction

The current predictive modeling method for implementing chronic obstructive pulmonary disease (COPD) and asthma care management has three major gaps

COPD and asthma are chronic respiratory diseases incurring heavy healthcare burdens on society, healthcare systems, and payers. In the U.S., COPD affects over 6.5% of adults, is the third leading cause of death, and incurs 700,000 inpatient stays, 1.5 million emergency room visits, and U.S. \$32 billion in cost annually [1]. Asthma affects 8% of adults [2] and 9.6% of children [3, 4], and incurs 3,630 deaths, 493,000 inpatient stays, 1.8 million emergency room visits, and U.S. \$56 billion in cost annually [2, 5]. As a service intended to prevent poor outcomes like hospitalization, care management is widely adopted to provide tailored preventive care for COPD and asthma patients. Purchased by most large employers and offered by almost all private health plans [6-8], care management is a collaborative process to evaluate, plan, implement, coordinate, and monitor options and services in order to accommodate an individual's health and service needs [9]. In care management, a care manager calls a patient regularly, helps arrange for medical appointments, and schedules health and related services. Appropriate use of care management can cut hospital use (emergency room visits and inpatient stays) by up to 40% [7, 10-15], lower cost by up to 15% [11-16], and enhance patient adherence, quality of life, and satisfaction by 30-60% [10].

Predictive models are widely used, e.g., by health plans in 9 of 12 regions [17], as the best method [18] to warn of poor outcomes and to identify COPD and asthma patients for care management [6-8]. Multiple models have been built for predicting the health outcomes of individual COPD and asthma patients [19-28]. But, current modeling methods have three major gaps restricting their effectiveness due to inadequate use of temporal features showing trends and early health change. A temporal feature, like the slope of pulmonary function across the last year, is an independent variable formed by transforming longitudinal attribute(s).

Gap 1: Existing models often give late warnings, making care reactive and suboptimal

Existing models for predicting the health outcomes of individual COPD and asthma patients seldom use temporal features [19-28] and often give late warnings, making care reactive and missing opportunities for clinical and therapy teams to intervene early to reduce the risk of poor outcomes. A health risk is often identified at a relatively late stage of declining health, when the chance of a poor outcome is high and resolving the issue is difficult and costly. A typical model predicts patient outcomes in the next, say, 12 months. For patients with imminent poor outcomes, this does not warn early enough. If a patient will actually be hospitalized for COPD next week, intervening now could be too late to avoid the hospitalization. If temporal features were used, this patient could be identified a few weeks or months earlier, when health decline is still at an early stage, resolving the issue is easier, and preventing hospitalization is likely.

Gap 2: Typical models have low accuracy, leading to suboptimal and wasted care

Models for predicting a patient's health outcome and cost typically have low accuracy. When projecting the health outcome of a patient, the accuracy measure of area under the receiver operating characteristic curve (AUC) is typically much lower than 0.8 [19-28]. When projecting the healthcare cost of a patient, the accuracy measure of R^2 is typically lower than 25% [29, 30] and the mean error is as big as the mean cost [31]. These large errors in prediction results create difficulty in properly aligning care management's use with the patients needing it the most [10].

Care management can require over U.S. \$5,000 per person per year [11] and usually enrolls only 1-3% of patients because of resource limits [32]. For patients predicted to have the worst outcomes or the largest costs [10, 33], care managers review patient charts and manually make allocation decisions.

A small percentage of patients use most of the healthcare resources and costs. The upper 20% of patients use 80% of the resources and costs. The upper 1% use 25% [19, 32, 34]. Accurately identifying patients at high risk for poor outcomes or large healthcare costs is critical for effective targeted application of care management resources. Yet, Weir *et al.* [33] showed that in the upper 10% of patients who actually spent the largest healthcare costs, over 60% of them were not included in the upper 10%

risk group identified by a predictive model. In the upper 1% of patients who actually spent the largest healthcare costs, around 50% and over 80% of them were not included in the identified upper 10% and 1% risk groups, respectively. Assume the care management program could take 1% of all patients. In this case, even if the care managers could afford to examine the upper 10% risk group found by the predictive model and manually made correct decisions for enrollment, the care managers would still not find around half of the upper 1% of patients who spent the largest healthcare costs. For COPD and asthma, if we could identify 10% more of the upper 1% of patients who spent the largest healthcare costs and enroll them in care management, we could boost outcomes and spare possibly up to U.S. \$120 million in COPD care [1] and U.S. \$210 million in asthma care each year [19-21]. In general, due to the large patient population, a small boost in accuracy will benefit numerous patients and have a big positive impact.

Current models for predicting the health outcomes and healthcare costs of individual COPD and asthma patients have low prediction accuracy for several reasons:

- 1) Many temporal features with high predictive power are either frequently unused in an existing model or yet to be found. Google recently applied long short-term memory (LSTM) [35], one kind of deep neural network, to all of the attributes in the electronic health record to automatically learn temporal features from longitudinal data [36]. For forecasting each of the three outcomes: long hospital stay, unanticipated readmissions within 30 days, and in-hospital mortality, this increased the AUC by ~+10% [36]. Multiple other studies [37-39] showed similar results for a variety of clinical prediction tasks. This aligns with what has taken place in areas like video classification, natural language processing, and speech recognition, where temporal features that LSTM automatically learned from data outperform those mined from data by other methods or specified by experts [40, 41].
- 2) Although >40 risk factors for undesirable outcomes in COPD/asthma have been identified [19, 20, 23, 28, 42-48], an existing model usually uses only a few (e.g., <10) [19-24, 26-28]. Existing models were often constructed based on data obtained from clinical trials or old-fashioned electronic health records collecting limited variables [49]. No published model adopts all of the known risk factors available in modern electronic health records collecting extensive variables [49].
- 3) Environmental information like air quality and weather variables are known to impact COPD and asthma outcomes [43, 50-52], but with rare exceptions [25], are infrequently used in existing models.

Gap 3: Typical models give no information on why a patient is at high risk for a poor outcome and on possible interventions to mitigate risk, often causing busy care managers to spend more time and to miss suited interventions

Before enrolling a patient, care managers need to know why the patient is at high risk for a poor outcome and about possible interventions to mitigate risk. Complex predictive models, which include most machine learning models like LSTM, give no explanatory or prescriptive information. Frequently, a patient's records have many variables on hundreds of pages accumulated over a long period of time [53]. Unlike physicians who see patients from time to time, care managers often have not previously seen the records when needing to make enrollment decisions. When the model offers no explanation, busy care managers often spend extra time reviewing the records to find the reasons. This is time-consuming and difficult.

A care manager may use subjective, variable judgment to form a care plan, but may miss some suited interventions because of two factors:

- 1) Several reasons can make a patient at high risk for a poor outcome. Each reason is shown by a feature combination as a risk pattern. For instance, the ratio of inhaled steroid to beta agonist dispensing to the patient decreased over 12 months AND sulfur dioxide level was ≥ 3 parts per million for ≥ 5 days in the past week. Many features exist. Like any human, an ordinary care manager can deal with ≤ 9 information items simultaneously [54], making it difficult to identify all reasons from numerous possible feature combinations.
- 2) Huge variation in practice, often by 1.6-5.6 times, appears across care managers, facilities, and regions [34, 55-58]. Missing suited interventions can degrade outcomes. Typical automatic explanation methods [59, 60] do not handle longitudinal attributes and cannot fully address these issues.

To fill the gaps, we will provide accurate early warnings of poor outcomes and suggest tailored interventions in a data-driven and comprehensible way

To fill the gaps for more COPD and asthma patients to receive appropriate and timely care, we will 1) use temporal features to provide accurate early warnings of poor outcomes, and assess potential impact on prediction accuracy, risk warning timeliness, and outcomes; 2) automatically identify actionable temporal risk factors for each patient at high risk for future hospital use and assess the impact on prediction accuracy and outcomes; 3) assess actionable information's impact on clinicians' acceptance of early warnings and on perceived care plan quality. Here, actionable information refers to the explanations and their linked interventions provided by our automated approach.

Innovation

This study will lead to several innovations. We will develop new, general informatics techniques. We will transform care management for COPD and asthma by directing it to the patients needing it in a more timely fashion and more precisely than current methods:

- a) We will build models to predict a patient's hospital use earlier and more accurately than current models, which often give late warnings and have low accuracy.
- b) We will be the first to semi-automatically extract predictive and clinically meaningful temporal features from longitudinal medical data. This process helps us address data quality issues, and automatically find and drop uninformative variables. All of these boost model accuracy and generalizability, and reduce the effort needed to build models usable in clinical practice. Currently, to build such models, clinicians typically need to manually identify such features, which is difficult and time-consuming.
- c) We will be the first to provide rule-based automatic explanations of machine learning prediction results directly on longitudinal data. Explanations are critical for care managers to understand the results to make appropriate care management enrollment and intervention decisions. Compared with other forms of automatic explanations like that used in Rajkomar *et al.* [36], rule-based explanations are easier to understand and can more directly suggest actionable interventions. Most automatic explanation methods [60], including our prior one [59], for machine learning prediction results cannot handle longitudinal attributes. Also, our prior method [59] gives explanations for a limited portion of patients. We will improve our prior method, handle longitudinal attributes, and expand automatic explanations' coverage of patients.
- d) We will be the first to automatically identify actionable temporal risk factors and suggest interventions based on inclusion of objective data. Currently, care managers use subjective, variable judgment to manually form care plans. Some suited interventions for patients at high risk for a poor outcome get missed. Also, care managers provide finite input on the patient to the other clinical care team members. With automatic explanations and suggested interventions in hand, care managers can pass this tailored information to the other clinical care team members so they can act accordingly. This could transform the care management process and make it more effective via closer collaboration between care managers and the other clinical care team members.
- e) Current models for predicting the health outcomes of individual COPD and asthma patients were built mostly using a small number of patients (e.g., <1,000) or variables (e.g., <10) [19-28], making it difficult to identify many predictive features and the interactions among them. Air quality and weather variables impact COPD and asthma outcomes [43, 50-52], but are rarely used in existing models. The predictive power of many known risk factors for undesirable outcomes is unused. Also, many predictive features have not yet been found. In contrast, we will use many patients and variables, enabling us to identify more predictive features and the interactions among them. The variables will include air quality, weather, and patient variables, cover many known risk factors for undesirable outcomes, and be used to find new predictive features in a data-driven way. Many features are new, capturing trends that existing models rarely touch.

To build and validate models for predicting the health outcomes of individual COPD and asthma patients, we will use data from four different electronic health record systems HELP, HELP2, Cerner, and Epic. This boosts model generalizability. In contrast, every existing model for predicting the health outcomes of individual COPD and asthma patients was built using data from only one electronic health record system [19-28].

In short, this study is significant as it will produce new techniques to advance machine learning for clinical applications and potentially transform preventive care for more patients to receive appropriate and timely care. Wide use of these techniques could boost outcomes and save resources.

2. Methods

Computing environment

All experiments will be done on a secure computer cluster at University of Washington Medicine (UWM) that is encrypted and password-protected. With proper authorization, all of the UWM care manager and physician test participants and research team members can log into this computer cluster from their UWM computers. We will install Oracle database, R, Weka [61], and TensorFlow [62] to be used in the study on the computer cluster. Weka is a major open-source machine learning toolkit. It incorporates many popular machine learning algorithms including both base and ensemble algorithms, feature selection techniques, and methods for dealing with imbalanced classes [63]. TensorFlow is Google's open-source deep neural network package.

Data sets

We will employ clinical and administrative data from the enterprise data warehouses (EDWs) of three leading healthcare systems Intermountain Healthcare (IH), Kaiser Permanente Southern California (KPSC), and UWM, as well as publicly

available air quality and weather data. All of the data to be used are structured. We will use all patients' data, which are needed for computing healthcare system features [64, 65], rather than only COPD and asthma patients' data. As the largest healthcare system in Utah, IH has 185 clinics and 22 hospitals. The EDW of IH contains numerous variables [66]. In this study, we will start with using the following of these variables: "admission date and time; age; orders (medications, labs, exams, immunizations, imaging, counseling, etc.), including order name, ordering provider, performing date, and result date; allergies; barriers (hearing, language, learning disability, mental status, religion, vision, etc.); cause of death; chief complaint; death date; diagnoses; discharge date; exam result; facility seen for the patient visit; gender; health insurance; healthcare cost; height; home address; immunizations; lab test result; language(s) spoken; medication refills; primary care physician as listed in the electronic medical record; problem list; procedure date; procedures; provider involved in the visit; race/ethnicity; referrals; religion; visit type (inpatient, outpatient, urgent care, or emergency department); vital signs; weight" [65]. An IH data analyst will download a de-identified IH data set, encrypt it, and transfer it to the secure computer cluster. The IH data set has information on clinical encounters in the previous 14 years (2005-2018). For the previous five years, the IH data for adults cover over 5,786,414 clinical encounters and 878,448 adult patients (age \geq 18) per year. The IH data for children cover over 1,557,713 clinical encounters and 360,698 pediatric patients (age $<$ 18) per year. COPD prevalence is \sim 4.1% in the IH adult population. Asthma prevalence is \sim 8.6% and \sim 7.6% in the IH adult and pediatric population, respectively. The IH data set provides the electronic record of care for \sim 60% of adults and \sim 95% of children in Utah [56, 67]. IH devotes many resources to maintain data integrity and accuracy. Because of its huge size and variable richness, the data set provides many advantages for us to explore the proposed predictive models.

KPSC and UWM have similar strengths. KPSC is the largest integrated healthcare system in Southern California, providing care to \sim 16% of residents in 227 clinics and 15 hospitals [68]. A KPSC data analyst will download a de-identified KPSC data set, encrypt it, and transfer it to the secure computer cluster. The KPSC data set has information on clinical encounters in the previous 10 years (2009-2018). For the previous five years, the KPSC data for adults cover over 9,448,987 clinical encounters and 2,890,027 adult patients per year. The KPSC data for children cover more than 1,380,900 clinical encounters and 975,249 pediatric patients per year. COPD prevalence is \sim 4.1% in the KPSC adult population. Asthma prevalence is \sim 10.8% and \sim 10.9% in the KPSC adult and pediatric population, respectively.

As the largest academic healthcare system in Washington, UWM has 12 clinics and 4 hospitals. A UWM data analyst will download a de-identified UWM data set, encrypt it, and transfer it to the secure computer cluster. The UWM data set has information on adult patient encounters in the previous 7 years (2012-2018). The UWM data cover over 1,714,196 clinical encounters and 277,289 adult patients per year. COPD prevalence among patients is \sim 4.1%. Asthma prevalence is \sim 9%.

In addition to the clinical and administrative data, we will use 11 air quality and weather variables, which were recorded over the previous 14 years (2005-2018) by the monitoring stations in the regions served by UWM, IH, and KPSC and are available from federal data sources [69, 70]. These variables include ozone, sulfur dioxide, particulate matter up to 10 μ m in size and 2.5 μ m in size, nitrogen dioxide, temperature, carbon monoxide, wind speed, relative humidity, precipitation, and dew point.

In the following, we sketch our techniques. Our design paper [71] describes the ideas in more detail. In this study, for each technique, we will flesh out its technical details, do computer coding, tune its parameters, and test it. The discussion below focuses on COPD. Whenever we mention COPD, the same applies to asthma.

Aim 1: Use temporal features to provide accurate early warnings of poor outcomes and assess impact on prediction accuracy.

We will semi-automatically extract predictive and clinically meaningful temporal features from patient, air quality, and weather data, and build models to predict a patient's health outcome. Each feature involves one or more raw variables. The number of possible features is almost infinite. Also, factors like environmental variables beyond air quality and weather can influence patient outcomes. This study does not intend to exhaust all of the possible features and factors that can influence patient outcomes, and achieve the highest possible prediction accuracy in theory. Rather, our purpose is to show that using temporal features can improve risk warning timeliness, prediction accuracy, and care management. A non-trivial boost in health outcomes can greatly benefit society. As is adequate for our target decision support application and typical with predictive modeling, our study focuses on associations.

Data pre-processing

We will write Oracle database SQL queries and R and Java programs for data pre-processing. Our source code will be made freely available on a project Web site hosted by UWM. In our future publications on this study's results, we will describe all of the decisions made for data pre-processing, such as the thresholds used for determining the physiologically impossible and invalid values of an attribute. We will transform all of the data sets into the Observational Medical Outcomes Partnership (OMOP) common data model format [72] and its related standardized terminologies [73]. We will extend the data model to include patient, air quality, and weather variables that are in our data sets, but not covered by the original data model. We will

adopt conventional techniques like imputation to manage missing values and to find, rectify, or drop invalid values [74, 75]. To avoid using too many longitudinal attributes, we will employ grouper models like the Diagnostic Cost Group system to merge diseases, drugs, and procedures [31, 34]. We will use the method given in our paper [71] to select the most relevant lab tests.

Variables: We will use patient, air quality, and weather variables. The patient variables include standard variables like diagnoses that the clinical predictive modeling literature [34, 55, 74] has studied, and many known risk factors for undesirable COPD outcomes listed in Bahadori *et al.* [45]. For air quality and weather variables, we will do spatial interpolation [76] to obtain their daily average values at the patient’s home address from those at regional monitoring stations [77].

COPD and asthma cases and outcomes

As test cases, we will develop and test our approach using (i) COPD, (ii) pediatric asthma, and (iii) adult asthma. For COPD, we will adjust the criteria used by the Centers for Medicare and Medicaid Services and National Quality Forum [78-80] to incorporate outpatient and emergency room visit data [81] to find COPD patients. A patient is deemed to have COPD if he/she is ≥ 40 and has

- 1) 1 outpatient visit diagnosis code of COPD (ICD-9 (International Classification of Diseases, Ninth Revision): 491.21, 491.22, 491.8, 491.9, 492.8, 493.20, 493.21, 493.22, 496; ICD-10 (International Classification of Diseases, Tenth Revision): J41.8, J42, J43.*, J44.*) and ≥ 1 prescription of tiotropium within 6 months of the outpatient visit,
- 2) ≥ 2 outpatient or ≥ 1 emergency room visit diagnosis codes of COPD (ICD-9: 491.21, 491.22, 491.8, 491.9, 492.8, 493.20, 493.21, 493.22, 496; ICD-10: J41.8, J42, J43.*, J44.*),
- 3) ≥ 1 hospital primary discharge diagnosis code of COPD (ICD-9: 491.21, 491.22, 491.8, 491.9, 492.8, 493.20, 493.21, 493.22, 496; ICD-10: J41.8, J42, J43.*, J44.*), or
- 4) ≥ 1 hospitalization with a primary discharge diagnosis code of respiratory failure (ICD-9: 518.81, 518.82, 518.84, 799.1; ICD-10: J80, J96.0*, J96.2*, J96.9*, R09.2) and a secondary discharge diagnosis code of acute exacerbation of COPD (ICD-9: 491.21, 491.22, 493.21, 493.22; ICD-10: J44.0, J44.1).

The outcome measure is whether a patient used the hospital (inpatient stay and emergency room visit) with a primary diagnosis of COPD (ICD-9: 491.21, 491.22, 491.8, 491.9, 492.8, 493.20, 493.21, 493.22, 496; ICD-10: J41.8, J42, J43.*, J44.*) in the subsequent year.

For asthma, we will use Schatz *et al.*’s [20, 82, 83] method to find asthma patients. A patient is deemed to “have asthma if he/she has 1) ≥ 1 diagnosis code of asthma (ICD-9 493.*; ICD-10 J45/J46.*) or 2) ≥ 2 asthma-related medication dispensing records (excluding oral steroids) in a one-year period, including β -agonists (excluding oral terbutaline), inhaled steroids, other inhaled anti-inflammatory drugs, and oral leukotriene modifiers.” [84] The outcome measure is whether a patient used the hospital with a primary diagnosis of asthma (ICD-9 493.*; ICD-10 J45/J46.*) in the subsequent year.

Temporal feature extraction

We will use a new method to semi-automatically extract predictive and clinically meaningful temporal features from longitudinal data. These features will be used to build the final predictive model and to automatically identify actionable temporal risk factors for each patient at high risk for future hospital use. Our new method is semi-automatic, as its final step involves a human to extract features through visualization [71]. It generalizes to many clinical applications and is sketched as follows, with more details described in our design paper [71].

Our method uses LSTM [35], a type of deep neural network that models long-range dependencies and often reaches higher prediction accuracy than other algorithms [40]. A lot of work has been performed using LSTM to construct predictive models on medical data [36-39, 85]. LSTM performs computations on a sequence of input vectors from the same patient, one vector after another. Every input vector \bar{x}_t is marked by a time step t . After finishing the whole sequence, LSTM gains results that are combined with static attributes like gender [86] to predict the outcome of the patient. Every input vector contains information of one patient visit like vital signs and diagnoses. The sequence length can differ across patients. This helps increase model accuracy, because LSTM can use as much of each patient’s information as possible, without dropping information to make every patient have the same length of history. In addition, this enables us to make timely predictions on new patients without waiting until every patient acquires history of a certain length. With information from only one visit, LSTM can start making projections on the patient.

As Figure 1 shows, an LSTM network includes a sequence of units, one for each time step. In the figure, each unit is denoted by a rounded rectangle. \otimes represents the element-wise multiplication. \oplus represents the element-wise sum. A unit contains an input gate i_t , a hidden state h_t , an output gate o_t , a forget gate f_t , and a memory cell c_t . The memory cell maintains long-term memory and keeps summary information from all of the previous inputs. Every element of the memory cell vector \bar{c}_t represents some learned temporal feature. As shown by Karpathy *et al.* [87], only $\sim 10\%$ of the memory cell vector elements could be interpreted [88]. This is because LSTM puts no limit on the number of input vector elements that can connect to every memory cell vector element. All of the input vector elements could be adopted in every element of the input and forget gates’ activation

vectors, and connect to every memory cell vector element. Consequently, no limit is put on the number of attributes utilized in every learned temporal feature.

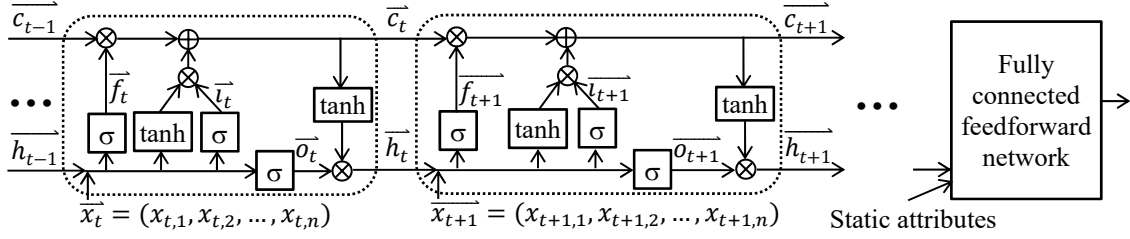


Figure 1. A long short-term memory network.

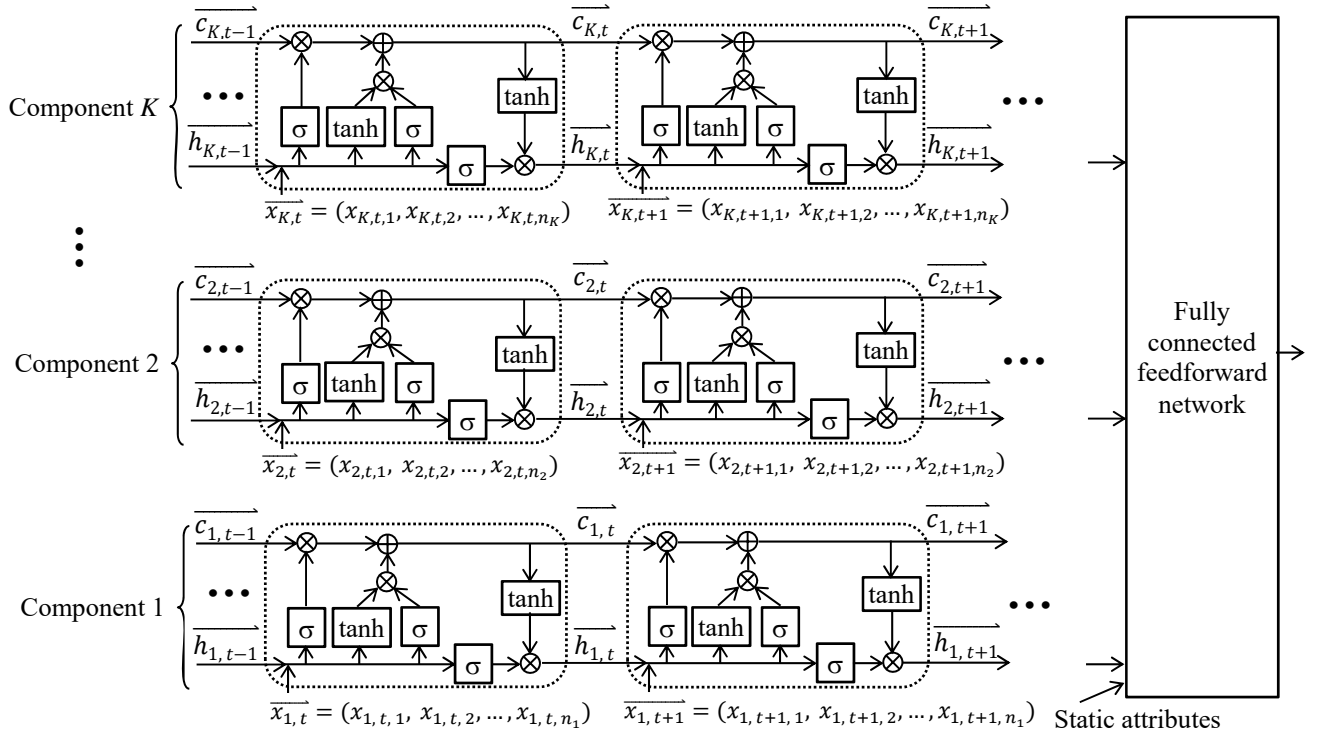


Figure 2. A multi-component long short-term memory network having K components.

It is difficult to understand a feature that involves many attributes. To address this issue, we will use multi-component LSTM (MCLSTM), a new type of LSTM that can automatically drop uninformative attributes. As Figure 2 shows, an MCLSTM has several component LSTM networks, each using some rather than all of the longitudinal attributes. By limiting the number of attributes connecting to every memory cell vector element, more memory cell vector elements will depict clinically meaningful and more generalizable temporal features. Since LSTM often produces more accurate models than other algorithms [36-39], the learned features tend to be predictive. As patient attributes are collected at a different frequency from air quality and weather attributes, we specify certain component networks for the former and the others for the latter. To let data tell which component network uses which attributes, we use a new exclusive group Lasso (least absolute shrinkage and selection operator) regularization method. It combines exclusive Lasso [89, 90] and group Lasso [91] to reach two goals jointly. First, in each component network, every attribute competes with every other attribute. When one is employed, the others are less likely to be employed. Second, in each component network, all of the input vector weight matrix elements connecting to the same attribute tend to become non-zero (or zero) concurrently. Non-zero means the attribute is employed. We will use TensorFlow [62] to train MCLSTM, and use our prior method [84, 92] to automate hyper-parameter value selection.

Kale *et al.* [93-97] showed that in a deep neural network, we can use training instances that incur the highest activations of a neuron to find clinically meaningful features. After training the MCLSTM network, we proceed as follows to identify zero or

more such features from every memory cell vector element at the final time step of the network. First, we find several training instances that incur the highest activations in the memory cell vector element. Second, in each of those training instances, we find one or more segments of the input vector sequence termed effective segments, each tending to represent a useful temporal feature. Third, we partition all spotted effective segments into multiple clusters, and visualize each cluster separately to identify zero or more clinically meaningful temporal features. As shown in Wang *et al.* [98], such visualization could help us find and address data quality issues like an implausible order of events, boosting model accuracy. For each identified feature, Dr. Luo and a clinician in our team will jointly arrive at an exact mathematical definition of an extracted feature. Many extracted features capture trends more precisely than the raw features learned by LSTM. This also boosts model accuracy.

Aim 1's final predictive models

We will use the extracted temporal features to convert longitudinal data to tabular data, with one column per feature, and add static features. Healthcare system features like the number of a physician's patients of a given race can boost model accuracy [64, 65, 99] and are included as static features. We will employ Weka [61] to construct Aim 1's final predictive models. As shown in Aim 4, these models are suitable for automatic explanation. We will use supervised algorithms and our prior method [84, 92] to automatically select the algorithm, feature selection technique, imbalanced class handling method, and hyper-parameter values among all of the applicable ones. We will do manual fine-tuning if needed.

Using historical data up to the prediction time point, we will build three sets of models, one for each of three combinations: COPD at IH, KPSC, and UWM. For each of IH, KPSC, and UWM, the corresponding set of COPD models will be built for all of the COPD patients in that healthcare system. Unlike integrated healthcare systems IH and KPSC, UWM has most of its patients referred from other healthcare systems and fairly incomplete data on many of its patients. To reduce incomplete data's impact on model accuracy, we will use our prior constraint-based method [100] to find patients tending to receive most of their care at UWM, and build models on and apply models to them. Previously, we showed that a good constraint for all types of UWM patients on average is that the patient lives within 5 miles of a UWM hospital and has a UWM primary care physician [100]. Yet, the optimal distance threshold could vary across various types of patients because of their different characteristics. Intuitively, a UWM COPD patient is likely to keep using and get a large portion of his/her care from UWM, even if the patient lives at some distance away from the closest UWM hospital. In comparison, a patient who visited a UWM emergency room once due to a car accident may no longer use UWM after that visit. We will use the approach in our prior work [100] to find an optimal distance threshold for COPD patients. As noted earlier, we will develop and test our techniques on asthma as well.

Accuracy evaluation and justification of the sample size

The discussion below is for IH data. The cases with KPSC and UWM data are similar. As we need to compute outcomes for the subsequent year, we essentially possess 13 years of IH data over the past 14 years. We will train and test models in a usual way. We will do stratified 10-fold cross validation [61] on the data in the first 12 years to train models and to estimate their accuracy. The data in the 13th year will be employed to gauge the performance of the best models, mirroring future use in clinical practice. We will select the best model using the standard performance metric AUC [61]. A care management program typically enrolls 1-3% of COPD patients [32]. Of the upper 1% of COPD patients the model projects to be at the highest risk of using the hospital, we will report the percentage of patients using the hospital in the subsequent year. For a program taking 1% of COPD patients based on the model's prediction results, this percentage reflects the degree of correct enrollment. To find the variables vital for high accuracy, we will conduct backward elimination [74] to remove features on the condition that accuracy does not drop >0.02 . We will compare the variables vital for high accuracy on IH data with those on KPSC and UWM data. Using the variables available in both the IH and KPSC/UWM data sets, we will build a best predictive model on IH data, and compare the model's accuracy on IH data with that on KPSC/UWM data.

We will test the hypothesis that using our techniques can boost model accuracy. To do this, we will use a two-sided Z-test to compare the AUCs of two predictive models built in a way like that in Obuchowski [101]. The first predictive model will use the best machine learning algorithm and take all features essential for high accuracy. The second model will be adapted from those in the literature. For each predictive model for hospital usage reported in the literature [19-28], we will retrain it on our data set using the attributes appearing in both the original model and our data set. The most accurate one of the retrained models will be the second model. Our hypothesis is:

Null hypothesis: The first model reaches the same AUC as the second.

Alternative hypothesis: The first model reaches a higher AUC than the second.

The categorical outcome variable of hospital usage has two possible values (classes). To the best of our knowledge, every predictive model for hospital usage reported in the literature reaches an AUC <0.8 [19-28]. "Using a two-sided Z-test at a significance level of 0.05 and assuming for both classes a correlation coefficient of 0.6 between the two models' prediction results, a sample size of 137 instances per class has 90% power to detect a difference of 0.1 in AUC between the two models," [65] like an increase of AUC from 0.8 to 0.9. The IH data in the 13th year include around 35,000 COPD patients, offering

enough power to test our hypothesis. This conclusion remains valid if the actual correlation coefficient differs somewhat from the assumed one.

Aim 2: Assess using temporal features' impact on risk warning timeliness.

The discussion below is for IH data. The cases with KPSC and UWM data and with asthma are similar.

Outcome of the number of days of early warning the model provides for the patient and the estimation approach

Consider a predictive model and a patient who used the hospital on date D in the 14th year. The outcome is the number of days of early warning the model provides for the patient. To measure the number, we find the first date D' ($D-365 \leq D' \leq D-1$) such that if we use D' as the prediction time point and input the patient's history up to D' into the model, the model predicts hospital use in the subsequent year. In this case, the model warns the first hospital use k ($0 \leq k \leq D-D'$) days in advance, with $D'+k$ being the first day between D' and D when the patient used the hospital. k is the outcome number. Otherwise, if the model still predicts no hospital use when we reach $D-1$, the model warns zero day in advance and zero is the outcome number. We expect using our techniques will raise the outcome number. We will assess the outcome on the cohort of COPD patients who ever used the hospital during the 14th year. For these patients, the average number of days of early warning given by the model shows how timely it warns.

Outcome evaluation and justification of the sample size

We will test the hypothesis that using our techniques can boost risk warning timeliness. To do this, for the patient cohort, we will use an F-test to compare the number of days of early warnings provided by the two models mentioned in Aim 1's accuracy evaluation section, assuming a Poisson model with an offset of 365 days. Our hypothesis is:

Null hypothesis: The number of days of early warning provided by the first model is the same as that provided by the second.

Alternative hypothesis: The number of days of early warning provided by the first model is larger than that provided by the second.

Assuming the number of days of early warning has an exponential distribution and employing an F-test at a one-sided significance level of 0.05, a sample size of 600 patients offers 80% power to detect a minimum raise of 27.8 days of early warning by the first model, when the second model warns on average 180 days in advance. About 2,000 COPD patients ever used IH hospitals during the 14th year, giving enough power to test our hypothesis. The conclusion remains valid if the actual situation differs somewhat from the assumed one.

For Aims 1 and 2, our goal is to reach a boost of ≥ 0.1 in accuracy and ≥ 30 days in risk warning timeliness, respectively. If we cannot reach this goal on the entire COPD patient group, we will construct distinct models for differing patient subgroups. The patient subgroups are described by characteristics like age or co-morbidity, which are often independent variables in the original predictive models. If we still cannot reach this goal, we will do subanalyses to find the patient subgroups, for which our predictive models show good performance, and then apply our ultimate predictive models only to these patient subgroups.

Aim 3: Assess using temporal features' potential impact on outcomes via simulations.

To assess the value of a predictive model for future clinical deployment, we need to appraise care management outcomes if the model is adopted, and decide how to generalize the predictive model to other sites gathering differing sets of variables. Our predictive models will be constructed on IH, KPSC, and UWM data. Our simulations will guide how to employ the predictive models in other healthcare systems. No prior study has decided the variables most crucial for COPD and asthma model generalization. We will apply our simulation method to care management of 1) COPD patients, 2) asthmatic children, and 3) asthmatic adults.

Outcomes of the number of inpatient stays and the number of emergency room visits in the subsequent year and the estimation approach

The number of inpatient stays in the subsequent year is the primary outcome. The number of emergency room visits in the subsequent year is the secondary outcome. The following discussion focuses on IH data and inpatient stays. The cases with KPSC and UWM data and/or emergency room visits can be handled similarly. From statistics reported in the literature [102, 103], we will obtain the percentage of inpatient stays p a care management program can help avoid. Given a set of variables, we will adopt the same method used in Aim 1 to train a predictive model on the data in the first 12 years. For the data in the 13th year, we will gather prediction results, then estimate the outcome. Consider a patient who will have n_e inpatient stays in the subsequent year without enrolling in the program. If the patient gets enrolled, for each inpatient stay of the patient, we will simulate whether it will occur or not based on probability $1-p$. The gross outcome estimate will be the sum of the estimated outcomes of all patients. Adopting a similar method, we will find the minimum accuracy required for the predictive model to be valuable in clinical practice.

Sensitivity analysis

IH, KPSC, and UWM collect many variables. Another healthcare system may collect fewer. To ensure generalizability, we will evaluate various variable combinations and obtain the estimated outcomes when the revised model is adopted. These estimates will pinpoint crucial variables. If a crucial variable is not available in a given healthcare system, these estimates can hint alternative variables having minimal adverse impact on the outcomes.

We will employ a variable grouping approach relating variables likely to co-exist, such as those linked in a lab test panel, according to the judgment of our clinical experts. We will create and post a table showing many possible combinations of variables by groups, encompassing the trained parameters and the simulated outcomes of the predictive model. A healthcare system wanting to deploy the model can employ this table to estimate the expected outcomes in the system's data environment, as well as to determine the variables to be gathered. The table has three columns, one for each of IH, KPSC, and UWM. Many variables collected by IH, KPSC, and UWM and used in this study are commonly available in many other systems. Thus, all variables used in each of many rows in the table will already exist in these systems.

Outcome evaluation and justification of the sample size

The following discussion focuses on IH data. The cases with KPSC and UWM data are similar. We will employ McNemar's test to compare the paired-sample outcomes reached by the two predictive models mentioned in Aim 1's accuracy evaluation section. We will test two hypotheses: using our techniques will link to a potential drop in 1) inpatient stays and 2) emergency room visits in the subsequent year. Our primary hypothesis is:

Null hypothesis: The number of inpatient stays in the subsequent year reached by the first model is the same as that reached by the second.

Alternative hypothesis: The number of inpatient stays in the subsequent year reached by the first model is smaller than that reached by the second.

Among the patients truly at high risk for future hospital use, the first model will find some missed by the second and vice versa. Assuming the former cuts inpatient stays in the subsequent year by 5% and the latter increases them by 1%, at a one-sided significance level of 0.05, a sample size of 251 instances provides 80% power to verify the primary hypothesis. The IH data in the 13th year cover about 35,000 COPD patients, offering enough power to test the primary hypothesis.

Aim 4: Automatically identify actionable temporal risk factors for each patient at high risk for future hospital use and assess the impact on prediction accuracy and outcomes.

Care managers currently give finite input on the patient to the other clinical care team members. Due to bandwidth constraint, care managers can afford to examine only a finite number of patients top ranked by the predictive model - those whose projected risk for future hospital use is over a given threshold like the 95th percentile. For those patients, we will automatically explain early warnings, identify actionable temporal risk factors, and suggest tailored interventions. This helps care managers make enrollment decisions and form tailored care plans. This also enables care managers to pass actionable information on to the other members in the clinical care teams and collaborate more closely with them. To implement the new function, we will improve our prior method [59] to automatically explain a machine learning model's prediction results without incurring any accuracy loss. For non-longitudinal tabular data, our prior method separates explanation and prediction by employing two models simultaneously, each for a distinct purpose. The first model gives predictions to maximize accuracy. The second employs class-based association rules mined from historical data to explain the first model's results. Our prior method cannot handle longitudinal attributes and has not yet been applied to COPD, asthma, or care management.

As mentioned in Aim 1, we will use temporal features to convert longitudinal data to tabular data, with one column per feature. Then we can apply our prior automatic explanation method [59]. Each patient is represented by the same set of features and is marked as either high risk for future hospital use or not. From historical data, our method mines association rules linked to high risk. An example rule is: the ratio of inhaled steroid to beta agonist dispensing to the patient decreased over 12 months AND sulfur dioxide level was ≥ 3 parts per million for ≥ 5 days in the past week \rightarrow high risk. The first item on the left hand side of the rule is an actionable temporal risk factor. Two interventions for the first item are to 1) assess COPD controller medication compliance, and change, prescribe, or raise the dose of the medication if needed; and 2) assess the patient for COPD triggers and ensure the patient stays away from them. Our paper [71] listed several interventions for a few other temporal risk factors. By discussion and consensus, the clinical experts in our team will check the mined rules and drop those making little or no sense clinically. For every rule that remains, our clinical team will mark the actionable temporal risk factors in it and list zero or more interventions that address the reason shown by the rule.

At the time of prediction, for every patient our most accurate model projects to be at high risk for future hospital use, we will find and show all of the association rules whose left hand side conditions the patient satisfies, and list the interventions linking

to these rules as our suggestions. Each rule shows a reason why high risk is anticipated for the patient. Users of our automatic explanation function can give feedback to help us find and drop unreasonable rules [64].

Boost automatic explanations' coverage of patients, model accuracy, and generalizability

For a non-trivial portion of patients, our prior automatic explanation method [59] cannot explain the prediction results of the model. Our prior method employs a conventional approach to mine association rules at a specific level of two parameters: minimum confidence and support. This approach is suboptimal for imbalanced data. There, the outcome variable takes the high-risk value for future hospital use much more often than the low-risk one. Adopting the same minimum support for both values is inadequate [104]. If the minimum support is too small, the rule mining process will form many overfitted rules, making it daunting for clinicians to check all of the mined rules. If the minimum support is large, we cannot identify enough rules for the high-risk value. Consequently, for many patients projected to be at high risk for future hospital use, we cannot explain the prediction results of the model.

To enlarge automatic explanations' coverage of patients, we will use a new technique. It generalizes to many clinical applications and is sketched as follows, with more details given in our design paper [71]. We will use Paul *et al.*'s [104] approach to mine association rules, by replacing support by value-specific support termed commonality. This has two advantages. First, the rule mining process produces fewer overfitted rules, cutting the time clinicians need to check the mined rules. Second, we obtain more rules for the high-risk value of the outcome variable. Thus, for more patients projected to be at high risk for future hospital use, we can explain the prediction results of the model.

Using automatic explanations and the method described in our paper [64], we will find and drop uninformative features, and retrain the predictive model. For the model, this can boost its accuracy, as well as make it generalize better to other healthcare systems beyond where it was originally built. On non-medical data, Ribeiro *et al.* [105] showed a similar method with a narrower scope boosted model accuracy by ~10%.

Performance evaluation

We will compare the association rules obtained from IH, KPSC, and UWM data. The following discussion focuses on IH data. The cases with KPSC and UWM data are similar. We will do analyses similar to those in Aim 1 to compare using our new techniques in Aim 4 vs. the current method of offering no explanation. We will compare the outcomes of the number of inpatient stays and the number of emergency room visits in the subsequent year and the accuracy reached by two models: the best ones produced in Aims 1 and 4, respectively. We will test three hypotheses: using our new techniques in Aim 4 will link to a potential drop in 1) inpatient stays and 2) emergency room visits in the subsequent year, and 3) boost prediction accuracy. Our primary hypothesis is:

Null hypothesis: The number of inpatient stays in the subsequent year reached by the second model is the same as that reached by the first.

Alternative hypothesis: The number of inpatient stays in the subsequent year reached by the second model is smaller than that reached by the first.

We will employ McNemar's test to compare the paired-sample outcomes reached by the two models.

Among the patients truly at high risk for future hospital use, the second model will find some cases missed by the first and vice versa. Assuming the former cuts inpatient stays in the subsequent year by 2.5% and the latter increases them by 0.5% and using McNemar's test, at a one-sided significance level of 0.05, a sample size of 503 instances provides 80% power to verify the primary hypothesis. The IH data in the 13th year cover ~35,000 COPD patients, offering enough power to test the primary hypothesis.

To assess our technique's impact on model generalizability, we will compare two predictive models' accuracy on KPSC/UWM data. The first model is the best one produced in Aim 1 on IH data using the variables available in both the IH and KPSC/UWM data sets. The second is produced by using our technique to drop uninformative features from the first model and retrain it on IH data. We will develop and test our techniques on asthma as well.

Aim 5: Assess actionable information's impact on clinicians' acceptance of early warnings and on perceived care plan quality.

As an essential preparatory step for future clinical deployment, we will evaluate actionable information's impact on UWM care managers and physicians' decision making in a test setting. For physicians, we will use primary care physicians, pulmonologists, and allergists managing COPD patients. The discussion below focuses on care managers. The case of evaluating with 10 physicians is similar.

Subject recruitment

As an operational project at UWM, we are working on COPD outcome prediction and can access ~25 UWM care managers for adults. Via making announcements in their email lists and personal contact, we will recruit 10 care managers. We will adopt

purposeful sampling to ensure adequate variability in work experience [106]. All evaluation test participants will give consent and be up-to-date on privacy and information security policy training required by UWM. Participants will obtain pseudonyms connecting their responses to questions to protect privacy. After completing the task, each will obtain U.S. \$2,400 as compensation for participation for ~40 hours' work. We will conduct two experiments.

Experiment 1

Procedures

From the IH data in the 13th year, we will randomly select 400 IH COPD patients who used the hospital in the subsequent year, and automatically explain the prediction results of the best IH model built in Aim 4. We will use patients outside of UWM to help ensure no care manager is aware of any of those patients' outcome in the subsequent year. We will show every care manager a different subset of 40 patients and proceed in three steps:

- (1) Step 1: For every patient, we will present the historical de-identified patient attributes and the prediction result to the care manager, and ask him/her to record the enrollment decision and interventions, if any, that he/she plans to use on the patient. For the historical patient attributes, we will show the static attributes' values at the top, followed by the longitudinal attributes' values in reverse chronological order. No care manager will see any automatic explanation in this step.
- (2) Step 2: For every patient, we will present the automatic explanations and their linked interventions to the care manager, and survey him/her using both semi-structured and open-ended questions. The automatic explanations will appear as a list of association rules. Below each rule is the list of interventions linked to the rule. The questions will include whether these explanations would change the enrollment decision for the patient, whether he/she believes they would improve care plan quality, their usefulness on a 1-7 scale with anchors of not at all/very useful, and their perceived trustworthiness on a 1-7 scale with anchors of not at all/completely. Our questionnaire will embrace a text field for writing comments.
- (3) Step 3: We will use the standard Technology Acceptance Model (TAM) satisfaction questionnaire [107] to survey the care manager about the automatic explanations. A technology is unimportant unless people will accept and use it. Developed based on multiple well-accepted behavioral theories, TAM is the most widely adopted model of people's acceptance and usage of a technology. The TAM satisfaction questionnaire will measure the perceived usefulness and the perceived ease of use of automatic explanations. Perceived usefulness is known to link strongly to future usage intentions and to actual function usage [108, 109]. Multiple studies have demonstrated the validity and reliability of the TAM satisfaction questionnaire [110, 111].

Analysis

We will adopt the inductive approach described in Patton *et al.* [106, 112] to conduct qualitative analysis. Care managers' textual comments will be put into ATLAS.ti qualitative analysis software [113]. Three people in our research team will independently highlight quotations on prediction results and automatic explanations for all records. Quotations will be examined, classified into pre-codes, and merged into categories via discussion and negotiated consensus in several iterations. We will find general themes via synthesis of categories. The quantitative analyses will include giving descriptive statistics for every quantitative outcome measure. We will test the hypothesis that for the patients who will use the hospital in the next year, giving actionable information will improve the perceived care plan quality. Our hypothesis is:

Null hypothesis: For the patients who will use the hospital in the next year, the care manager does not believe that showing the automatic explanations and their linked interventions would improve care plan quality.

Alternative hypothesis: For the patients who will use the hospital in the next year, the care manager believes that showing the automatic explanations and their linked interventions would improve care plan quality

We will fit a random effect logistic model to account for correlation among the outcomes of the same care manager on whether the perceived care plan quality is improved.

Justification of the sample size

Assuming a moderate intra-class correlation of 0.1 within the same care manager on the outcome of whether the perceived care plan quality is improved, a sample size of 40 instances per care manager for 10 care managers is equivalent to totally 82 independent instances after adjusting for the clustering effect. At a two-sided significance level of 0.05, we will have 80% power to identify a 9.7% increase in the odds of improving the perceived care plan quality with actionable information. A similar conclusion holds if the actual correlation differs somewhat from the assumed one.

If giving actionable information has no significant impact on the perceived care plan quality on the whole group of COPD patients, we will do subanalyses to find those patient subgroups on which significant impact occurs.

Experiment 2

Procedures

We will randomly partition the 10 care managers into two disjoint groups: the intervention group and the control group. Each group has five care managers. From the IH data in the 13th year, we will randomly select 200 IH COPD patients who used the hospital in the subsequent year and whose data are unused in Experiment 1, and automatically explain the prediction results of the best IH model built in Aim 4. For each group, we will show every care manager in the group a different subset of 40 patients. With random assignment, each patient is shown to two care managers, one in the intervention group and the other in the control group. In the control group, for every patient, we will show the care manager the historical de-identified patient attributes and the prediction result, but no automatic explanation. In the intervention group, for every patient, we will show the care manager the historical de-identified patient attributes, the prediction result, the automatic explanations and their linked interventions. In both groups, we will ask the care managers to record their enrollment decisions.

Analysis

We will test the hypothesis that for a patient who will use the hospital in the next year, giving actionable information will increase the likelihood that a care manager decides to enroll the patient in care management. Our hypothesis is:

Null hypothesis: For a patient who will use the hospital in the next year, the likelihood that a care manager in the intervention group decides to enroll the patient in care management is the same as that in the control group.

Alternative hypothesis: For a patient who will use the hospital in the next year, the likelihood that a care manager in the intervention group decides to enroll the patient in care management is higher than that in the control group.

We will fit a random effect logistic model to compare care managers' enrollment decision outcomes between the intervention group and the control group, while accounting for correlation among the enrollment decision outcomes of the same care manager and correlation between the enrollment decision outcomes of the same patient examined by two care managers.

Justification of the sample size

Assuming a moderate intra-class correlation of 0.1 within the same care manager and within the same patient examined by two care managers on the enrollment decision outcome, a sample size of 40 instances per care manager for 10 care managers is equivalent to totally 74 independent instances after adjusting for the clustering effect. At a two-sided significance level of 0.05, we will have 80% power to identify a 9.7% boost in the intervention group in the odds that a care manager decides to enroll the patient in care management. A similar conclusion holds if the actual correlations differ somewhat from the assumed ones.

As mentioned right before Aim 1, the above discussion focuses on COPD. Whenever we mention COPD, the same applies to asthma and will be developed and tested on asthma also in Aims 1-5.

Ethics approval

We have obtained from IH, UWM, and KPSC institutional review board approvals for this study.

3. Results

We are currently downloading clinical and administrative data from the EDWs of UWM, KPSC, and IH. We plan to start data analysis in 2020 and finish our study in 2025.

4. Discussion

Clinical use of our results

Care managers collaborate with the other members in the clinical care teams. We will automatically explain early warnings and suggest possible interventions to help clinical care teams form tailored care plans on the grounds of objective data. This could facilitate clinicians to review structured data in patient charts faster, and enable closer collaboration between care managers and the other members in the clinical care teams. Once our methods find patients at the largest projected risks for future hospital use and provide explanations, clinicians will check patient charts, examine factors like social dimensions and potential for improvement [102], and make care management enrollment and intervention decisions.

As time goes by, both the feature patterns linked to high risk for future hospital use and patient status keep changing. In clinical practice, we can re-apply our techniques regularly to the latest clinical, administrative, air quality, and weather data sets to move patients into and out of care management and to find new feature patterns over time.

As in the case with LSTM, with information from only one visit, our proposed predictive models can start making projections on the patient. Yet, all else being equal, we would expect the prediction accuracy and risk warning timeliness reached by our models to improve as the length of patient history increases.

Generalizability

We will semi-automatically extract predictive and clinically meaningful temporal features from longitudinal data, solving an open computer science challenge [60]. Both our feature extraction and automatic explanation methods will help drop uninformative variables, reducing the variables used in the model. This boosts model generalizability and partly addresses the limitation that one study cannot afford to test models on all U.S. patients. As Gupta *et al.* [114] showed, many extracted features represent general properties of the attributes used in the features, and can be valuable for other predictive modeling tasks. Using the extracted features to build a temporal feature library to aid feature reuse, we can cut down the effort required to construct models for other predictive modeling tasks.

The principles of our techniques are general, depending on no unique characteristic of a specific disease, patient cohort, or healthcare system. Care management is also widely used for patients with diabetes and heart diseases [34], where our techniques could be used. Our simulation will find out how to generalize a predictive model to other sites gathering differing sets of variables, and those variables most crucial for generalization. We will use data from three healthcare systems IH, KPSC, and UWM to illustrate our techniques on the cases of COPD and asthma patients. These healthcare systems include two integrated systems (IH and KPSC), an academic system with most patients referred from other systems (UWM), and many heterogeneous facilities. These facilities cover 41 hospitals and 424 clinics spread over three large geographic areas, ranging from rural and community urban clinics staffed by a variety of clinicians including physicians, nurses, therapists, and advanced practice practitioners with limited resources to metropolitan tertiary care hospitals staffed by sub-specialists. These systems use four different electronic health record systems: IH uses Cerner, HELP, and HELP2; UWM uses Cerner and Epic; KPSC uses Epic. Variation in scope of services, staff composition, geographic location, cultural background, patient population, healthcare system type, and electronic health record system allows us to find factors generalizable to other facilities nationwide. Our models will be based on the OMOP common data model [72] and its related standardized terminologies [73], which standardize clinical and administrative variables from ≥ 10 large U.S. healthcare systems [115, 116]. At a minimum, our models will apply to those systems using OMOP.

After extension, our techniques can be applied to various decision support applications and diseases and advance clinical machine learning: 1) More precise models giving earlier warnings will boost decision support tools for managing limited resources, like planning for healthcare resource allocation [117], and automatically finding patients tending to be readmitted soon, triggering home visits by nurses to cut readmissions. 2) Using our techniques can boost prediction accuracy and risk warning timeliness of other outcomes like missed appointments [118], patient satisfaction [119], and adherence to treatment [120]. This will help target resources, such as reminder phone calls to cut missed appointments [118], or interventions to boost adherence to treatment [120].

We expect our more accurate predictive models giving earlier warnings to have value for clinical practice. Future studies will test our techniques on some other patient cohorts and diseases, implement our techniques at UWM, IH, and KPSC for care management for COPD and asthma, and evaluate the impacts in randomized controlled trials.

In summary, the techniques that will be developed in this study will advance machine learning for many clinical applications and help transform preventive care to be more efficient, effective, and timely. This will boost outcomes and save resources.

Acknowledgments

We thank Peter Tarczy-Hornoch, Sean D. Mooney, Trevor A. Cohen, David Crosslin, and Philip J. Brewster for helpful discussions. GL, BLS, CK, SH, XS, MS, RSZ, GHD, and FLN were partially supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number R01HL142503. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

GL was mainly responsible for the paper. He conceptualized and designed the study, performed the literature review, and wrote the paper. FLN, BLS, CK, DHA, MAM, KAS, MS, RSZ, and GHD offered feedback on study design and medical issues and revised the paper. SH took part in retrieving the Intermountain Healthcare data set and interpreting its detected peculiarities. XS took part in conceptualizing and writing the statistical analysis sections. All authors read and approved the final manuscript.

Conflicts of interest

None declared.

Abbreviations

AUC: area under the receiver operating characteristic curve
COPD: chronic obstructive pulmonary disease
EDW: enterprise data warehouse
ICD-9: International Classification of Diseases, Ninth Revision

ICD-10: International Classification of Diseases, Tenth Revision
IH: Intermountain Healthcare
KPSC: Kaiser Permanente Southern California
Lasso: least absolute shrinkage and selection operator
LSTM: long short-term memory
MCLSTM: multi-component LSTM
OMOP: Observational Medical Outcomes Partnership
TAM: Technology Acceptance Model
UWM: University of Washington Medicine

References

1. Ford ES, Murphy LB, Khavjou O, Giles WH, Holt JB, Croft JB. Total and state-specific medical and absenteeism costs of COPD among adults aged ≥ 18 years in the United States for 2010 and projections through 2020. *Chest* 2015;147(1):31-45. PMID:25058738
2. Asthma. <https://www.cdc.gov/nchs/fastats/asthma.htm>. Archived at <http://www.webcitation.org/75kW25twQ>
3. Akinbami LJ, Moorman JE, Liu X. Asthma prevalence, health care use, and mortality: United States, 2005-2009. *Natl Health Stat Report* 2011;(32):1-14. PMID:21355352
4. Akinbami LJ, Moorman JE, Bailey C, Zahran HS, King M, Johnson CA, Liu X. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. *NCHS Data Brief* 2012;(94):1-8. PMID:22617340
5. Asthma in the US. <https://www.cdc.gov/vitalsigns/asthma/>. Archived at <http://www.webcitation.org/75kW929CX>
6. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, Blumenthal D. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. *J Gen Intern Med* 2007;22 Suppl 3:391-5. PMID:18026807
7. Caloyeras JP, Liu H, Exum E, Broderick M, Mattke S. Managing manifest diseases, but not health risks, saved PepsiCo money over seven years. *Health Aff (Millwood)* 2014;33(1):124-31. PMID:24395944
8. Nelson L. Lessons from Medicare's demonstration projects on disease management and care coordination. https://www.cbo.gov/sites/default/files/112th-congress-2011-2012/workingpaper/WP2012-01_Nelson_Medicare_DMCC_Demonstrations_1.pdf. Archived at <http://www.webcitation.org/6agZcFxD1>
9. Commission for Case Manager Certification. Definition and philosophy of case management. <http://ccmcertification.org/about-us/about-case-management/definition-and-philosophy-case-management>. Archived at <http://www.webcitation.org/6agZuMc1N>
10. Levine SH, Adams J, Attaway K, Dorr DA, Leung M, Popescu P, Rich J. Predicting the financial risks of seriously ill patients. California HealthCare Foundation. <https://www.chcf.org/publication/predicting-the-financial-risks-of-seriously-ill-patients/>. Archived at <http://www.webcitation.org/75kTo00VY>
11. Rubin RJ, Dietrich KA, Hawk AD. Clinical and economic impact of implementing a comprehensive diabetes management program in managed care. *J Clin Endocrinol Metab* 1998;83(8):2635-42. PMID:9709924
12. Greineder DK, Loane KC, Parks P. A randomized controlled trial of a pediatric asthma outreach program. *J Allergy Clin Immunol* 1999;103(3 Pt 1):436-40. PMID:10069877
13. Kelly CS, Morrow AL, Shults J, Nakas N, Strobe GL, Adelman RD. Outcomes evaluation of a comprehensive intervention program for asthmatic children enrolled in Medicaid. *Pediatrics* 2000;105(5):1029-35. PMID:10790458
14. Axelrod RC, Zimbro KS, Chetney RR, Sabol J, Ainsworth VJ. A disease management program utilizing life coaches for children with asthma. *J Clin Outcomes Manag* 2001;8(6):38-42.
15. Dorr DA, Wilcox AB, Bruncker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. *J Am Geriatr Soc* 2008;56(12):2195-202. PMID:19093919
16. Beaulieu N, Cutler DM, Ho K, Isham G, Lindquist T, Nelson A, O'Connor P. The business case for diabetes disease management for managed care organizations. *Forum for Health Economics & Policy* 2006;9(1):1-37.
17. Mays GP, Claxton G, White J. Managed care rebound? Recent changes in health plans' cost containment strategies. *Health Aff (Millwood)*. 2004;Suppl Web Exclusives:W4-427-36. PMID:15451964
18. Curry N, Billings J, Darin B, Dixon J, Williams M, Wennberg D. Predictive Risk Project Literature Review. London: King's Fund. http://www.kingsfund.org.uk/sites/files/kf/field/field_document/predictive-risk-literature-review-june2005.pdf. Archived at <http://www.webcitation.org/6aga2XBFC>
19. Schatz M, Nakahiro R, Jones CH, Roth RM, Joshua A, Petitti D. Asthma population management: development and validation of a practical 3-level risk stratification scheme. *Am J Manag Care* 2004;10(1):25-32. PMID:14738184

20. Schatz M, Cook EF, Joshua A, Petitti D. Risk factors for asthma hospitalizations in a managed care organization: development of a clinical prediction rule. *Am J Manag Care* 2003;9(8):538-47. PMID:12921231
21. Lieu TA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high-risk children with asthma. *Am J Respir Crit Care Med* 1998;157(4 Pt 1):1173-80. PMID:9563736
22. Lieu TA, Capra AM, Quesenberry CP, Mendoza GR, Mazar M. Computer-based models to identify high-risk adults with asthma: is the glass half empty or half full? *J Asthma* 1999;36(4):359-70. PMID:10386500
23. Forno E, Fuhlbrigge A, Soto-Quirós ME, Avila L, Raby BA, Brehm J, Sylvia JM, Weiss ST, Celedón JC. Risk factors and predictive clinical scores for asthma exacerbations in childhood. *Chest* 2010;138(5):1156-65. PMID:20472862
24. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur Respir Rev* 2017;26(143). pii: 160061. PMID:28096287
25. Orchard P, Agakova A, Pinnock H, Burton CD, Sarran C, Agakov F, McKinstry B. Improving prediction of risk of hospital admission in chronic obstructive pulmonary disease: application of machine learning to telemonitoring data. *J Med Internet Res* 2018;20(9):e263. PMID:30249589
26. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. *J Med Internet Res* 2017;19(3):e69. PMID:28270380
27. Loymans RJB, Debray TPA, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Schermer TRJ, Assendelft WJJ, Timp M, Chung KF, Sousa AR, Sont JK, Sterk PJ, Reddel HK, Ter Riet G. Exacerbations in adults with asthma: a systematic review and external validation of prediction models. *J Allergy Clin Immunol Pract* 2018;6(6):1942-52.e15. PMID:29454163
28. Miller MK, Lee JH, Blanc PD, Pasta DJ, Gujrathi S, Barron H, Wenzel SE, Weiss ST. TENOR risk score predicts healthcare in adults with severe or difficult-to-treat asthma. *Eur Respir J* 2006;28(6):1145-55. PMID:16870656
29. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analyzing health care utilization and costs. *Annu Rev Public Health* 1999;20:125-44. PMID:10352853
30. Hileman G, Steele S. Accuracy of claims-based risk scoring models. <https://www.soa.org/research-reports/2016/2016-accuracy-claims-based-risk-scoring-models/>. Archived at <http://www.webcitation.org/75kUQVCAe>
31. Ash A, McCall N. Risk assessment of military populations to predict health care cost and utilization. http://www.rti.org/pubs/tricare_riskassessment_final_report_combined.pdf. Archived at <http://www.webcitation.org/6aga7wYZC>
32. Axelrod RC, Vogel D. Predictive modeling in health plans. *Disease Management & Health Outcomes* 2003;11(12):779-87. doi:10.2165/00115677-200311120-00003
33. Weir S, Aweh G, Clark RE. Case selection for a Medicaid chronic care management program. *Health Care Financ Rev* 2008;30(1):61-74. PMID:19040174
34. Duncan I. *Healthcare Risk Adjustment and Predictive Modeling*. Winsted, CT: ACTEX Publications Inc.; 2011. ISBN:1566987695
35. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation* 1997;9(8):1735-80. doi:10.1162/neco.1997.9.8.1735
36. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, Liu PJ, Liu X, Marcus J, Sun M, Sundberg P, Yee H, Zhang K, Zhang Y, Flores G, Duggan GE, Irvine J, Le Q, Litsch K, Mossin A, Tansuwan J, Wang D, Wexler J, Wilson J, Ludwig D, Volchenboum SL, Chou K, Pearson M, Madabushi S, Shah NH, Butte AJ, Howell M, Cui C, Corrado GS, Dean J. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine* 2018;1:18. doi:10.1038/s41746-018-0029-1
37. Lipton ZC, Kale DC, Elkan C, Wetzel RC. Learning to diagnose with LSTM recurrent neural networks. In: *Proceedings of the International Conference on Learning Representations*. 2016 Presented at: International Conference on Learning Representations; May 2-4, 2016; San Juan, Puerto Rico p. 1-18.
38. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med* 2017;89:248-55. PMID:28843829
39. Razavian N, Marcus J, Sontag D. Multi-task prediction of disease onsets from longitudinal laboratory tests. In: *Proceedings of the Machine Learning in Health Care Conference*. 2016 Presented at: Machine Learning in Health Care Conference; August 19-20, 2016; Los Angeles, CA p. 73-100.
40. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press; 2016. ISBN:0262035618
41. Combi C, Keravnou-Papailiou E, Shahar Y. *Temporal Information Systems in Medicine*. New York, NY: Springer; 2010. ISBN:9781441965424
42. Forno E, Celedón JC. Predicting asthma exacerbations in children. *Curr Opin Pulm Med* 2012;18(1):63-9. PMID:22081091
43. Schatz M. Predictors of asthma control: what can we modify? *Curr Opin Allergy Clin Immunol* 2012;12(3):263-8. PMID:22517290

44. Coyle YM. Predictors of acute asthma relapse: strategies for its prevention. *J Asthma* 2003;40(3):217-24. PMID:12807164
45. Bahadori K, FitzGerald JM. Risk factors of hospitalization and readmission of patients with COPD exacerbation - systematic review. *Int J Chron Obstruct Pulmon Dis* 2007;2(3):241-51. PMID:18229562
46. Fleming L. Asthma exacerbation prediction: recent insights. *Curr Opin Allergy Clin Immunol* 2018;18(2):117-23. PMID:29406359
47. Buelo A, McLean S, Julious S, Flores-Kim J, Bush A, Henderson J, Paton JY, Sheikh A, Shields M, Pinnock H; ARC Group. At-risk children with asthma (ARC): a systematic review. *Thorax*. 2018;73(9):813-24. PMID:29871982
48. Puranik S, Forno E, Bush A, Celedón JC. Predicting severe asthma exacerbations in children. *Am J Respir Crit Care Med* 2017;195(7):854-9. PMID:27710010
49. Evans RS. Electronic health records: then, now, and in the future. *Yearb Med Inform* 2016;Suppl 1:S48-61. PMID:27199197
50. Dick S, Doust E, Cowie H, Ayres JG, Turner S. Associations between environmental exposures and asthma control and exacerbations in young children: a systematic review. *BMJ Open* 2014;4(2):e003827. PMID:24523420
51. Li J, Sun S, Tang R, Qiu H, Huang Q, Mason TG, Tian L. Major air pollutants and risk of COPD exacerbations: a systematic review and meta-analysis. *Int J Chron Obstruct Pulmon Dis* 2016;11:3079-91. PMID:28003742
52. Hansel NN, McCormack MC, Kim V. The effects of air pollution and temperature on COPD. *COPD* 2016;13(3):372-9. PMID:26683097
53. Halamka JD. Early experiences with big data at an academic medical center. *Health Aff (Millwood)* 2014;33(7):1132-8. PMID:25006138
54. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956;63(2):81-97. PMID:13310704
55. Iezzoni LI. Risk adjustment for measuring health care outcomes, 4th ed. Chicago, IL: Health Administration Press; 2013. ISBN:1567934374
56. James BC, Savitz LA. How Intermountain trimmed health care costs through robust quality improvement efforts. *Health Aff* 2011;30(6):1185-91. PMID:21596758
57. The Dartmouth Atlas of Health Care. <http://www.dartmouthatlas.org/data/topic/topic.aspx?cat=21>. Archived at <http://www.webcitation.org/6agaBuCzH>
58. Pacific Business Group on Health. Advancing physician performance measurement: Using administrative data to assess physician quality and efficiency. http://www.pbgh.org/storage/documents/reports/PBGHP3Report_09-01-05final.pdf. Archived at <http://www.webcitation.org/6agaGjTgH>
59. Luo G. Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. *Health Inf Sci Syst* 2016;4:2. PMID:26958341
60. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Comput Surv* 2019;51(5):93. doi:10.1145/3236009
61. Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Burlington, MA: Morgan Kaufmann; 2016. ISBN:0128042915
62. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker PA, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X. TensorFlow: a system for large-scale machine learning. In: *Proceedings of the USENIX Symposium on Operating Systems Design and Implementation*. 2016 Presented at: USENIX Symposium on Operating Systems Design and Implementation; October 8-10, 2018; Carlsbad, CA p. 265-83.
63. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 2002;16:321-57. doi:10.1613/jair.953
64. Luo G, Sward K. A roadmap for optimizing asthma care management via computational approaches. *JMIR Med Inform* 2017;5(3):e32. PMID:28951380
65. Luo G, Stone BL, Sakaguchi F, Sheng X, Murtaugh MA. Using computational approaches to improve risk-stratified patient management: rationale and methods. *JMIR Res Protoc* 2015;4(4):e128. PMID:26503357
66. Evans RS, Lloyd JF, Pierce LA. Clinical use of an enterprise data warehouse. *AMIA Annu Symp Proc* 2012:189-98. PMID:23304288
67. Byington CL, Reynolds CC, Korgenski K, Sheng X, Valentine KJ, Nelson RE, Daly JA, Osguthorpe RJ, James B, Savitz L, Pavia AT, Clark EB. Costs and infant outcomes after implementation of a care process model for febrile infants. *Pediatrics* 2012;130(1):e16-24. PMID:22732178
68. Koebnick C, Langer-Gould AM, Gould MK, Chao CR, Iyer RL, Smith N, Chen W, Jacobsen SJ. Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. *Perm J* 2012;16(3):37-41. PMID:23012597

69. Air quality data homepage of the United States Environmental Protection Agency. <https://www.epa.gov/outdoor-air-quality-data>. Archived at <http://www.webcitation.org/77x6RHRbH>
70. MesoWest homepage. <https://mesowest.utah.edu>. Archived at <http://www.webcitation.org/77x6T5G7z>
71. Luo G. A roadmap for semi-automatically extracting predictive and clinically meaningful temporal features from medical data for predictive modeling. *Global Transitions* 2019;1:61-82. doi:10.1016/j.glt.2018.11.001
72. Observational Medical Outcomes Partnership (OMOP) Common Data Model homepage. <https://www.ohdsi.org/data-standardization/the-common-data-model/>. Archived at <http://www.webcitation.org/75kWBuPaS>
73. Observational Medical Outcomes Partnership (OMOP) standardized vocabularies. <http://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:sidebar>. Archived at <http://www.webcitation.org/75kWjC13b>
74. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer; 2009. ISBN:038777243X
75. Pyle D. *Data Preparation for Data Mining*. San Francisco, CA: Morgan Kaufmann; 1999. ISBN:1558605290
76. Bivand RS, Pebesma E, Gómez-Rubio V. *Applied Spatial Data Analysis with R*, 2nd ed. New York, NY: Springer; 2013. ISBN:9781461476177
77. Luo G, Stone BL, Fassl B, Maloney CG, Gesteland PH, Yerram SR, Nkoy FL. Predicting asthma control deterioration in children. *BMC Med Inform Decis Mak* 2015;15(1):84. PMID:26467091
78. Lindenauer PK, Grosso LM, Wang C, Wang Y, Krishnan JA, Lee TA, Au DH, Mularski RA, Bernheim SM, Drye EE. Development, validation, and results of a risk-standardized measure of hospital 30-day mortality for patients with exacerbation of chronic obstructive pulmonary disease. *J Hosp Med* 2013;8(8):428-35. PMID:23913593
79. National Quality Forum. NQF #1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization. http://www.qualityforum.org/Projects/n-r/Pulmonary_Endorsement_Maintenance/1891_30_Day_RSRR_COPD.aspx. Archived at <http://www.webcitation.org/75kXWso5s>
80. Nguyen HQ, Chu L, Amy Liu IL, Lee JS, Suh D, Korotzer B, Yuen G, Desai S, Coleman KJ, Xiang AH, Gould MK. Associations between physical activity and 30-day readmission risk in chronic obstructive pulmonary disease. *Ann Am Thorac Soc* 2014;11(5):695-705. PMID:24713094
81. Cooke CR, Joo MJ, Anderson SM, Lee TA, Udriș EM, Johnson E, Au DH. The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease. *BMC Health Serv Res* 2011;11:37. PMID:21324188
82. Desai JR, Wu P, Nichols GA, Lieu TA, O'Connor PJ. Diabetes and asthma case identification, validation, and representativeness when using electronic health data to construct registries for comparative effectiveness and epidemiologic research. *Med Care* 2012;50 Suppl:S30-5. PMID:22692256
83. Wakefield DB, Cloutier MM. Modifications to HEDIS and CSTE algorithms improve case recognition of pediatric asthma. *Pediatr Pulmonol* 2006;41(10):962-71. PMID:16871628
84. Luo G, Stone BL, Johnson MD, Tarczy-Hornoch P, Wilcox AB, Mooney SD, Sheng X, Haug PJ, Nkoy FL. Automating construction of machine learning models with clinical big data: proposal rationale and methods. *JMIR Res Protoc* 2017;6(8):e175. PMID:28851678
85. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform* 2018;22(5):1589-604. PMID:29989977
86. Velickovic P, Karazija L, Lane ND, Bhattacharya S, Liberis E, Liò P, Chieh A, Bellahsen O, Vegreville M. Cross-modal recurrent models for weight objective prediction from multimodal time-series data. In: *Proceedings of the EAI International Conference on Pervasive Computing Technologies for Healthcare*. 2018 Presented at: EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2018; New York, NY p. 178-86.
87. Karpathy A, Johnson J, Li F. Visualizing and understanding recurrent networks. In: *Proceedings of the International Conference on Learning Representations Workshop*. 2016 Presented at: International Conference on Learning Representations Workshop; May 2-4, 2016; San Juan, Puerto Rico p. 1-12.
88. Krakovna V, Doshi-Velez F. Increasing the interpretability of recurrent neural networks using hidden Markov models. In: *Proceedings of the ICML Workshop on Human Interpretability in Machine Learning*. 2016 Presented at: ICML Workshop on Human Interpretability in Machine Learning; June 23, 2016; New York, NY p. 46-50.
89. Zhou Y, Jin R, Hoi SCH. Exclusive Lasso for multi-task feature selection. In: *Proceedings of the International Conference on Artificial Intelligence and Statistics*. 2010 Presented at: International Conference on Artificial Intelligence and Statistics; May 13-15, 2010; Sardinia, Italy p. 988-95.
90. Campbell F, Allen GI. Within group variable selection through the exclusive Lasso. *Electron J Statist* 2017;11(2):4220-57. doi:10.1214/17-EJS1317

91. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Statist Soc B* 2006;68(1):49-67. doi:10.1111/j.1467-9868.2005.00532.x
92. Zeng X, Luo G. Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection. *Health Inf Sci Syst* 2017;5(1):2. PMID:29038732
93. Kale DC, Che Z, Bahadori MT, Li W, Liu Y, Wetzel R. Causal phenotype discovery via deep networks. *AMIA Annu Symp Proc* 2015;2015:677-86. PMID:26958203
94. Tang Z, Shi Y, Wang D, Feng Y, Zhang S. Memory visualization for gated recurrent neural networks in speech recognition. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2017 Presented at: *IEEE International Conference on Acoustics, Speech and Signal Processing*; March 5-9, 2017; New Orleans, LA p. 2736-40. doi:10.1109/ICASSP.2017.7952654
95. Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PLoS One* 2013;8(6):e66341. PMID:23826094
96. Che Z, Kale DC, Li W, Bahadori MT, Liu Y. Deep computational phenotyping. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015 Presented at: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; August 10-13, 2015; Sydney, Australia p. 507-16. doi:10.1145/2783258.2783365
97. Kale D, Che Z, Liu Y, Wetzel R. Computational discovery of physiomes in critically ill children using deep learning. In: *Proceedings of the Workshop on Data Mining for Medical Informatics*. 2014 Presented at: *Workshop on Data Mining for Medical Informatics*; Nov 15, 2014; Washington, DC p. 1-2.
98. Wang TD, Wongsuphasawat K, Plaisant C, Shneiderman B. Visual information seeking in multiple electronic health records: design recommendations and a process model. In: *Proceedings of the ACM International Health Informatics Symposium*. 2010 Presented at: *ACM International Health Informatics Symposium*; November 11-12, 2010; Arlington, VA p. 46-55. doi:10.1145/1882992.1883001
99. Neuvirth H, Ozery-Flato M, Hu J, Laserson J, Kohn MS, Ebadollahi S, Rosen-Zvi M. Toward personalized care management of patients at risk: the diabetes case study. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2011 Presented at: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; August 21-24, 2011; San Diego, CA p. 395-403. doi:10.1145/2020408.2020472
100. Luo G, Tarczy-Hornoch P, Wilcox AB, Lee ES. Identifying patients who are likely to receive most of their care from a specific health care system: demonstration via secondary analysis. *JMIR Med Inform* 2018;6(4):e12241. PMID:30401670
101. Obuchowski NA, McClish DK. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Stat Med* 1997;16(13):1529-42. PMID:9249923
102. Duncan I. *Managing and Evaluating Healthcare Intervention Programs*, 2nd ed. Winsted, CT: ACTEX Publications; 2014. ISBN:1625421125
103. Wang LY, Zhong Y, Wheeler L. Direct and indirect costs of asthma in school-age children. *Prev Chronic Dis* 2005;2(1):A11. PMID:15670464
104. Paul R, Groza T, Hunter J, Zankl A. Inferring characteristic phenotypes via class association rule mining in the bone dysplasia domain. *J Biomed Inform* 2014;48:73-83. PMID:24333481
105. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016 Presented at: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; August 13-17, 2016; San Francisco, CA p. 1135-44. doi:10.1145/2939672.2939778
106. Patton MQ. *Qualitative Research & Evaluation Methods*, 3rd ed. Thousand Oaks, CA: SAGE Publications; 2001. ISBN:0761919716
107. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 1989;13(3):319-40. doi:10.2307/249008
108. Davis FD, Venkatesh V. Toward preprototype user acceptance testing of new information systems: implications for software project management. *IEEE Trans Engineering Management* 2004;51(1):31-46. doi:10.1109/TEM.2003.822468
109. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies* 1993;38(3):475-87. doi:10.1006/imms.1993.1022
110. Adams DA, Nelson RR, Todd PA. Perceived usefulness, ease of use, and usage of information technology: a replication. *MIS Quarterly* 1992;16(2):227-47. doi:10.2307/249577
111. Hendrickson AR, Massey PD, Cronan TP. On the test-retest reliability of perceived usefulness and perceived ease of use scales. *MIS Quarterly* 1993;17(2):227-30. doi:10.2307/249803
112. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation* 2006;27(2):237-46. doi:10.1177/1098214005283748

113. ATLAS.ti qualitative analysis software. <http://www.atlasti.com/index.html>. Archived at <http://www.webcitation.org/6oOjqY7rj>
114. Gupta P, Malhotra P, Vig L, Shroff G. Transfer learning for clinical time series analysis using recurrent neural networks. In: Proceedings of the KDD Workshop on Machine Learning for Medicine and Healthcare. 2018 Presented at: KDD Workshop on Machine Learning for Medicine and Healthcare; August 20, 2018; London, United Kingdom p. 1-4.
115. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong ICK, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li Y, Stang PE, Madigan D, Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015;216:574-8. PMID:26262116
116. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012;19(1):54-60. PMID:22037893
117. Sun Y, Heng BH, Tay SY, Seow E. Predicting hospital admissions at emergency department triage using routine administrative data. *Acad Emerg Med* 2011;18(8):844-50. PMID:21843220
118. Chariatte V, Berchtold A, Akre C, Michaud PA, Suris JC. Missed appointments in an outpatient clinic for adolescents, an approach to predict the risk of missing. *J Adolesc Health* 2008;43(1):38-45. PMID:18565436
119. Jackson JL, Chamberlin J, Kroenke K. Predictors of patient satisfaction. *Soc Sci Med* 2001;52(4):609-20. PMID:11206657
120. Saha Y, Singal R. Predicting patient adherence: why and how. <http://www.cognizant.com/InsightsWhitepapers/Predicting-Patient-Adherence-Why-and-How.pdf>. Archived at <http://www.webcitation.org/6agaJvbtY>