Using Computational Methods to Improve Integrated Disease Management for Asthma and Chronic Obstructive Pulmonary Disease: Protocol for a Secondary Analysis

Gang Luo1, PhD; Bryan L Stone2, MD, MS; Xiaoming Sheng3, PhD; Shan He4, PhD; Corinna Koebnick5, PhD; 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
3College of Nursing, University of Utah, 10 South 2000 East, Salt Lake City, UT 84112, USA
4Care Transformation and Information Systems, Intermountain Healthcare, Parkway Building, 3930 Parkway Boulevard, West Valley City, UT 84120, USA
5Department of Research & Evaluation, Kaiser Permanente Southern California, 100 S. Los Robles, 6th Floor, Pasadena, CA 91101, USA

luogang@uw.edu, bryan.stone@hsc.utah.edu, xiaoming.sheng@utah.edu, shan.he@imail.org, corinna.koebnick@kp.org, 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: Asthma and chronic obstructive pulmonary disease (COPD) both put a heavy burden on healthcare. About 1/4 of asthma and COPD patients are prone to exacerbations, which can be greatly reduced by preventive care via integrated disease management that has a limited service capacity. To do this well requires a predictive model for exacerbation-proneness, but no such model exists. It would be suboptimal to build such models using the current model building approach for asthma and COPD, which has 2 gaps due to rarely factoring in temporal features showing early health changes and general directions. 1) Existing models for other asthma and COPD outcomes rarely use more advanced temporal features, such as the slope of the number of days to albuterol refill, and are inaccurate. 2) Existing models seldom show the reason a patient is deemed high-risk and potential interventions to cut the risk, making already occupied clinicians expend more time on chart review and overlook suitable interventions. Regular automatic explanation methods cannot deal with temporal data and well address this issue.

Objective: To let more asthma and COPD patients obtain suitable and timely care to avert exacerbations, we will implement comprehensible computational methods to accurately predict exacerbation-proneness and recommend customized interventions.

Methods: We will: a) use temporal features to accurately predict exacerbation-proneness; b) automatically find modifiable temporal risk factors for every high-risk patient; c) assess actionable warning’s impact on clinicians’ decisions of using integrated disease management to prevent exacerbation-proneness.

Results: We have obtained most of the clinical and administrative data of asthma patients from 3 prominent American healthcare systems. We are retrieving the other clinical and administrative data, mostly of COPD patients, needed for the study. We intend to complete the study in 6 years.

Conclusions: Our results will help make asthma and COPD care more proactive, effective, and efficient, improving outcomes and saving resources.

Keywords: asthma; chronic obstructive pulmonary disease; decision support techniques; forecasting; machine learning

Introduction

The gap in identifying exacerbation-prone asthma and COPD patients for preventive care

In the U.S., 9.6% of children and 8% of adults have asthma, leading to 1.8 million emergency department visits, 493,000 inpatient stays, US $56 billion in cost, and 3,630 deaths every year [1-4]. About 6.5% of adults have chronic obstructive pulmonary disease (COPD), the #3 cause of death leading to 1.5 million emergency department visits, 0.7 million inpatient stays, and US $32 billion in cost every year [5]. A main goal in managing asthma and COPD patients is to reduce exacerbations, which expend ~40%-75% of their total care cost [6-8] and drive their lung function decline [9]. About 1/4 of asthma and COPD patients are exacerbation-prone [10-14], meaning that a patient has in a year ≥2 systemic corticosteroid orders, or ≥1 emergency department visit or inpatient stay for asthma or COPD with systemic corticosteroid treatment (Figure 1) [10,13,15]. These patients incur ~2/3 of all exacerbations [12,13,16] and suffer from low quality of life; sleep disturbance; limitations of daily activities impacting independence, relationships, family life, socialization, and career; anxiety; distress; missed work with lost earnings; missed school; high care costs; high hospital use; intubation; and death [10,17-19]. Even brief use of systemic corticosteroids to treat exacerbations can greatly raise risk of venous thromboembolism, sepsis, and fracture [20,21].

Figure 1. Determining when an asthma or chronic obstructive pulmonary disease (COPD) patient becomes exacerbation-prone.

Many healthcare systems and health plans use predictive models as the best method [22] to find high-risk patients for preventive care to improve outcomes and save resources [23-25]. For instance, this is the case with the health plans in 9 of 12 American metropolitan communities mentioned in Mays et al. [26]. Yet, no model exists to predict exacerbation-proneness, which only partly correlates with disease severity [16]. Exacerbation-prone patients are currently identified after exacerbations occur, making it too late to do integrated disease management (IDM). IDM is defined as “a group of coherent interventions,
designed to prevent or manage 1 or more chronic conditions using a community wide, systematic and structured multidisciplinary approach potentially employing multiple treatment modalities.” [27] IDM typically has several components, such as self-management education, skills training, care management, and structured follow-up [28,29]. Having a limited service capacity [29-33], IDM can lower hospital use by up to 40%, cut cost by up to 31%, greatly reduce symptoms, and enhance treatment adherence, patient satisfaction, and quality of life by 30-60% [26,28-32,34-42]. Neither patient registries nor dashboards are able to identify exacerbation-prone patients before exacerbations occur, and thus to apply IDM in a timely manner. A patient registry tracks a given patient cohort, but cannot do predictions. Although many attributes are often needed to gain high prediction accuracy [43-45], a dashboard tracks only a few attributes. To have prediction capability, a dashboard needs to be supported by a predictive model in the backend. To guide IDM’s use and prevent exacerbations, models for exacerbation-proneness are needed. This cannot be well done with the current model building approach for other asthma and COPD outcomes, which has 2 major gaps due to limited use of temporal features showing early health changes and general directions [46-94]. Each temporal feature is an independent variable computed on 1 or more longitudinal attributes, like the slope of pulmonary function last year, the slope of body mass index last year, the number of days in the previous week during which SO2 level was ≥4 parts per million, and whether the patient’s filling frequency of oral corticosteroid prescription rose over time. Although this study focuses on exacerbation-prone asthma and COPD as use cases, the proposed computing techniques and software can be harnessed to forecast outcomes of other diseases like congestive heart failure and diabetes, with temporal features like the slopes of cardiac function and blood glucose level over time.

Gap 1: Low prediction accuracy
Existing models for predicting an individual asthma or COPD patient’s health outcome typically have low accuracy [46-94], Loymans et al.’s systematic review [52] and our review [43] showed that for forecasting hospital use (emergency department visits and inpatient stays) for asthma in asthma patients, each prior model excluding Zein et al.’s models [58] has an area under the receiver operating characteristic curve (AUC) within 0.61-0.81, a sensitivity within 25%-49%, and a positive predictive value within 4%-22% [46-57]. Both Zein et al.’s models [58] and our recent new models [43-45] have similarly higher accuracy, but are still not good enough for well aligning preventive care with the patients needing it the most. The case with COPD is similar [59-94].

Existing models for predicting asthma and COPD outcomes typically have low accuracy for several reasons:
1) Existing models use elementary temporal features like the count of inpatient stays and ever intubated last year, but rarely use more advanced temporal features like the slope of the number of days to albuterol refill showing general directions. Many highly predictive temporal features are yet to be found or unused. In 2018, Google used all of the attributes in the electronic medical record along with long short-term memory (LSTM) [95,96], 1 kind of deep neural network, to discover temporal features automatically from longitudinal data [97]. This raised the AUC by ~+10% for projecting each of long hospital stay, in-hospital mortality, and unanticipated readmissions in 30 days [97]. Several other studies [98-100] obtained similar results for various clinical prediction tasks. This matches recent progress in areas like video classification, speech recognition, and natural language processing, where temporal features LSTM automatically discovered from data beat those that experts provided or other temporal and sequential pattern mining methods [101-104] mined from data. Xiang et al.’s LSTM model for predicting asthma outcome [57] had a low AUC of 0.7 because it used only 3 types of attributes and mostly inpatient data without much outpatient data, not because LSTM is ineffective.
2) Although >100 potential risk factors for poor outcomes in asthma and COPD are known [50-52,105-112], a typical prior model uses only a few (e.g., ≤17) [46-57,59-93]. None of the published models adopts all established risk factors contained in contemporary electronic medical records [113].
3) Weather and air quality variables impact asthma and COPD outcomes [114-117], but are seldom employed in existing models.

Gap 2: No information given on the reason a patient is deemed high-risk and potential interventions to cut the risk
To do preventive care well, clinicians need to know the reason a patient is deemed high-risk and potential interventions to cut the risk. Sophisticated predictive models, including the bulk of machine learning models like LSTM, are black boxes and give no such information, although explanation is critical for users’ acceptance, satisfaction, trust, and decision correctness [118-121]. Often, a patient’s clinical records include numerous variables on many pages recorded over multiple years [122]. As the model gives no explanation, already occupied clinicians need to expend extra time on chart review to identify the reasons. This is hard and time-consuming. In fact, the black box issue has been a major reason for the slow adoption of machine learning in clinical practice, despite machine learning often produces the highest prediction accuracy among all predictive modeling methods [33,123-127].

A clinician could develop a care plan using subjective, variable clinical judgment. But, this care plan often misses some suitable interventions because:
1) Big practice variation, frequently by 1.6-5.6 times, shows up across facilities, clinicians, and regions [128-135].
A patient can become high-risk for many reasons, each shown by a risk pattern given by a feature combination, e.g., the SO$_2$ level was ≥4 parts per million for ≥4 days in the previous week AND the number of days to albuterol refill rose over 12 months. Many features and feature combinations exist. A clinician is a human, can typically process ≤9 information items at once [136], and can easily miss some key reasons for which the patient is high-risk. Outcomes can degrade if suitable interventions are unused. Regular automatic explanation methods [137-140] cannot deal with longitudinal data and well address this issue.

Our proposed solutions
To let more asthma and COPD patients obtain suitable and timely care to prevent exacerbations, we will 1) employ temporal features to develop the first set of models to accurately predict exacerbation-prone asthma and COPD; 2) automate finding modifiable temporal risk factors for every high-risk patient; and 3) assess actionable warning’s impact on clinicians’ decisions of using IDM to prevent exacerbation-proneness.

Innovation
We will develop new techniques to automate extracting temporal features from and explaining machine learning predictions on longitudinal data. We will improve preventive care, notably for asthma and COPD, by steering it to the patients who need it more precisely and more timely than current risk modeling methods:

a) To the best of our knowledge, this study will construct the first set of models to predict which asthma and COPD patients will be exacerbation-prone. Currently, these patients are found after exacerbations occur, making it too late to do IDM. This is a major public health issue [29,31,32]. Our models can improve IDM and guide its use to avert exacerbations. Compared with the current model building method for other asthma and COPD outcomes that often produces low accuracy, our model building method will lead to more accurate predictions.

b) To the best of our knowledge, this will be the first study to extract comprehensible and predictive temporal features semi-automatically from longitudinal data without needing any manually pre-specified pattern template required by many sequential and temporal pattern mining methods [102-104]. This helps raise model accuracy and cut the effort taken to construct clinically usable models. At present, clinicians usually have to manually find such features to construct such models. This is time-consuming and hard. Prior models for asthma and COPD rarely use more advanced temporal features like slope [46-94]. Also, although current deep neural network methods can automatically discover temporal features, the discovered features are hidden in neurons and often incomprehensible, making it hard to explain the predictions [137,138].

c) To the best of our knowledge, this will be the first study to automate giving rule-formed explanations for machine learning predictions straightly on longitudinal data. Clinicians need explanations to understand the predictions and decide IDM enrollment and interventions. Rule-formed explanations are easier to comprehend and can better hint actionable interventions than other forms of automatic explanations. Most automatic explanation methods [137,138] for machine learning predictions cannot deal with longitudinal data. Our prior automatic explanation method [140-142] is no exception. It has 5 hyper-parameters whose effective values vary by modeling problem and data set. A computing expert often needs several months to do many trials to laboriously find these values for a data set. We will improve our prior method to deal with longitudinal data and automatically and efficiently select hyper-parameter values, so healthcare researchers with limited computing expertise can use our method with low overhead.

d) To the best of our knowledge, this will be the first study to automate finding modifiable temporal risk factors and recommending interventions on the basis of objective data, making IDM more efficient and effective. At present, clinicians rely on subjective, variable judgment to create care plans manually and overlook some suitable interventions for patients at high risk.

e) To the best of our knowledge, this will be the first study to assess actionable warning’s impact on clinicians’ decisions of using IDM to prevent exacerbation-proneness.

Methods
Computing resources
We will do all of the experiments on a password-protected and encrypted computer cluster hosted at the University of Washington Medicine (UWM). With appropriate authorization and using their university computers, all research team members and test participants at UWM can remotely access this computer cluster.

Data sets
All of the data this study will use are structured. We will use clinical and administrative data stored in the enterprise data warehouses of 3 prominent American healthcare systems: UWM, Kaiser Permanente Southern California (KPSC), and Intermountain Healthcare (IH). We will use >200 clinical and administrative variables listed in our papers’ [43-45] appendices, with differing names of the same concept in distinct electronic medical record systems already manually matched by us. These
variables cover a wide range of aspects, such as patient demographics, encounters, medications, laboratory tests, diagnoses, procedures, vital signs, and allergies. We can form temporal features of most variables, which are longitudinal with timestamps.

In Utah, IH is the largest healthcare system owning 24 hospitals and 215 clinics. As in our prior work on asthma outcome prediction [43-45], an IH data analyst will run Oracle database queries to retrieve a de-identified IH data set (e.g., shift dates, replace identifiers and ages≥90); and use Secure Shell (SSH) to encrypt it and transfer it to the password-protected and encrypted computer cluster, where we will do analysis. The IH data set covers patient encounters in 2005-2020. For the previous 5 years, data for children cover >5,000 pediatric asthma patients (age<18) per year. Data for adults cover >14,000 adult asthma patients (age≥18) and >6,000 adult COPD patients per year. IH expends many resources on data integrity and accuracy. Because of its large size and variable richness [143], the data set offers us many advantages to explore the proposed methods.

UWM and KPSC have similar strengths. In Washington, UWM is the largest academic healthcare system owning 4 hospitals and 12 clinics for adults. A UWM data analyst will execute SQL Server database queries to retrieve a de-identified UWM data set (e.g., shift dates, replace identifiers and ages≥90); and use SSH to encrypt it and transfer it to the password-protected and encrypted computer cluster. The UWM data set covers adult patient encounters in 2011-2020. For the previous 5 years, data cover >12,000 adult asthma patients and >5,000 adult COPD patients per year.

In Southern California, KPSC is the largest integrated healthcare system owning 15 hospitals and 231 clinics [144]. A KPSC data analyst will run database queries to retrieve a de-identified KPSC data set (e.g., shift dates, replace identifiers and ages≥90); and use SSH to encrypt it and transfer it to the password-protected and encrypted computer cluster. The KPSC data set covers patient encounters in 2009-2020. For the previous 5 years, data for children cover >77,000 pediatric asthma patients per year. Data for adults cover >172,000 adult asthma patients and >78,000 adult COPD patients per year.

Besides the clinical and administrative data, we will adopt 11 weather and air quality variables we have downloaded from public sources [145,146]: daily mean PM₂.₅, daily maximum 8-hour CO, daily mean PM₁₀, daily maximum 8-hour O₃, daily maximum 1-hour NO₂, daily maximum 1-hour SO₂, hourly mean precipitation, hourly mean relative humidity, hourly mean wind speed, hourly mean temperature, and hourly mean dew point. These variables were recorded over 16 years (2005-2020) by monitoring stations residing in the areas covered by IH, UWM, and KPSC.

The following discussion focuses on asthma. Whenever we refer to asthma, the same applies to COPD.

Aim 1: Employ temporal features to predict exacerbation-prone asthma and COPD accurately.

We will extract comprehensible and predictive temporal features semi-automatically from patient, weather, and air quality data, and construct models to predict exacerbation-proneness. Each feature uses ≥1 raw variable. There is an almost infinite number of possible features. Traits of a pediatric patient’s parents and other factors could also impact patient outcomes. Our goal is not to test all possible useful features and obtain the theoretically maximum possible prediction accuracy. Instead, we intend to show that temporal features can be used to improve prediction accuracy and IDM. We will create a separate model for every disease and healthcare system pair. This study will focus on associations, as is sufficient for decision support for IDM and common with predictive modeling.

Data pre-processing

We will convert all data sets into the Observational Medical Outcomes Partnership (OMOP) common data model format [147] and its linked standardized terminologies [148]. Much of UWM data is already in this format. IH and KPSC have put their data into an internal normalized format similar to this one. We will expand the data model to embrace patient, weather, and air quality variables that the original data model misses, but exist in our data sets. We will employ the method described in our paper [149] to choose the most pertinent laboratory tests. To cut the number of features, we will use the AHRQ Clinical Classifications Software (CCS) system [150,151] to merge diseases, use the Berenson-Eggers Type of Service (BETOS) system [152] to merge procedures, and use the HIC/GC3 (HIC: Hierarchical Ingredient Code) system [153] to merge drugs. We will adopt the method used in our prior work [43-45] to identify, correct, or delete invalid values. To deal with missing values, we will test various imputation techniques [154,155], such as last observation carried forward, replacement with mean values, and replacement with median values, and use the technique that works the best.

Patient, weather, and air quality variables will be used. The patient variables will cover standard variables studied in the clinical predictive modeling literature [128,129,154], such as diagnoses, and >100 known potential risk factors for poor asthma outcomes listed in our papers [43-45,156]. One such risk factor is the frequency of nighttime awakening recorded on the validated Asthma Control Test questionnaire [157] in the electronic medical record system. For weather and air quality variables, we will do inverse distance weighting spatial interpolation [158] to compute their daily average values at the patient’s residence zip code from their values at local monitoring stations, as we and others did before for asthma outcome prediction [159-161].

Asthma and COPD cases and outcomes
We will implement and test our method using (i) pediatric asthma, (ii) adult asthma, and (iii) COPD. We will use our prior method [44] adapted from the literature [47,162,163] to identify asthma patients. We deem a patient to have asthma in a given year if the patient has ≥1 asthma diagnosis code (International Classification of Diseases, Ninth Revision [ICD-9] 493.x or International Classification of Diseases, Tenth Revision [ICD-10] J45/J46.x) in the year. The outcome is whether the patient became exacerbation-prone (i.e., had ≥2 systemic corticosteroid orders, or ≥1 emergency department visit or inpatient stay with a principal diagnosis of asthma and systemic corticosteroid treatment) in the following year [10,15].

We will use our prior method [164] adapted from the literature [165-168] to identify COPD patients. As Figure 2 shows, we deem a patient to have COPD if the patient is ≥40 and fulfills “any of these 4 conditions:

1) an outpatient visit diagnosis code of COPD (ICD-9: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; ICD-10: J42, J41.8, J44.x, J43.x) followed by ≥1 prescription of long-acting muscarinic antagonist (aclidinium, glycopyrrolate, tiotropium, and umeclidinium) within 6 months,

2) ≥1 emergency department or ≥2 outpatient visit diagnosis codes of COPD (ICD-9: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; ICD-10: J42, J41.8, J44.x, J43.x),

3) ≥1 inpatient stay discharge having a principal diagnosis code of COPD (ICD-9: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; ICD-10: J42, J41.8, J44.x, J43.x), and

4) ≥1 inpatient stay discharge having a principal diagnosis code of respiratory failure (ICD-9: 518.82, 518.81, 799.1, 518.84; ICD-10: J96.0x, J91.8, J96.9x, J96.2x, R09.2) and a secondary diagnosis code of acute COPD exacerbation (ICD-9: 491.22, 491.21, 493.22, 493.21; ICD-10: J44.1, J44.0).” [164]

The outcome is whether the patient became exacerbation-prone (i.e., had ≥2 systemic corticosteroid orders, or ≥1 emergency department visit or inpatient stay with a principal diagnosis of COPD and systemic corticosteroid treatment) in the following year [13].

Extracting temporal features

We will adopt the method described in our design paper [149] to extract comprehensible and predictive temporal features semi-automatically from longitudinal data. In Aim 1, we will employ the extracted features to construct the final predictive models. In Aim 2, we will use the extracted features to automate finding modifiable temporal risk factors for every high-risk patient. The main idea of our temporal feature extraction method is to build a so-called multi-component LSTM deep neural network model on longitudinal data, use a so-called exclusive group Lasso (least absolute shrinkage and selection operator) regularization method to restrict the number of attributes employed in each component LSTM network, and then do visualization to identify comprehensible temporal features from certain cell vector elements in each component LSTM network. The final step of using visualization to identify temporal features and providing their definitions involves humans and is semi-automatic. All of the other steps are automatic. Our temporal feature extraction method is general and works for many
clinical applications. It has never been implemented in computer code before. Also, some of its technical details are not given in our design paper [149]. In this study, we will fill in all of the missing technical details and code and test this method.

The final predictive models in Aim 1

We will employ the extracted temporal features like the slope of the number of days to albuterol refill to transform longitudinal data into tabular data, producing 1 column per temporal feature, and add static features. We will put no artificial upper or lower bound and use as many features as needed (likely 10s to 100s based on our prior experience [43-45]). Our data are relatively balanced [10-14]. We will harness Weka [169], a major open-source machine learning toolkit, to create the final models in Aim 1. As Aim 2 shows, these models suit for automatic explanation. Weka implements many classic machine learning algorithms and feature selection techniques. We will adopt supervised algorithms and our prior method [170] to automate selection of the machine learning algorithm, feature selection technique, and hyper-parameter values out of all applicable ones. When needed, we will perform fine-tuning manually.

We will employ past data up to the prediction time point to construct 5 sets of models, 1 set for each of 5 combinations: pediatric asthma at IH and KPSC, and adult asthma at IH, UWM, and KPSC. UWM has rather incomplete data on many of its patients partly because most of its patients are referred from elsewhere. To reduce the impact of incomplete data on model performance, we will harness our prior constraint-based method [164,171] to identify patients apt to get most of their care from UWM, and construct models for them. As mentioned earlier, we will also implement and test our method on COPD.

Evaluating model performance and power analysis

The discussion below focuses on IH data. The cases with UWM and KPSC data are analogous. Since we need to calculate outcomes in the following year, we effectively have 15 years of IH data over the prior 16 years. We will train and test models in a standard way. On the first 14 years’ data, we will do stratified 10-fold cross validation [169] to train models and gauge their performance. On the 15th year’s data, we will appraise the performance of the best models, reflecting future use in practice. We will employ the standard performance metric AUC [169] to choose the best model and record its AUC. We will show the model’s accuracy, sensitivity, specificity, and positive and negative predictive values when the cutoff threshold of binary classification varies from the top 1% to the top 50% of asthma patients with the highest predicted risk. To find the variables essential for reaching high model performance, backward elimination [154] will be adopted to remove features as long as AUC drops ≤0.002. We will compare the variables essential for reaching high model performance on IH data with those on UWM and KPSC data. Gender’s predictive power will be explicitly checked. We will use the variables appearing in both the UWM/KPSC and IH data sets to construct a best model on IH data, and compare its performance on UWM/KPSC data with that on IH data.

We will test the hypothesis that adopting our techniques could enhance model performance twice, once for adults and once for children. To do this, we will compare the AUCs of 2 predictive models built using the attributes in our data set and the best machine learning algorithm. The first model will harness all features essential for reaching high model performance. The second model will be done in the same way as our recent model for predicting hospital use for asthma [44] related to exacerbation-proneness. We anticipate the second model to have an AUC around our recent model’s AUC of 0.86. Our hypothesis is:

1) Null hypothesis: The second model has identical AUC as the first model.
2) Alternative hypothesis: The second model has a smaller AUC than the first model.

The categorical outcome variable of exacerbation-proneness has 2 values (classes). According to the standard method developed by Obuchowski and McClish for AUC-related sample size computation [172], using a 2-sided Z-test at a significance level of 0.05 and assuming for both classes a Pearson correlation coefficient of 0.6 between the 2 models’ predictions, a sample size of 464 instances per class provides 90% power to identify an AUC difference of 0.05 between the 2 models. The 15th year’s IH data cover >5,000 children and >14,000 adults with asthma, offering enough power to test our hypothesis. If the real correlation coefficient is different from the assumed one by no more than a moderate degree, the conclusion would remain valid.

Sensitivity analysis

IH, UWM, and KPSC all record a lot of variables. Another healthcare system could record fewer variables. We will test miscellaneous variable combinations and assess the performance of the corresponding modified models. This will help us ensure generalizability and identify critical variables. If a healthcare system does not record a particular critical variable, the assessed performance numbers can suggest alternative variables with minimal degrade of model performance.

Based on our clinical experts’ judgment, we will merge variables apt to co-occur, such as the variables appearing in a lab test panel, into groups. We will form and publish a table listing possible combinations of variables by groups, accompanied by the performance numbers and the trained parameters of the corresponding predictive models. A healthcare system interested in deploying the model can use the table to assess expected model performance for their data environment and determine variables
to be recorded. The table contains a distinct column for each of IH, UWM, and KPSC. A lot of variables recorded by IH, UWM, and KPSC and employed in this study are common and recorded by many other healthcare systems. Hence, these healthcare systems already have all of the variables appearing in each of many rows in the table.

**Aim 2: Automate finding modifiable temporal risk factors for every high-risk patient.**

For patients with predicted risk over a fixed bar like the 75th percentile, we will automate explaining warnings, finding modifiable temporal risk factors, and recommending customized interventions. This will help clinicians make decisions on IDM enrollment and develop customized care plans. To create the new function, we will enhance our prior method [140] of automatically explaining machine learning predictions with no loss of model performance. Our prior method cannot deal with longitudinal data, has hard-to-tune hyper-parameters, and has not been used on COPD or IDM before.

**Explanation method**

As Aim 1 shows, we will employ temporal features to transform longitudinal data into tabular data, producing 1 column per temporal feature. Our prior automatic explanation method [140] can then be used. Each patient is labeled either high risk or not high risk. Our method mines from past data association rules tied to high risk. One example rule is: the SO2 level was ≥4 parts per million for ≥4 days in the previous week AND the number of days to albuterol refill rose over the prior 12 months → the patient is high risk. The second item on the rule’s left hand side is a modifiable temporal risk factor. Three interventions for it are to 1) assess the patient on asthma triggers and make sure the patient avoids them; 2) evaluate compliance with asthma controller medications and prescribe, modify, or increase the doses of the medications if necessary; and 3) create a new asthma action plan to use more aggressive interventions when the patient is in the yellow zone [173]. Our paper [149] presented multiple interventions for several more temporal risk factors. Through discussion and consensus, our clinical team will examine the mined rules and remove those that make no or little clinical sense. For each rule left, our clinical team will identify the modifiable temporal risk factors in it and provide zero or more evidence-based interventions from the literature addressing the reason it gives. The rules are used to give explanations instead of predictions.

At prediction time, for each patient our most accurate model (initially resulting from Aim 1) marks high risk, we will identify and present all association rules tied to high risk and whose left hand side conditions are fulfilled by the patient, as well as show the rules’ linked interventions as our recommendations. Every rule presents a reason the patient is predicted high-risk. Users of the automatic explanation function could provide input to facilitate us identify and remove unreasonable rules [174].

Automatically and efficiently selecting hyper-parameter values

Our previous automatic explanation method [140-142] uses 5 hyper-parameters. Their effective values differ by modeling problem and data set. In our prior work [140-142], for each data set, a computing expert took several months to do many trials to laboriously find these values. To cut this overhead and to allow healthcare researchers with no extensive computing background to use our method, we will extend the progressive sampling-based approach, which we previously developed for expediting automatic machine learning model selection [170], to automatically and efficiently select the 5 hyper-parameters’ values. On average, our progressive sampling-based approach performs search 2 orders of magnitude faster than the modern Auto-Weka automatic selection approach [170,175]. Our approach generalizes to many clinical applications.

We will also develop our techniques on COPD.

**Aim 3: Assess actionable warning’s impact on clinicians’ decisions of using integrated disease management to prevent exacerbation-proneness.**

To prepare for future clinical use, in a UWM test setting, we will assess actionable warning’s impact on clinicians’ decisions of using IDM on asthma patients to prevent exacerbation-proneness, and UWM physicians’ (primary care doctors, pulmonologists, and allergists) and nurses’ subjective opinions of automatic explanations.

**Recruiting subjects**

As an UWM operational project, we are building asthma outcome prediction models and have access to ~700 physicians and ~1,700 nurses managing adult asthma patients. Through personal contact and advertising in their email lists, we will recruit 20 test participants (10 physicians and 10 nurses) with purposeful sampling to guarantee enough variability in their work experience [176]. Every test participant will offer consent before participation and be current on UWM’s policy training on information security and privacy. To protect privacy, every test participant will receive a pseudonym linking his or her responses. Upon task completion, each physician and each nurse will receive $2,300 and $1,200, respectively, as compensation for participation for ~20 hours’ work.

**Procedures**
Using the 15th year’s (2019) IH data, we will randomly select 800 IH adult asthma patients and automatically explain the predictions of the best performing IH model formed in Aim 1. Using patients outside of UWM can help ensure no test participant knows the outcome of any of these patients in the following year. We will present a distinct subset of 40 patients to each test participant and proceed in 4 steps.

(1) Step 1: For each patient, we will display to the test participant the 2005-2019 de-identified patient data in reverse chronological order like in the electronic medical records, and ask the test participant to write down the IDM enrollment decision (Y/N) and any interventions that the test participant plans to adopt on the patient.

(2) Step 2: For each patient, we will display to the test participant the 2005-2019 de-identified patient data, the prediction, the automatic explanations, and the interventions connected to them. We will ask the test participant to write down his or her IDM enrollment decision (Y/N) on the patient after seeing the prediction and the explanations; the linked interventions he or she agrees with; those he or she disagrees with; and the interventions that he or she comes up with in Step 1, but whose concepts are missed by the linked interventions.

(3) Step 3: Perceived usefulness closely links to future usage intentions as well as actual function usage [177,178]. Using the classic Technology Acceptance Model satisfaction questionnaire [179], we will survey the test participant to know his or her perceived ease of use and usefulness of automatic explanations.

(4) Step 4: We will do a focus group with 10 randomly chosen test participants to assess what helps them use or prevents them from using the automatic explanations in clinical practice, and why they agree or disagree with the automatically recommended interventions.

Quantitative and qualitative analyses

Quantitative analyses: We will give descriptive statistics on each quantitative outcome measure, including the mean and the standard deviation of each of the following: 1) the number of times that a test participant changes his or her IDM enrollment decision on a patient after seeing the prediction and the explanations; 2) the number of linked interventions for a patient a test participant agrees with; 3) the number of linked interventions for a patient a test participant disagrees with; 4) the number of interventions that a test participant comes up with for a patient in Step 1, but whose concepts are missed by the linked interventions; and 5) the rating of every item in the Technology Acceptance Model satisfaction questionnaire. We will test the hypothesis that giving actionable warnings will improve clinicians’ decisions of using IDM to prevent exacerbation-proneness, i.e., the degree of IDM enrollment decision matching whether the patient will become exacerbation-prone in the following year. Our hypothesis is:

1) Null hypothesis: The degree of IDM enrollment decision matching whether the patient will become exacerbation-prone in the following year in Step 2 is the same as that in Step 1.

2) Alternative hypothesis: The degree of IDM enrollment decision matching whether the patient will become exacerbation-prone in the following year in Step 2 is larger than that in Step 1.

We will fit a random effect logistic model that accounts for correlation among the outcomes related to the same test participant.

Power analysis: Assuming a modest intra-class correlation of 0.1 within the same test participant on the outcome, a sample size of 40 patients per test participant for the 20 test participants is equivalent to a total of 82 independent patients after factoring in the clustering effect. We will have, at a 2-sided significance level of 0.05, 80% power to detect a 9.7% rise in the chances of improving clinicians’ decisions of using IDM with actionable warnings. If the real correlation is different from the assumed one by no more than a moderate degree, a similar conclusion would hold.

Qualitative analyses: Using the inductive method in Patton et al. [176,180], test participants’ comments recorded in text during the focus group will be loaded into ATLAS.ti qualitative analysis software [181]. Three people from our research team will highlight quotations independently. Through discussion and negotiated consensus in multiple iterations, these people will review quotations, categorize quotations into pre-codes, merge codes into categories, and synthesize categories to identify general themes.

Exploring for other diseases

Preventive care is also widely adopted for patients with heart diseases and diabetes. To explore what will be needed to generalize our techniques to predict outcomes of these diseases in the future, we will do 2 phases of focus groups, each phase with a distinct set of 6 UWM clinical experts on these diseases, and add more phases if not reaching saturation.

As stated immediately before Aim 1, the discussion above concentrates on asthma. Whenever we refer to asthma, the same applies to COPD and will be implemented and tested on COPD also in Aims 1 and 2, but not in Aim 3.

Ethics Approval

We have received from the UWM institutional review board approval for this study and are applying for approvals from IH and KPSC.
Results
We have downloaded 2005-2020 weather and air quality data from public sources [145,146]. For the clinical and administrative data, Dr. Luo at UWM has obtained the 2005-2018 asthma patient data from IH [44], the 2009-2018 asthma patient data from KPSC [45], and the 2011-2018 asthma patient data from UWM [43]. We are retrieving from IH, UWM, and KPSC the other clinical and administrative data, mostly of COPD patients, needed for the study. We intend to complete the study in 6 years.

Discussion
Using Our Results in Clinical Practice
IH, UWM, KPSC, and many other places do IDM and use inaccurate predictive models with AUC <0.8 and sensitivity ≤49% for preventive care via care management [22,24-26,46-57,59-93]. In a way similar to our recent work of using IH, UWM, and KPSC data to greatly raise prediction accuracy for hospital use for asthma [43-45] related to exacerbation-proneness, we expect our models predicting exacerbation-proneness to be more accurate than those models, benefit many patients, and have practical value. We will automate explaining warnings and recommending interventions to aid clinicians to review structured data in patient clinical records faster and create customized care plans based on objective data. After our methods find patients with the greatest predicted risks and offer explanations, clinicians will review patient clinical records, look at factors like social dimensions [182], and make IDM enrollment and intervention decisions. As feature patterns linked to high risk and patient status keep changing, our techniques can be used continuously to move patients out of and into IDM and to discover new feature patterns.

Besides making the predictive model more accurate, using temporal features showing early health changes and general directions could also boost warning timeliness. If a patient will be admitted into the hospital for COPD or asthma, but the model would not predict this until 1 week before the hospital admission, intervening at that time could be too late to avoid the admission. If the model uses suitable temporal features and runs continuously, this patient could be found several weeks earlier, when health decline just begins and preventing hospital admission is likely.

Generalizability
Predictive models vary by diseases and other factors and could be dissimilar to each other. Yet, our proposed methods and software for extracting temporal features and automatically explaining machine learning predictions are general with no reliance on any special property of a specific healthcare system, disease, or patient cohort. Given a new data set with a different disease, set of variables, patient cohort, or prediction target, one can plug in our software to extract temporal features and to automatically explain machine learning predictions. Besides being used for asthma and COPD patients, preventive care is also widely adopted for heart disease and diabetes patients [128], where our techniques could be harnessed, e.g., to predict hospital use. Our sensitivity analysis results in Aim 1 can be used to identify critical variables and figure out how to generalize a predictive model to a healthcare system recording a different set of variables from IH, UWM, and KPSC.

We will employ data retrieved from 3 healthcare systems UWM, IH, and KPSC to demonstrate our techniques on asthma and COPD patients. These systems include an academic system that has most of its patients referred from elsewhere (UWM), 2 integrated systems (IH and KPSC), and 42 hospitals and 458 clinics. Spreading across 3 large geographic areas, these heterogeneous facilities range from tertiary care hospitals in large cities served by sub-specialists to community rural and urban clinics served by general practitioners and family physicians with limited resources. These healthcare systems use 4 distinct electronic medical record systems: KPSC uses Epic; UWM uses Epic and Cerner; IH uses HELP, HELP2, and Cerner. Variation in healthcare system type, patient population, geographic location, cultural background, staff composition, electronic medical record system, and scope of services enables us to identify factors that generalize to other facilities nationwide. The OMOP common data model [147] and its linked standardized terminologies [148] standardize administrative and clinical variables from ≥10 major American healthcare systems [183,184]. Our models will be based on OMOP and apply to these healthcare systems using OMOP.

With appropriate extension, our techniques can be adopted for miscellaneous diseases and decision support applications and improve clinical machine learning. For example, using our techniques can enhance prediction accuracy of other outcomes such as no show [185], hospital use [186], and treatment adherence [187]. This will facilitate us to target resources, such as telephone reminders to cut no shows [185], home visits by nurses and care management to cut hospital use [186], and interventions to boost treatment adherence [187].

We can use the features extracted by our temporal feature extraction method to create a feature library to ease feature reuse [188]. This will help cut the effort taken to create predictive models for other modeling projects.

Significance Thresholds
In both the “Evaluating model performance and power analysis” and “Quantitative and qualitative analyses” sections, we use the widely adopted significance level of 0.05 to do power analysis. The statistics community has debated a lot about \( p \) value and its dichotomization [189-191]. Setting a threshold for \( p \) value is essential for power analysis and sample size estimation [189]. Also, to the best of our knowledge, no consensus has been reached on what the best alternative is if \( p \) values and statistical significance are not used [189]. Following the advice given by Amrhein et al. [191], after obtaining this study’s research results, we will report the actual \( p \) values, treat them as continuous measures of evidence against the null hypotheses rather than as parts of binary decision rules, and acknowledge that multiple independent studies are needed to provide stronger support for or against our hypotheses.

Conclusion

Our results will help make IDM for asthma and COPD more proactive, effective, and efficient, improving outcomes and saving resources. Future studies will evaluate our methods on heart diseases, diabetes, and other diseases, deploy our methods at UWM, KPSC, and IH for IDM for asthma and COPD, and test the performance against current IDM practice.

Acknowledgments

We thank Peter Tarczy-Hornoch and Siyang Zeng for helpful discussions. GL, BS, XS, SH, CK, and FN 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. FN offered feedback on study design and medical issues, participated in doing the literature review, and revised the paper. BS offered feedback on study design and medical issues and revised the paper. XS took part in conceptualizing and writing the statistical analysis sections. CK took part in retrieving the Kaiser Permanente Southern California asthma patient data set and interpreting its detected peculiarities. SH took part in retrieving the Intermountain Healthcare data set and interpreting its detected peculiarities. 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
- ICD-9: International Classification of Diseases, Ninth Revision
- ICD-10: International Classification of Diseases, Tenth Revision
- IDM: integrated disease management
- IH: Intermountain Healthcare
- KPSC: Kaiser Permanente Southern California
- LSTM: long short-term memory
- OMOP: Observational Medical Outcomes Partnership
- SSH: Secure Shell
- UWM: University of Washington Medicine

References

17. Foster JM, McDonald VM, Guo M, Reddel HK. "I have lost in every facet of my life": the hidden burden of severe asthma. Eur Respir J 2017;50(3):1700765. PMID:28931662
49. 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):359-370. PMID:10386500
64. Yii ACA, Loh CH, Tiew PY, Xu H, Taha AAM, Koh J, Tan J, Lapperre TS, Anzueto A, Tee AKH. A clinical prediction model for hospitalized COPD exacerbations based on "treatable traits". Int J Chron Obstruct Pulmon Dis 2019;14:719-728. PMID:30988606


113. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2016;Suppl 1:S48-61. PMID:27199197


122. Halamka JD. Early experiences with big data at an academic medical center. Health Aff (Millwood) 2014;33(7):1132-1138. PMID:25006138
134. Gifford E, Foster EM. Provider-level effects on psychiatric inpatient length of stay for youth with mental health and substance abuse disorders. Med Care 2008;46(3):240-246. PMID:18388838
136. 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
141. Tong Y, Messinger AI, Luo G. Testing the generalizability of an automated method for explaining machine learning predictions on asthma patients' asthma hospital visits to an academic healthcare system. IEEE Access 2020;8:195971-195979. PMID:33240737
143. Evans RS, Lloyd JF, Pierce LA. Clinical use of an enterprise data warehouse. AMIA Annu Symp Proc 2012;189-198. PMID:23304288


191. Amrhein V, Korner-Nievergelt F, Roth T. The earth is flat (p > 0.05): significance thresholds and the crisis of unreplicable research. PeerJ 2017;5:e3544. PMID:28698825