Forecasting Future Asthma Hospital Encounters of Patients with Asthma in an Academic Health Care System: Predictive Model Development and Secondary Analysis Study

Yao Tong[1], MS; **Amanda I. Messinger**[2], MD; **Adam B. Wilcox**[1], PhD; **Sean D. Mooney**[1], PhD; **Giana H. Davidson**[3,4], MD, MPH; **Pradeep Suri**[5,6,7], MD, MS; **Gang Luo**[1], PhD

[1]Department of Biomedical Informatics and Medical Education, University of Washington, UW Medicine South Lake Union, 850 Republican Street, Building C, Box 358047, Seattle, WA 98109, USA

[2]The Breathing Institute, Department of Pediatrics, University of Colorado School of Medicine, Children's Hospital Colorado, East 16th Avenue, Box 395, Aurora, CO 80045, USA

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

[4]Department of Health Services, University of Washington, 1959 N.E. Pacific, Box 356410, Seattle, WA 98195, USA

[5]Seattle Epidemiologic Research and Information Center & Division of Rehabilitation Care Services, VA Puget Sound Health Care System, 1660 S Columbian Way, Seattle, WA 98108, USA

[6]Clinical Learning, Evidence, and Research (CLEAR) Center, University of Washington, 4333 Brooklyn Ave, Box 359455, Seattle, WA 98195, USA

[7]Department of Rehabilitation Medicine, University of Washington, Ninth & Jefferson Building, 325 Ninth Avenue, Box 359612, Seattle, WA 98104, USA

yaotong@uw.edu, amanda.messinger@childrenscolorado.org, abwilcox@uw.edu, sdmooney@uw.edu, ghd@uw.edu, pradeep.suri@va.gov, luogang@uw.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 98109, USA
Phone: 1-206-221-4596
Fax: 1-206-221-2671
Email: luogang@uw.edu

Abstract

Background: Asthma affects a large proportion of the population and leads to a lot of hospital encounters covering both hospitalizations and emergency department visits every year. To lower the number of such encounters, many healthcare systems and health plans deploy predictive models to prospectively find patients at high risk and offer them care management services for preventive care. Yet, the previous models do not have enough accuracy to serve this purpose well. Embracing the modeling strategy of examining many candidate features, we newly built a machine learning model to forecast asthmatic patients' future asthma hospital encounters at Intermountain Healthcare, a non-academic healthcare system. This model is more accurate than the previous published models. But, it is unclear how well our modeling strategy generalizes to academic healthcare systems, whose patient composition is different from Intermountain Healthcare's.

Objective: This study evaluates our modeling strategy's generalizability to University of Washington Medicine (UWM), an academic healthcare system.

Methods: All of the adult asthmatic patients who visited UWM facilities between 2011 and 2018 served as the patient cohort. We considered 234 candidate features. Through secondary analysis of 82,888 UWM data instances from 2011 to 2018, we built a machine learning model to forecast asthmatic patients' asthma hospital encounters in the subsequent 12 months.

Results: Our UWM model yielded an area under the receiver operating characteristic curve (AUC) of 0.902. When placing the cutoff point for making binary classification at the top 10.00% (1,464/14,644) of asthmatic patients with the biggest forecasted risk, our UWM model yielded an accuracy of 90.60% (13,268/14,644), a sensitivity of 70.18% (153/218), and a specificity of 90.91% (13,115/14,426).

Conclusions: Our modeling strategy showed excellent generalizability to UWM, leading to a model with an AUC that is higher than all of the AUCs previously reported in the literature for forecasting asthma hospital encounters. After further optimization, our model could be employed to facilitate efficient and effective allocation of asthma care management resources to improve outcomes.

International Registered Report Identifier (IRRID): PRR2-10.2196/5039

Keywords: Asthma; forecasting; machine learning; patient care management; risk factors

Introduction

Background

In the United States, 7.7% of people have asthma, which is responsible for 188,968 hospitalizations, 1,776,851 emergency department (ED) visits, and 3,441 deaths annually [1]. To reduce asthma hospital encounters covering both hospitalizations and ED visits, many healthcare systems and health plans deploy predictive models to prospectively find patients at high risk and offer them care management services for preventive care. University of Washington Medicine (UWM), Intermountain Healthcare, and Kaiser Permanente Northern California [2] are three examples of such healthcare systems. Examples of such health plans include those in nine of 12 metropolitan communities [3]. Once a patient is deemed high risk and enrolled in a care management program, a care manager will regularly assess the patient's asthma control, adjust the patient's asthma medications if necessary, and help the patient make appointments for health and related services. Using effective care management, as many as 40% of future hospital encounters by asthmatic patients can be avoided [4-7].

Due to its limited service capacity, a care management program normally enrolls at most 3% of patients with a particular condition [8]. To maximize the benefits of this resource intensive program, it is crucial for the program to enroll only the highest-risk patients. After all, the deployed predictive model's accuracy (or lack thereof) places an upper bound on the program's effectiveness. Before us, several research groups have built multiple models for forecasting asthmatic patients' future asthma hospital encounters. Every such model examined only a few features [2,9-22]. Overlooking some important features in the model degrades model accuracy, making the model miss at least half of the patients who will experience future asthma hospital encounters and incorrectly forecast future asthma hospital encounters for many other asthmatic patients. These errors result in impaired patient outcomes and wasted healthcare spending [23]. In non-medical fields, people frequently adopt the modeling strategy of examining many candidate features to enhance machine learning models' accuracy [24-27]. Embracing this modeling strategy for medical data, we newly built a machine learning model to forecast asthmatic patients' future asthma hospital encounters at Intermountain Healthcare, a non-academic healthcare system [23]. Our Intermountain Healthcare model raised the area under the receiver operating characteristic curve (AUC) to 0.859, which is higher than every previous published model's AUC by 0.049 or more. While this progress is encouraging, it is unclear how well our modeling strategy generalizes to academic healthcare systems, which normally care for more complex and sicker patients than non-academic healthcare systems [28].

Objective

This study evaluates our modeling strategy's generalizability to UWM, an academic healthcare system. Similar to our Intermountain Healthcare model [23], our UWM model employs clinical and administrative data to forecast asthmatic patients' future asthma hospital encounters covering both hospitalizations and ED visits. There are two possible values of the categorical dependent variable: whether the asthmatic patient will incur asthma hospital encounters in the subsequent 12 months or not. This paper reports the development and evaluation of our UWM model.

Our contributions

This study makes three innovative contributions:

1) We conducted the first evaluation of our modeling strategy's generalizability to an academic healthcare system.
2) We evaluated the predictive power of 71 new features, which were unused in our prior study [23], for forecasting asthma hospital encounters.
3) We evaluated the generalizability of our Intermountain Healthcare model to UWM, as well as the generalizability of our UWM model to Intermountain Healthcare. To the best of our knowledge, this is the first time that model generalizability has been evaluated in both directions. Previously, model generalizability was evaluated solely in one direction, by assessing the performance of a model built using one site's data on another site's data [17].

Methods

Study design and ethics approval

UWM's and Intermountain Healthcare's institutional review boards approved this secondary analysis study on clinical and administrative data.

Patient cohort

UWM is the biggest academic healthcare system in Washington State. Its enterprise data warehouse contains clinical and administrative data of 3 hospitals and 12 clinics for adults. Our patient cohort covered the adult asthmatic patients (age≥18) who visited any of these UWM facilities between 2011 and 2018. We defined a patient to have asthma in a specific year if the encounter billing database contained at least one asthma diagnosis code (International Classification of Diseases, Ninth Revision [ICD-9]: 493.0x, 493.1x, 493.8x, 493.9x; International Classification of Diseases, Tenth Revision [ICD-10]: J45.x) record on the patient in that year [10,29,30]. As the sole exclusion criterion, we eliminated those patients who passed away in that year.

Prediction target (a.k.a. the dependent variable)

The prediction target came from our previous study [23]. We defined an asthma hospital encounter as a hospitalization or an ED visit that has asthma as its principal diagnosis (ICD-9: 493.0x, 493.1x, 493.8x, 493.9x; ICD-10: J45.x). As Figure 1 shows, for each patient deemed to have asthma in a specific year, we used any asthma hospital encounter at UWM in the subsequent 12 months, i.e., the 12 months after the end of this year, as the outcome of interest. We adopted the patient's data by the end of this year to forecast the patient's outcome in the subsequent 12 months.

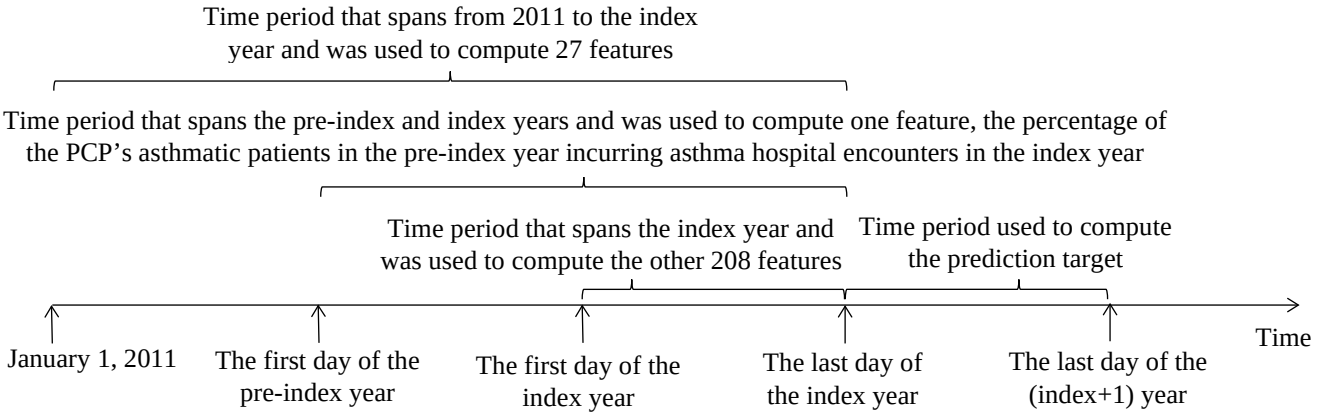

Figure 1. The time periods used to compute the features and the prediction target for an (index year, patient) pair.

Data set

The UWM enterprise data warehouse supplied a structured data set, which contained clinical and administrative data on our patient cohort's encounters at the 3 UWM hospitals and 12 UWM clinics between 2011 and 2019.

Features (a.k.a. independent variables)

Similar to what we did previously [23], we examined 234 candidate features describing a wide variety of characteristics. Table 1 of Multimedia Appendix 1 describes these features calculated on the structured attributes in our data set, with the 71 new features not used in our prior study [23] marked in italics. Throughout this paper, every mention of the number of a particular kind of items like medications counts multiplicity whenever the word *differing* is absent. For instance, consider a patient who was ordered medications twice in a given year. The first time medications 1 and 2 were ordered for the patient. The second time medications 2 and 3 were ordered for the patient. Then the total number of medications ordered for the patient in this year is four. The total number of differing medications ordered for the patient in this year is three.

Every input data instance to the predictive model addresses a unique (index year, patient) pair and is used to forecast the patient's outcome in the subsequent 12 months, i.e., the 12 months after the end of the index year. For that pair, we computed the patient's age and primary care provider (PCP) on the last day of the index year. The PCP identified was the patient's last PCP recorded in the electronic medical record system on or before the last day of the index year. As Figure 1 shows, adopting the data in the pre-index and index years, we computed one feature: the percentage of the PCP's asthmatic patients in the pre-index year incurring asthma hospital encounters in the index year. Using the data from 2011 to the index year, we computed 25 features: the number of years from the first encounter related to asthma in the data set, the number of years from the first encounter related to chronic obstructive pulmonary disease in the data set, family history of asthma, 15 features related to the problem list, and seven allergy features. We derived the other 208 features on the data in the index year.

Data analysis

Data preparation

Our UWM data set included peak expiratory flow values, which were absent in the Intermountain Healthcare data set adopted in our prior study [23]. Adopting the lower and upper bounds supplied by a clinical expert in our team, we deemed all peak expiratory flow values over 700 biologically implausible. Adopting the data preparation approach used in our prior paper [23] and this criterion, we pinpointed biologically implausible values, marked them missing, and normalized data. As the outcome of interest came from the subsequent year, our data set included eight years of effective data (2011-2018) over the nine-year period of 2011-2019. To be consistent with future model use in practice, we used the 2011-2017 data to train models and the 2018 data to evaluate model performance.

Performance metrics

Table 1. The confusion matrix.

Outcome class	Future asthma hospital encounters	No future asthma hospital encounter
Forecasted future asthma hospital encounters	True positive (TP)	False positive (FP)
Forecasted no future asthma hospital encounter	False negative (FN)	True negative (TN)

As presented in Table 1 and the formulas below, we evaluated model performance using six standard metrics: accuracy, AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

accuracy = (TP+TN)/(TP+TN+FP+FN),
sensitivity = TP/(TP+FN),
specificity = TN/(TN+FP),
positive predictive value = TP/(TP+FP),
negative predictive value = TN/(TN+FN).

We performed 1,000-fold bootstrap analysis [31] to calculate the six performance measures' 95% confidence intervals. For instance, we computed our final UWM model's performance measures on each bootstrap sample of the 2018 data. The 2.5^{th} and 97.5^{th} percentiles of the 1,000 values we obtained for every performance metric gave the corresponding performance measure's 95% confidence interval. We rendered the receiver operating characteristic curve to show the sensitivity-specificity tradeoff.

Classification algorithms

As in our prior paper [23], our predictive models were built using Weka Version 3.9 [32]. Weka is a core open-source software package for data mining and machine learning. It integrates a large number of popular feature selection techniques and machine learning algorithms. We checked the extreme gradient boosting (XGBoost) machine learning classification

algorithm [33] implemented in the software package XGBoost4J [34], as well as the 39 native classification algorithms in Weka listed in our previous paper's [23] online appendix. As an efficient and scalable realization of gradient boosting, XGBoost is a form of an ensemble of decision trees. Since XGBoost accepts only numerical features, we used one-hot encoding to transform categorical features to numerical features before giving them to XGBoost. We employed the 2011-2017 training data and the automatic machine learning model selection method developed in our prior work [35] to automatically select the feature selection technique, classification algorithm, data balancing method for handling imbalanced data, and hyper-parameter values among all of the pertinent ones. On average, our method can reduce model error rate by 11% and search time by 28 times than the modern Auto-WEKA automatic machine learning model selection method [35,36].

This study mainly evaluated our modeling strategy's generalizability to UWM via employing the UWM training set to train multiple models and then checking their performance on the UWM test set. In addition, we conducted two experiments to evaluate our models' generalizability across health systems.

Evaluating the generalizability of our Intermountain Healthcare model to UWM

In the first experiment, we evaluated the generalizability of our Intermountain Healthcare model to UWM. Previously, we developed both a simplified model and a full model on the Intermountain Healthcare data set [23]. Our simplified Intermountain Healthcare model uses the top 21 features whose importance values calculated by XGBoost on that data set are ≥0.01 [23]. Compared to our full Intermountain Healthcare model using 142 features, our simplified Intermountain Healthcare model retained nearly all of its predictive power. Our UWM data set contained the top 21 features and missed some other features adopted in our full Intermountain Healthcare model. We evaluated our simplified Intermountain Healthcare model's performance on the UWM test set twice. The first time, we retrained our simplified Intermountain Healthcare model on the UWM training set. The second time, we did no retraining and just directly applied our original simplified Intermountain Healthcare model trained on the Intermountain Healthcare training set.

Evaluating the generalizability of our UWM model to Intermountain Healthcare

In the second experiment, we evaluated the generalizability of our UWM model to Intermountain Healthcare. We employed a simplified UWM model, which used only the top features whose importance values calculated by XGBoost on the UWM training set were ≥0.01. For any top feature that was newly introduced in this study and was unused in our prior study [23], we computed the feature on the Intermountain Healthcare data set. We evaluated our simplified UWM model's performance on the Intermountain Healthcare test set twice. The first time, we retrained our simplified UWM model on the Intermountain Healthcare training set. The second time, we did no retraining and just directly applied our simplified UWM model trained on the UWM training set.

Results

Demographic and clinical characteristics of our patient cohort

Each data instance addresses a unique (index year, patient) pair. Table 2 and Table 3 show the demographic and clinical characteristics of our UWM patient cohort during 2011-2017 and 2018, respectively. The characteristics are similar across the two time periods. During 2011-2017 and 2018, 1.74% (1,184/68,244) and 1.49% (218/14,644) of data instances were linked to asthma hospital encounters in the subsequent 12 months, respectively.

Table 2. Demographic and clinical characteristics of the asthmatic patients at UWM during 2011-2017.

Characteristic	Data instances (*N*=68,244), *n* (%)	Data instances connecting to asthma hospital encounters in the subsequent 12 months (*N*=1,184), *n* (%)	Data instances connecting to no asthma hospital encounter in the subsequent 12 months (*N*=67,060), *n* (%)
Age			
<40	23,459 (34.38)	466 (39.36)	22,993 (34.29)
40 to 65	33,889 (49.66)	583 (49.24)	33,306 (49.67)
65+	10,896 (15.97)	135 (11.40)	10,761 (16.05)
Gender			
Male	24,198 (35.46)	551 (46.54)	23,647 (35.26)
Female	44,046 (64.54)	633 (53.46)	43,413 (64.74)
Race			
American Indian or Alaska native	1,358 (1.99)	32 (2.70)	1,326 (1.98)
Asian	5,721 (8.38)	96 (8.11)	5,625 (8.39)

Black or African American	8,420 (12.34)	520 (43.92)	7,900 (11.78)
Native Hawaiian or other Pacific islander	673 (0.99)	14 (1.18)	659 (0.98)
White	47,747 (69.97)	507 (42.82)	47,240 (70.44)
Unknown or not reported	4,325 (6.34)	15 (1.27)	4,310 (6.43)
Ethnicity			
Hispanic	3,526 (5.17)	82 (6.93)	3,444 (5.14)
Non-Hispanic	56,309 (82.51)	1,062 (89.70)	55,247 (82.38)
Unknown or not reported	8,409 (12.32)	40 (3.38)	8,369 (12.48)
Insurance			
Private	40,009 (58.63)	424 (35.81)	39,585 (59.03)
Public	28,787 (42.18)	756 (63.85)	28,031 (41.80)
Self-paid or charity	1,366 (2.00)	65 (5.49)	1,301 (1.94)
No. of years from the first encounter related to asthma in the data set			
≤3	60,873 (89.20)	986 (83.28)	59,887 (89.30)
>3	7,371 (10.80)	198 (16.72)	7,173 (10.70)
Asthma medication prescription			
Inhaled corticosteroid	28,889 (42.33)	626 (52.88)	28,263 (42.15)
Inhaled corticosteroid/long-acting beta-2 agonist combination	22,015 (32.26)	499 (42.15)	21,516 (32.08)
Leukotriene modifier	8,171 (11.97)	201 (16.98)	7,970 (11.88)
Long-acting beta-2 agonist	12,293 (18.01)	374 (31.59)	11,919 (17.77)
Mast cell stabilizer	47 (0.07)	4 (0.34)	43 (0.06)
Short-acting inhaled beta-2 agonist	47,808 (70.05)	1,010 (85.30)	46,798 (69.79)
Systemic corticosteroid	18,699 (27.40)	614 (51.86)	18,085 (26.97)
Comorbidity			
Allergic rhinitis	11,449 (16.78)	172 (14.53)	11,277 (16.82)
Anxiety or depression	19,885 (29.14)	372 (31.42)	19,513 (29.10)
Bronchopulmonary dysplasia	1 (0.00)	0 (0.00)	1 (0.00)
Chronic obstructive pulmonary disease	3,826 (5.61)	133 (11.23)	3,693 (5.51)
Cystic fibrosis	61 (0.09)	1 (0.08)	60 (0.09)
Eczema	3,891 (5.70)	66 (5.57)	3,825 (5.70)
Gastroesophageal reflux	12,291 (18.01)	238 (20.10)	12,053 (17.97)
Obesity	7,845 (11.50)	177 (14.95)	7,668 (11.43)
Sinusitis	7,261 (10.64)	89 (7.52)	7,172 (10.69)
Sleep apnea	4,556 (6.68)	88 (7.43)	4,468 (6.66)
Smoking status			
Current smoker	14,081 (20.63)	255 (21.54)	13,826 (20.62)
Former smoker	15,530 (22.76)	221 (18.67)	15,309 (22.83)
Never smoker or unknown	38,633 (56.61)	708 (59.80)	37,925 (56.55)

Table 3. Demographic and clinical characteristics of the asthmatic patients at UWM in 2018.

Characteristic	Data instances (*N*=14,644), *n* (%)	Data instances connecting to asthma hospital encounters in the subsequent 12 months (*N*=218), *n* (%)	Data instances connecting to no asthma hospital encounter in the subsequent 12 months (*N*=14,426), *n* (%)
Age			
<40	4,823 (32.94)	77 (35.32)	4,746 (32.90)
40 to 65	6,794 (46.39)	111 (50.92)	6,683 (46.33)
65+	3,027 (20.67)	30 (13.76)	2,997 (20.78)
Gender			
Male	5,238 (35.77)	100 (45.87)	5,138 (35.62)
Female	9,406 (64.23)	118 (54.13)	9,288 (64.38)
Race			
American Indian or Alaska native	281 (1.92)	8 (3.67)	273 (1.89)

Asian	1,325 (9.05)	18 (8.26)	1,307 (9.06)
Black or African American	1,570 (10.72)	79 (36.24)	1,491 (10.34)
Native Hawaiian or other Pacific islander	131 (0.89)	2 (0.92)	129 (0.89)
White	10,213 (69.74)	110 (50.46)	10,103 (70.03)
Unknown or not reported	1,124 (7.68)	1 (0.46)	1,123 (7.78)
Ethnicity			
Hispanic	850 (5.80)	20 (9.17)	830 (5.75)
Non-Hispanic	12,566 (85.81)	196 (89.91)	12,370 (85.75)
Unknown or not reported	1,228 (8.39)	2 (0.92)	1,226 (8.50)
Insurance			
Private	10,800 (73.75)	108 (49.54)	10,692 (74.12)
Public	8,023 (54.79)	182 (83.49)	7,841 (54.35)
Self-paid or charity	484 (3.31)	25 (11.47)	459 (3.18)
No. of years from the first encounter related to asthma in the data set			
≤3	10,566 (72.15)	124 (56.88)	10,442 (72.38)
>3	4,078 (27.85)	94 (43.12)	3,984 (27.62)
Asthma medication prescription			
Inhaled corticosteroid	6,177 (42.18)	108 (49.54)	6,069 (42.07)
Inhaled corticosteroid/long-acting beta-2 agonist combination	4,508 (30.78)	83 (38.07)	4,425 (30.67)
Leukotriene modifier	2,176 (14.86)	46 (21.10)	2,130 (14.77)
Long-acting beta-2 agonist	2,518 (17.19)	62 (28.44)	2,456 (17.02)
Mast cell stabilizer	14 (0.10)	1 (0.46)	13 (0.09)
Short-acting inhaled beta-2 agonist	9,704 (66.27)	164 (75.23)	9,540 (66.13)
Systemic corticosteroid	4,163 (28.43)	120 (55.05)	4,043 (28.03)
Comorbidity			
Allergic rhinitis	2,095 (14.31)	26 (11.93)	2,069 (14.34)
Anxiety or depression	4,346 (29.68)	62 (28.44)	4,284 (29.70)
Bronchopulmonary dysplasia	4 (0.03)	0 (0.00)	4 (0.03)
Chronic obstructive pulmonary disease	932 (6.36)	30 (13.76)	902 (6.25)
Cystic fibrosis	17 (0.12)	0 (0.00)	17 (0.12)
Eczema	743 (5.07)	11 (5.05)	732 (5.07)
Gastroesophageal reflux	2,657 (18.14)	46 (21.10)	2,611 (18.10)
Obesity	1,604 (10.95)	25 (11.47)	1,579 (10.95)
Sinusitis	1,372 (9.37)	15 (6.88)	1,357 (9.41)
Sleep apnea	1,499 (10.24)	24 (11.01)	1,475 (10.22)
Smoking status			
Current smoker	3,242 (22.14)	49 (22.48)	3,193 (22.13)
Former smoker	3,494 (23.86)	41 (18.81)	3,453 (23.94)
Never smoker or unknown	7,908 (54.00)	128 (58.72)	7,780 (53.93)

As the χ^2 two-sample test showed, for both the 2011-2017 and 2018 data, the data instances connecting to future asthma hospital encounters and those connecting to no future asthma hospital encounter exhibited the same distribution for anxiety or depression occurrence (P=.74 for the 2018 data and P=.09 for the 2011-2017 data), bronchopulmonary dysplasia occurrence (P=1.00), cystic fibrosis occurrence (P=1.00), eczema occurrence (P=1.00 for the 2018 data and P=.90 for the 2011-2017 data), gastroesophageal reflux occurrence (P=.29 for the 2018 data and P=.06 for the 2011-2017 data), and sleep apnea occurrence (P=.79 for the 2018 data and P=.32 for the 2011-2017 data). These two sets of data instances exhibited differing distributions for gender (P=.002 for the 2018 data and P<.001 for the 2011-2017 data), ethnicity (P<.001), insurance category (P<.001), race (P<.001), systemic corticosteroid prescription (P<.001), inhaled corticosteroid prescription (P=.02 for the 2018 data and P<.001 for the 2011-2017 data), inhaled corticosteroid/long-acting beta-2 agonist combination prescription (P=.02 for the 2018 data and P<.001 for the 2011-2017 data), short-acting inhaled beta-2 agonist prescription (P=.006 for the 2018 data and P<.001 for the 2011-2017 data), long-acting beta-2 agonist prescription (P<.001), leukotriene modifier prescription (P=.01 for the 2018 data and P<.001 for the 2011-2017 data), and chronic obstructive pulmonary disease occurrence (P<.001). For the 2011-2017 data, these two sets of data instances exhibited differing distributions for mast cell stabilizer prescription (P=.003), obesity

occurrence (*P*<.001), sinusitis occurrence (*P*<.001), allergic rhinitis occurrence (*P*=.04), and smoking status (*P*=.003). For the 2018 data, these two sets of data instances exhibited the same distribution for mast cell stabilizer prescription (*P*=.52), obesity occurrence (*P*=.89), sinusitis occurrence (*P*=.25), allergic rhinitis occurrence (*P*=.36), and smoking status (*P*=.19).

As the Cochran-Armitage trend test [37] showed, the data instances connecting to future asthma hospital encounters and those connecting to no future asthma hospital encounter exhibited the same distribution for age (*P*=.06) in the 2018 data and differing distributions for age (*P*<.001) in the 2011-2017 data. For both the 2018 and 2011-2017 data, these two sets of data instances exhibited differing distributions for the number of years from the first encounter related to asthma in the data set (*P*<.001).

Table 4 shows the number of asthmatic patients and the number of their visits in each year between 2011 and 2018.

Table 4. The number of asthmatic patients and the number of their visits in each year between 2011 and 2018.

Year	Number of asthmatic patients	Number of visits by asthmatic patients
2011	6,852	32,910
2012	7,768	40,730
2013	7,754	39,385
2014	9,785	58,953
2015	10,587	69,285
2016	12,072	78,605
2017	13,426	87,403
2018	14,644	94,875

Classification algorithm and features adopted by our final UWM model

Our automatic machine learning model selection method [35] selected the XGBoost classification algorithm [33]. XGBoost is a form of an ensemble of decision trees that can naturally deal with missing feature values. As described in Hastie *et al*. [38] in detail, XGBoost automatically calculates the importance value of each feature based on its apportioned contribution to the model. Our final UWM model was formed using XGBoost and 71 features displayed in descending order of their importance values in Table 2 of Multimedia Appendix 1. XGBoost automatically removed the other features because they had no additional predictive power.

Performance measures yielded by our final UWM model

On the UWM test set, our final model yielded an AUC of 0.902 (95% CI: 0.879-0.924). Figure 2 presents the model's receiver operating characteristic curve. Table 5 lists the model's performance measures when the cutoff point for making binary classification was placed at different top percentages of asthmatic patients having the biggest forecasted risk. When the cutoff point was placed at the top 10.00% (1,464/14,644), the model yielded an accuracy of 90.60% (13,268/14,644; 95% CI: 90.13-91.06), a sensitivity of 70.18% (153/218; 95% CI: 63.77-75.98), a specificity of 90.91% (13,115/14,426; 95% CI: 90.45-91.38), a PPV of 10.45% (153/1,464; 95% CI: 8.90-11.97), and an NPV of 99.51% (13,115/13,180; 95% CI: 99.39-99.62). Table 6 displays the model's confusion matrix in this case.

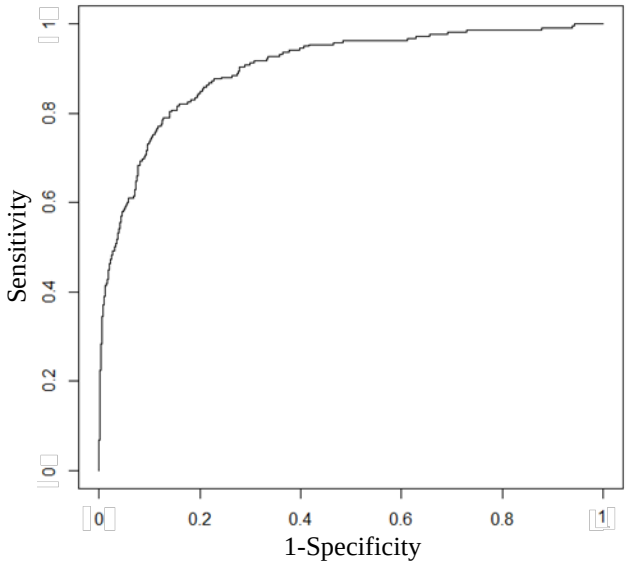

Figure 2. The receiver operating characteristic curve of our final UWM model.

Table 5. Our final UWM model's performance measures when the cutoff point for making binary classification was placed at different top percentages of asthmatic patients having the biggest forecasted risk.

Top percentage of asthmatic patients having the biggest forecasted risk (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
1	98.40 (14,410/14,644)	29.82 (65/218)	99.44 (14,345/14,426)	44.52 (65/146)	98.94 (14,345/14,498)
2	97.76 (14,316/14644)	41.74 (91/218)	98.61 (14,225/14,426)	31.16 (91/292)	99.12 (14,225/14,352)
3	96.92 (14,193/14,644)	47.25 (103/218)	97.67 (14,090/14,426)	23.46 (103/439)	99.19 (14,090/14,205)
4	96.02 (14,061/14,644)	50.46 (110/218)	96.71 (13,951/14,426)	18.80 (110/585)	99.23 (13,951/14,059)
5	95.17 (13,936/14,644)	55.50 (121/218)	95.76 (13,815/14,426)	16.53 (121/732)	99.30 (13,815/13,912)
6	94.28 (13,806/14,644)	59.17 (129/218)	94.81 (13,677/14,426)	14.69 (129/878)	99.35 (13,677/13,766)
7	93.33 (13,667/14,644)	61.01 (133/218)	93.82 (13,534/14,426)	12.98 (133/1,025)	99.38 (13,534/13,619)
8	92.39 (13,529/14,644)	62.84 (137/218)	92.83 (13,392/14,426)	11.70 (137/1,171)	99.40 (13,392/13,473)
9	91.58 (13,411/14,644)	69.27 (151/218)	91.92 (13,260/14,426)	11.47 (151/1,317)	99.50 (13,260/13,327)
10	90.60 (13,268/14,644)	70.18 (153/218)	90.91 (13,115/14,426)	10.45 (153/1,464)	99.51 (13,115/13,180)
15	85.88 (12,576/14,644)	79.36 (173/218)	85.98 (12,403/14,426)	7.88 (173/2,196)	99.64 (12,403/12,448)
20	80.99 (11,860/14,644)	83.03 (181/218)	80.96 (11,679/14,426)	6.18 (181/2,928)	99.68 (11,679/11,716)
25	76.12 (11,147/14,644)	87.61 (191/218)	75.95 (10,956/14,426)	5.22 (191/3,661)	99.75 (10,956/10,983)

Table 6. Our final UWM model's confusion matrix when the cutoff point for making binary classification was placed at the top 10.00% (1,464/14,644) of asthmatic patients having the biggest forecasted risk.

Outcome class	Future asthma hospital encounters	No future asthma hospital encounter

Forecasted future asthma hospital encounters	153	1,311
Forecasted no future asthma hospital encounter	65	13,115

Several features like family history of asthma were calculated on two or more years of data. When we dropped these features and checked solely those features calculated on one year of data, the model's AUC decreased from 0.902 to 0.899. If we employed only the top 17 features in Table 2 of Multimedia Appendix 1 whose importance values are ≥0.01 and ignored the other 217 features, the model's AUC decreased from 0.902 to 0.898 (95% CI: 0.874-0.919). In this case, when we placed the cutoff point for making binary classification at the top 10.00% (1,464/14,644) of asthmatic patients having the biggest forecasted risk, the model's accuracy decreased from 90.60% (13,268/14,644) to 90.59% (13,266/14,644; 95% CI: 90.11-91.06), sensitivity decreased from 70.18% (153/218) to 69.72% (152/218; 95% CI: 63.59-75.52), specificity remained at 90.91% (13,114/14,426; 95% CI: 90.42-91.37), PPV decreased from 10.45% (153/1,464) to 10.38% (152/1,464; 95% CI: 8.82-11.97), and NPV decreased from 99.51% (13,115/13,180) to 99.50% (13,114/13,180; 95% CI: 99.38-99.61).

Performance measures yielded by our simplified Intermountain Healthcare model on UWM data

When we did no retraining and applied our original simplified Intermountain Healthcare model trained on the Intermountain Healthcare training set [23] directly to the UWM test set, the model yielded an AUC of 0.861 (95% CI: 0.835-0.885). When we placed the cutoff point for making binary classification at the top 10.00% (1,464/14,644) of asthmatic patients having the biggest forecasted risk, the model yielded an accuracy of 90.29% (13,222/14,644; 95% CI: 89.81-90.77), a sensitivity of 59.63% (130/218; 95% CI: 53.39-65.68), a specificity of 90.75% (13,092/14,426; 95% CI: 90.28-91.20), a PPV of 8.88% (130/1,464; 95% CI: 7.46-10.34), and an NPV of 99.33% (13,092/13,180; 95% CI: 99.20-99.46).

After we used the UWM training set to retrain our simplified Intermountain Healthcare model [23], the retrained model yielded on the UWM test set an AUC of 0.874 (95% CI: 0.848-0.896). When we placed the cutoff point for making binary classification at the top 10.00% (1,464/14,644) of asthmatic patients having the biggest forecasted risk, the model yielded an accuracy of 90.34% (13,230/14,644; 95% CI: 89.85-90.80), a sensitivity of 61.47% (134/218; 95% CI: 54.63-67.66), a specificity of 90.78% (13,096/14,426; 95% CI: 90.32-91.23), a PPV of 9.15% (134/1,464; 95% CI: 7.62-10.66), and an NPV of 99.36% (13,096/13,180; 95% CI: 99.22-99.49).

Performance measures yielded by our simplified UWM model on Intermountain Healthcare data

Our simplified UWM model used only the top 17 features whose importance values are ≥0.01. When we did no retraining and applied our simplified UWM model trained on the UWM training set directly to the Intermountain Healthcare test set, the model yielded an AUC of 0.814 (95% CI: 0.798-0.830). When we placed the cutoff point for making binary classification at the top 10.00% (1,926/19,256) of asthmatic patients having the biggest forecasted risk, the model yielded an accuracy of 89.76% (17,285/19,256; 95% CI: 89.32-90.18), a sensitivity of 47.17% (383/812; 95% CI: 43.81-50.58), a specificity of 91.64% (16,902/18,444; 95% CI: 91.24-92.03), a PPV of 19.90% (383/1,925; 95% CI: 18.16-21.60), and an NPV of 97.52% (16,902/17,331; 95% CI: 97.28-97.75).

After we used the Intermountain Healthcare training set to retrain our simplified UWM model, the retrained model yielded on the Intermountain Healthcare test set an AUC of 0.846 (95% CI: 0.831-0.859). When we placed the cutoff point for making binary classification at the top 10.00% (1,926/19,256) of asthmatic patients having the biggest forecasted risk, the model yielded an accuracy of 90.11% (17,351/19,256; 95% CI: 89.64-90.56), a sensitivity of 51.23% (416/812; 95% CI: 47.55-54.49), a specificity of 91.82% (16,935/18,444; 95% CI: 91.43-92.21), a PPV of 21.62% (416/1,925; 95% CI: 19.81-23.41), and an NPV of 97.72% (16,935/17,331; 95% CI: 97.48-97.93).

Discussion

Principal results

We built a model on UWM data to forecast asthmatic patients' asthma hospital encounters in the subsequent 12 months. Table 7 reveals that our final UWM model yielded an AUC that is higher than the previously reported AUC of every existing model [2,9-23]. That is, our modeling strategy of examining many candidate features to enhance model accuracy showed excellent generalizability to UWM. After further optimization to boost its accuracy and to automatically give explanations of its predictions [39,40] to allow clinical interpretability, our UWM model could be employed to facilitate efficient and effective allocation of asthma care management resources to improve outcomes.

Table 7. A comparison of our final UWM model and several existing models for forecasting asthmatic patients' future hospitalizations and ED visits. "-" means that the initial paper showing the model did not give the performance measure.

Model	Prediction target	No. of data instances	No. of features	Classification algorithm	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC

			the model adopted						
Our final UWM model	Asthma hospital encounters	82,888	71	XGBoost	70.18	90.91	10.45	99.51	0.902
Our Intermountain Healthcare model [23]	Asthma hospital encounters	334,564	142	XGBoost	53.69	91.93	22.65	97.83	0.859
Loymans *et al*. [9]	Asthma exacerbation	611	7	Logistic regression	-	-	-	-	0.8
Schatz *et al*. [10]	Asthma-induced hospitalization in children	4,197	5	Logistic regression	43.9	89.8	5.6	99.1	0.781
Schatz *et al*. [10]	Asthma-induced hospitalization in adults	6,904	3	Logistic regression	44.9	87.0	3.9	99.3	0.712
Eisner *et al*. [11]	Asthma-induced hospitalization	2,858	1	Logistic regression	-	-	-	-	0.689
Eisner *et al*. [11]	Asthma-induced ED visit	2,415	3	Logistic regression	-	-	-	-	0.751
Sato *et al*. [12]	Severe asthma exacerbation	78	3	Classification and regression tree	-	-	-	-	0.625
Miller *et al*. [14]	Asthma hospital encounters	2,821	17	Logistic regression	-	-	-	-	0.81
Yurk *et al*. [16]	Lost day or hospital encounters for asthma	4,888	11	Logistic regression	77	63	82	56	0.78
Lieu *et al*. [2]	Asthma-induced hospitalization	16,520	7	Proportional-hazards regression	-	-	-	-	0.79
Lieu *et al*. [2]	Asthma-induced ED visit	16,520	7	Proportional-hazards regression	-	-	-	-	0.69
Lieu *et al*. [18]	Asthma hospital encounters	7,141	4	Classification and regression tree	49.0	83.6	18.5	-	-
Schatz *et al*. [19]	Asthma hospital encounters	14,893	4	Logistic regression	25.4	92.0	22.0	93.2	0.614
Forno *et al*. [21]	Severe asthma exacerbation	615	17	Scoring	-	-	-	-	0.75
Xiang *et al*. [22]	Asthma exacerbation	31,433	-	Recurrent neural network	-	-	-	-	0.70

In Table 2 of Multimedia Appendix 1, both the five most important features and multiple other features within the top 17 indicate loss of asthma control. It is important to note that loss of asthma control could be due in part to factors not well captured in our data, such as socioeconomic circumstances, variable management practices among providers, access to subspecialty clinicians, and non-adherence to medications and treatments. Variable asthma severity across patients over time also influences this process.

We checked 234 candidate features. Our final UWM model used 30.3% (71/234) of them. Despite being correlated with the outcome, many unused features had no extra predictive power on the UWM data set over the features adopted in our final UWM model.

When we did no retraining on the UWM data and directly applied our original simplified Intermountain Healthcare model trained on the Intermountain Healthcare training set [23], the model yielded an AUC of 0.861 on the UWM test set. This AUC is 0.041 lower than our final UWM model's AUC, but is still larger than the previously reported AUC of every existing model for forecasting asthmatic patients' future hospitalizations and ED visits (see Table 7). Hence, our simplified Intermountain Healthcare model showed excellent generalizability to UWM.

Compared to our full UWM model using 71 features, our simplified UWM model retained nearly all of its predictive power. When we did no retraining on the Intermountain Healthcare data and directly applied our simplified UWM model trained on the UWM training set, the model yielded an AUC of 0.814 on the Intermountain Healthcare test set. This AUC is 0.045 lower than our full Intermountain Healthcare model's AUC, but is still larger than the previously reported AUC of every existing model developed by others for forecasting asthmatic patients' future hospitalizations and ED visits (see Table 7). Hence, our simplified UWM model showed excellent generalizability to Intermountain Healthcare.

Comparison with the previous work

Researchers have built multiple models to forecast asthmatic patients' future hospitalizations and ED visits [2,9-23]. Table 7 compares our final UWM model with these models, which cover all of the relevant models described in Loymans *et al.*'s systematic review [17]. Our final UWM model's AUC is 0.902. Our Intermountain Healthcare model's AUC is 0.859. Every other existing model has a previously reported AUC ≤ 0.81 [2,9-22], which is lower than our final UWM model's AUC by at least 0.091.

It is important to consider the prevalence of the outcome of interest when comparing different predictive models' performance. Compared to other existing models, Yurk *et al.*'s model [16] achieved a higher sensitivity and PPV mainly because it adopted a differing prediction target: asthma hospital encounters or at least one day lost for diminished activities or missing work for asthma. This prediction target had a 54% prevalence rate in asthmatic patients and was therefore easier to forecast. If Yurk *et al.*'s model were employed to forecast asthma hospital encounters, an outcome that had a <2% prevalence rate in asthmatic patients, the model's sensitivity and PPV would likely drop.

Xiang *et al.*'s [22] recurrent neural network model reached a low AUC of 0.7 mainly because it used mostly inpatient data with little outpatient data; adopted only three types of attributes: medication, diagnosis, and demographics; and did not merge individual asthma medications into asthma medication categories such as nebulizer and short-acting beta-2 agonist. That is, the low AUC does not prove that the recurrent neural network is ineffective at predicting asthma outcomes, but is mainly due to incomplete data and insufficient feature modeling. In comparison, in building our final UWM model, we used both inpatient and outpatient data, adopted many types of attributes, and merged individual asthma medications into asthma medication categories to better capture and model the relationship among different asthma medications.

Excluding Yurk *et al.*'s model [16], every existing published model has a sensitivity ≤ 53.69%, which is significantly lower than our final UWM model's sensitivity of 70.18%. For the asthmatic patients who will have future asthma hospital encounters, sensitivity is the percentage of them identified by the model. The difference in sensitivity could have a significant impact on healthcare utilization. Due to asthma's high prevalence rate, for every 10% increase in the identified percentage of asthmatic patients who would have future asthma hospital encounters, up to 7,759 more hospitalizations and 71,074 more ED visits could be avoided in the U.S. each year with effective care management [1,4-7].

The targeted poor outcome's prevalence rate greatly impacts the PPV of any predictive model [41]. In our UWM test data set, 1.49% (218/14,644) of asthmatic patients had future asthma hospital encounters. When we placed the cutoff point for making binary classification at the top 10.00% (1,464/14,644) of asthmatic patients with the biggest forecasted risk, an impeccable model in theory would yield the highest possible PPV of 14.89% (218/1,464). Our final UWM model yielded a PPV of 10.45% (153/1,464), which is 70.18% of the highest possible PPV in theory. In comparison, our Intermountain Healthcare model achieved a PPV of 22.65% [23]. This is 53.69% of the highest possible PPV that an impeccable model in theory would yield on the Intermountain Healthcare test set. Lieu *et al.*'s [18] model yielded a PPV of 18.5% on a data set where 6.9% of asthmatic patients had future asthma hospital encounters. Schatz *et al.*'s [19] model yielded a PPV of 22.0% on a data set where 6.5% of asthmatic patients had future asthma hospital encounters. Compared to our case with UWM, both populations have a higher prevalence of asthma hospital encounters, which allows the PPV to be higher. Excluding these PPVs and Yurk *et al.*'s [16] model's PPV, no other existing published model's PPV exceeds 5.6%.

Our final UWM model and our Intermountain Healthcare model [23] have similar top features whose importance values are ≥0.01. In both models, many top features are related to prior ED visits and asthma medications. We had not identified several candidate features at the construction time of our Intermountain Healthcare model. They appeared as top features and affected the ranks and importance values of the other top features in our final UWM model.

Differing models in Table 7 were built using different patient cohorts and employ similar, but not necessarily identical prediction targets. Some features used in the models built by other researchers, such as certain features computed from patient reported outcomes and patient surveys, are unavailable in our UWM data set. Hence, we were unable to show the performance measures that the models built by other researchers would achieve on our UWM data set. Yet, we are confident that the techniques used by us improved prediction accuracy. Our final UWM model was built using a state-of-the-art machine learning algorithm, XGBoost. Compared to statistical methods such as logistic regression, machine learning can enhance prediction accuracy with less strict assumptions on data distribution [8,42,43]. Compared to the models built by other researchers, our final UWM model was built using more patients and a more extensive set of candidate features done with careful feature engineering, both of which are known to often help improve prediction accuracy [24-27,32]. As partial evidence for this, we

built predictive models for asthma hospital encounters using data from three healthcare systems UWM, Intermountain Healthcare [23], and Kaiser Permanente Southern California [44]. For each of the three healthcare systems, we started model building with around 20 candidate features and obtained unsatisfactory accuracy. This motivated us to examine several hundred candidate features. Ultimately, for each of the three healthcare systems, we built a model with an AUC that is higher than all of the AUCs other researchers previously reported in the literature for forecasting asthma hospital encounters [23,44]. This demonstrated the generalizability of our modeling strategy for forecasting asthma hospital encounters.

Considerations concerning the potential clinical use

Our final UWM model has an AUC that is higher than all of the AUCs previously reported in the literature for forecasting asthma hospital encounters, but still had a seemingly low PPV of 10.45% (153/1,464). Nevertheless, this model could provide value in clinical care:

(1) Healthcare systems like UWM, Intermountain Healthcare, and Kaiser Permanente Northern California [2] are using proprietary models to allocate asthma care management resources. These models and the models formerly built by others have similar performance measures. Our final UWM model has an AUC that is higher than the previously reported AUCs of all of these models.

(2) As explained above, even an impeccable model in theory would reach a low PPV because the poor outcome of interest has a low prevalence rate in our data set. For such an outcome, sensitivity better reflects the model's potential clinical value than PPV. Our final UWM model had a higher sensitivity than the previously reported sensitivity of every existing model using a comparable prediction target. It is important to note that while asthma hospital encounters have an overall low prevalence rate in the asthmatic patient population, they have significant financial and clinical impacts at both the population and the individual patient level.

(3) A PPV of 10.45% (153/1,464) is useful for finding high-risk asthmatic patients to receive low-cost preventive interventions. Below are four examples of such interventions: training the patient to record a diary about environmental triggers, coaching the patient to use an asthma inhaler correctly, coaching the patient to use a peak flow meter correctly and giving it to the patient to do self-monitoring of symptoms at home, and asking a nurse to do extra follow-up phone calls with the patient and/or the patient's caregiver. These interventions could have a significant impact on patient outcomes.

Our final UWM model employed 71 features. Reducing the number of features could ease clinical deployment of our model. To this end, if a minor decrease of prediction accuracy could be tolerated, one could adopt the top few features whose importance values are greater than a given threshold like 0.01 and drop the other features. The importance value of a feature varies across healthcare systems. Ideally, one should first calculate the features' importance values on a data set from the target healthcare system before choosing the features to retain.

As is typical with complex machine learning models, an XGBoost model using many features is hard to interpret. This can limit clinical understandability and adoption, particularly by clinicians who are resistant to using automated tools. In the future, we plan to adopt our previously developed method [39,40] to automatically explain the prediction results of our final UWM model.

Our final UWM model was constructed using XGBoost [33]. For binary classification on imbalanced data, XGBoost leverages a hyper-parameter scale_pos_weight to balance the two outcome classes' weights [45]. To maximize the AUC of our UWM model, our automatic model selection method [35] altered scale_pos_weight to a non-default value to balance the two outcome classes [46]. This incurs a side effect of significantly shrinking the model's forecasted probabilities of having future asthma hospital encounters to values much less than the actual probabilities [46]. This does not preclude us from choosing the top few percent of asthmatic patients having the greatest forecasted risk to receive various preventive interventions. To prevent this side effect from occurring, we could remain scale_pos_weight at its default value of one without doing any balancing. As a tradeoff, the model's AUC would decrease from 0.902 to 0.885 (95% CI: 0.861-0.907), yet even this decreased AUC is larger than all of the AUCs previously reported in the literature for forecasting asthma hospital encounters.

Limitations

This study has at least four limitations that could be interesting topics for future work:

(1) It is possible to raise model accuracy further using other features beyond those checked in this study. For example, features derived from environmental and physiological data gathered by intelligent wearable devices could have this potential.

(2) This study used purely structured data and checked only non-deep learning classification algorithms. It is possible to raise model accuracy further using deep learning as well as features derived from unstructured clinical notes using natural language processing techniques [40,47].

(3) Our UWM data set contained no data on patients' healthcare use outside of UWM. Hence, we limited the prediction target to asthma hospital encounters at UWM instead of asthma hospital encounters anywhere. Also, the features we checked were derived from patients' incomplete administrative and clinical data [48-51]. It would be worth investigating how model accuracy would vary if we have more complete administrative and clinical data of patients [52].

(4) This study evaluated our modeling strategy's generalizability to an academic healthcare system on a single outcome of a complex chronic disease. We recently showed that our modeling strategy also generalizes well to Kaiser Permanente Southern California for the same predictive modeling problem [44]. We plan to investigate our modeling strategy's generalizability to other diseases, outcomes, and healthcare systems in the future.

Conclusions

In the first evaluation of its generalizability to an academic healthcare system, our modeling strategy of examining many candidate features to enhance prediction accuracy showed excellent generalizability to UWM and led to a model with an AUC that is higher than all of the AUCs previously reported in the literature for forecasting asthma hospital encounters. After further optimization, our UWM model could be employed to facilitate efficient and effective allocation of asthma care management resources to improve outcomes.

Acknowledgments

We thank Katy Atwood for helping retrieve the UWM data set and Michael D. Johnson for useful discussions. GL, SDM, and GHD 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. YT did the work at the University of Washington when she was a visiting PhD student there.

Authors' contributions

YT participated in doing data analysis and writing the paper's first draft. GL conceptualized and designed the study, performed literature review, participated in doing data analysis, and rewrote the whole paper. AIM, ABW, SDM, GHD, and PS provided feedback on various medical issues, contributed to conceptualizing the presentation, and revised the paper.

Conflicts of interest

None declared.

Abbreviations:

AUC: Area Under the receiver operating characteristic Curve
ED: emergency department
FN: false negative
FP: false positive
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10: International Classification of Diseases, Tenth Revision
NPV: negative predictive value
PCP: primary care provider
PPV: positive predictive value
TN: true negative
TP: true positive
UWM: University of Washington Medicine
Weka: Waikato Environment for Knowledge Analysis
XGBoost: extreme gradient boosting

References

1. Centers for Disease Control and Prevention. Most recent national asthma data. 2020. https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm.
2. 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
3. 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
4. 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
5. 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
6. Kelly CS, Morrow AL, Shults J, Nakas N, Strope GL, Adelman RD. Outcomes evaluation of a comprehensive intervention program for asthmatic children enrolled in Medicaid. Pediatrics 2000;105(5):1029-35. PMID:10790458

7. 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.
8. Axelrod RC, Vogel D. Predictive modeling in health plans. Dis Manag Health Outcomes 2003;11(12):779-87. doi:10.2165/00115677-200311120-00003
9. Loymans RJ, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Assendelft WJ, Schermer TR, Chung KF, Sousa AR, Sterk PJ, Reddel HK, Sont JK, Ter Riet G. Identifying patients at risk for severe exacerbations of asthma: development and external validation of a multivariable prediction model. Thorax 2016;71(9):838-46. PMID:27044486
10. 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
11. Eisner MD, Yegin A, Trzaskoma B. Severity of asthma score predicts clinical outcomes in patients with moderate to severe persistent asthma. Chest 2012;141(1):58-65. PMID:21885725
12. Sato R, Tomita K, Sano H, Ichihashi H, Yamagata S, Sano A, Yamagata T, Miyara T, Iwanaga T, Muraki M, Tohda Y. The strategy for predicting future exacerbation of asthma using a combination of the Asthma Control Test and lung function test. J Asthma 2009;46(7):677-82. PMID:19728204
13. Osborne ML, Pedula KL, O'Hollaren M, Ettinger KM, Stibolt T, Buist AS, Vollmer WM. Assessing future need for acute care in adult asthmatics: the Profile of Asthma Risk Study: a prospective health maintenance organization-based study. Chest 2007;132(4):1151-61. PMID:17573515
14. Miller MK, Lee JH, Blanc PD, Pasta DJ, Gujrathi S, Barron H, Wenzel SE, Weiss ST; TENOR Study Group. TENOR risk score predicts healthcare in adults with severe or difficult-to-treat asthma. Eur Respir J 2006;28(6):1145-55. PMID:16870656
15. Peters D, Chen C, Markson LE, Allen-Ramey FC, Vollmer WM. Using an asthma control questionnaire and administrative data to predict health-care utilization. Chest 2006;129(4):918-24. PMID:16608939
16. Yurk RA, Diette GB, Skinner EA, Dominici F, Clark RD, Steinwachs DM, Wu AW. Predicting patient-reported asthma outcomes for adults in managed care. Am J Manag Care 2004;10(5):321-8. PMID:15152702
17. 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
18. 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
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. Grana J, Preston S, McDermott PD, Hanchak NA. The use of administrative data to risk-stratify asthmatic patients. Am J Med Qual 1997;12(2):113-9. PMID:9161058
21. 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
22. Xiang Y, Ji H, Zhou Y, Li F, Du J, Rasmy L, Wu S, Zheng WJ, Xu H, Zhi D, Zhang Y, Tao C. Asthma exacerbation prediction and risk factor analysis based on a time-sensitive, attentive neural network: retrospective cohort study. J Med Internet Res 2020;22(7):e16981. PMID:32735224
23. Luo G, He S, Stone BL, Nkoy FL, Johnson MD. Developing a model to predict hospital encounters for asthma in asthmatic patients: secondary analysis. JMIR Med Inform 2020;8(1):e16080. PMID:31961332
24. Mayfield J, McNamee P, Piatko CD. Named entity recognition using hundreds of thousands of features. In: Proceedings of the Seventh Conference on Natural Language Learning. 2003 Presented at: CoNLL'03; May 31-June 1, 2003; Edmonton, Canada p. 184-7.
25. Cao Y, Yu H, Abbott NL, Zavala VM. Machine learning algorithms for liquid crystal-based sensors. ACS Sens 2018; 3(11):2237-45. doi:10.1021/acssensors.8b00100
26. Zhai Y, Ong Y, Tsang IW. The emerging “big dimensionality.” IEEE Comput Intell Mag 2014;9(3):14-26. doi:10.1109/MCI.2014.2326099
27. Hansson K, Yella S, Dougherty M, Fleyeh H. Machine learning algorithms in heavy process manufacturing. American Journal of Intelligent Systems 2016;6(1):1-13. doi:10.5923/j.ajis.20160601.01
28. Liu LL, Forgione DA, Younis MZ. A comparative analysis of the CVP structure of nonprofit teaching and for-profit non-teaching hospitals. J Health Care Finance 2012;39(1):12-38. PMID:23155742
29. 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
30. 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

31. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating, 2nd ed. New York, NY: Springer; 2019. ISBN:3030163989
32. 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
33. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD'16; August 13-17, 2016; San Francisco, CA p. 785-94. doi:10.1145/2939672.2939785
34. XGBoost JVM package. 2020. https://xgboost.readthedocs.io/en/latest/jvm/index.html.
35. 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
36. Thornton C, Hutter F, Hoos HH, Leyton-Brown K. Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2013 Presented at: KDD'13; August 11-14, 2013; Chicago, IL p. 847-55. doi:10.1145/2487575.2487629
37. Agresti A. Categorical Data Analysis, 3rd ed. Hoboken, NJ: Wiley; 2012. ISBN:9780470463635
38. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY: Springer; 2016. ISBN:0387848576
39. 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
40. Luo G. A roadmap for semi-automatically extracting predictive and clinically meaningful temporal features from medical data for predictive modeling. Glob Transit 2019;1:61-82. PMID:31032483
41. Ranganathan P, Aggarwal R. Common pitfalls in statistical analysis: Understanding the properties of diagnostic tests - Part 1. Perspect Clin Res 2018;9(1):40-43. PMID:29430417
42. Kuhn M, Johnson K. Applied Predictive Modeling. New York, NY: Springer; 2013. ISBN:1461468485
43. Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS One 2014;9(2):e88225. PMID:24520356
44. Luo G, Nau CL, Crawford WW, Schatz M, Zeiger RS, Rozema E, Koebnick C. Developing a predictive model for asthma-related hospital encounters in patients with asthma in a large, integrated health care system: secondary analysis. JMIR Med Inform 2020;8(11):e22689. PMID:33164906
45. XGBoost parameters. 2020. https://xgboost.readthedocs.io/en/latest/parameter.html.
46. Notes on parameter tuning. 2020. https://xgboost.readthedocs.io/en/latest/tutorials/param_tuning.html.
47. Luo G, Stone BL, Koebnick C, He S, Au DH, Sheng X, Murtaugh MA, Sward KA, Schatz M, Zeiger RS, Davidson GH, Nkoy FL. Using temporal features to provide data-driven clinical early warnings for chronic obstructive pulmonary disease and asthma care management: protocol for a secondary analysis. JMIR Res Protoc 2019;8(6):e13783. PMID:31199308
48. Bourgeois FC, Olson KL, Mandl KD. Patients treated at multiple acute health care facilities: quantifying information fragmentation. Arch Intern Med 2010;170(22):1989-95. PMID:21149756
49. Finnell JT, Overhage JM, Grannis S. All health care is not local: an evaluation of the distribution of emergency department care delivered in Indiana. AMIA Annu Symp Proc 2011;2011:409-16. PMID:22195094
50. 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
51. Kern LM, Grinspan Z, Shapiro JS, Kaushal R. Patients' use of multiple hospitals in a major US city: implications for population management. Popul Health Manag 2017;20(2):99-102. PMID:27268133
52. Samuels-Kalow ME, Faridi MK, Espinola JA, Klig JE, Camargo CA Jr. Comparing statewide and single-center data to predict high-frequency emergency department utilization among patients with asthma exacerbation. Acad Emerg Med 2018;25(6):657-67. PMID:29105238

Appendix

Table 1. The list of candidate features considered in this study.

Feature category	Features
Features on patient demographics	Ethnicity (Hispanic or non-Hispanic); gender; age; marital status (married, single, partnered, divorced, widowed, or separated); race; and language.
Features related to laboratory tests	No. of laboratory tests; no. of laboratory tests having abnormal results; no. of days from the most recent laboratory test; whether an immunoglobulin E (IgE) test was performed; whether the greatest total serum IgE level is abnormally high; the greatest total serum IgE level; the largest percentage of blood eosinophils; and the largest blood eosinophil count.

Features related to vital signs	The mean diastolic blood pressure; the mean heart rate; the mean systolic blood pressure; the highest diastolic blood pressure; the mean temperature; the highest systolic blood pressure; the greatest heart rate; the mean respiratory rate; the greatest respiratory rate; the highest temperature; the mean peripheral capillary oxygen saturation (SpO_2); the lowest SpO_2; *the mean peak expiratory flow; the lowest peak expiratory flow*; the relative change of weight = (the most recently documented weight / the first documented weight - 1) × 100%; the largest body mass index (BMI); and the relative change of BMI = (the most recently documented BMI / the first documented BMI - 1) × 100%.
Features that are related to diagnoses and calculated solely on ICD-9 and ICD-10 diagnosis codes	No. of ICD-9 and ICD-10 diagnosis codes; no. of years from the first encounter related to asthma in the data set; *no. of primary or principal asthma diagnoses; no. of asthma diagnoses; whether the most recent asthma diagnosis is a primary or principal one; the severity of the most recent asthma diagnosis; the severity of the most severe asthma diagnosis; no. of diagnoses of asthma with status asthmaticus; no. of diagnoses of asthma with (acute) exacerbation; the exacerbation severity (uncomplicated, exacerbation, or asthmaticus) of the most recent asthma diagnosis; the greatest exacerbation severity of any asthma diagnosis; no. of days from the most recent asthma diagnosis; no. of days from the most recent diagnosis of asthma with (acute) exacerbation or status asthmaticus; no. of diagnoses of noncompliance with medication regimen; family history of asthma*; chronic obstructive pulmonary disease; no. of years from the first encounter related to chronic obstructive pulmonary disease in the data set; esophagitis; allergic rhinitis; anxiety or depression; ischemic heart disease; eczema; gastroesophageal reflux; sleep apnea; gastrostomy tube; obesity; Alzheimer's or Parkinson's disease; upper respiratory tract infection; decreased tone; increased tone; cystic fibrosis; immunoglobulin A (IgA) deficiency; pneumonia; vocal cord dysfunction; psoriasis; anaphylaxis; vasculitis; cirrhosis; gastrointestinal bleeding; gastrointestinal obstruction; inflammatory bowel disease; mental disorder; breathing abnormality like dyspnea; pregnancy; vitamin D deficiency; myocardial infarction; folate deficiency; congestive heart failure; malignancy; peripheral vascular disease; dementia; peptic ulcer disease; cerebrovascular disease; substance use; rheumatic disease; diabetes with chronic complication; renal disease; diabetes without chronic complication; bronchopulmonary dysplasia; moderate or severe liver disease; mild liver disease; paraplegia or hemiplegia; acquired immunodeficiency syndrome; and metastatic solid tumor.
Features that are related to diagnoses and calculated jointly on ICD-9 and ICD-10 procedure codes along with ICD-9 and ICD-10 diagnosis codes	Tracheostomy.
Features that are related to diagnoses and calculated jointly on Current Procedural Terminology (CPT) and Healthcare Common Procedure Coding System (HCPCS) procedure codes along with ICD-9 and ICD-10 diagnosis codes	Cataract; and sinusitis.
Features related to the problem list	*No. of active problems; no. of active problems of wheezing; no. of active problems of asthma with (acute) exacerbations; no. of active problems of asthma; no. of active problems of obesity; no. of active problems of congestive heart failure; no. of active problems of sleep apnea; no. of active problems of hypertension; no. of active problems of chronic obstructive pulmonary disease; no. of active problems of rhinitis; no. of active problems of diabetes; no. of active problems of anxiety/depression; no.of active problems of gastroesophageal reflux disease; no. of active problems about smoking; and the priority of the last active problem of asthma*.
Features related to medications	Total no. of medications ordered; no. of medication orders; total no. of differing medications ordered; total no. of medication refills permitted; total no. of units of medications ordered; total no. of asthma medications ordered; no. of asthma medication orders; total no. of differing

	asthma medications ordered; total no. of asthma medication refills permitted; total no. of units of asthma medications ordered; total no. of short-acting beta-2 agonists ordered; total no. of refills permitted for short-acting beta-2 agonists; total no. of units of short-acting beta-2 agonists ordered; total no. of systemic corticosteroids ordered; total no. of refills permitted for systemic corticosteroids; total no. of units of systemic corticosteroids ordered; *no. of asthma reliever orders; total no. of asthma relievers ordered; total no. of refills permitted for asthma relievers; total no. of differing asthma relievers ordered; total no. of units of asthma relievers ordered*; total no. of units of asthma relievers ordered that are neither short-acting beta-2 agonists nor systemic corticosteroids; total no. of asthma relievers ordered that are neither short-acting beta-2 agonists nor systemic corticosteroids; *no. of asthma controller orders; total no. of asthma controllers ordered; total no. of units of asthma controllers ordered; total no. of differing asthma controllers ordered; total no. of refills permitted for asthma controllers*; total no. of inhaled corticosteroids ordered; total no. of refills permitted for inhaled corticosteroids; total no. of units of inhaled corticosteroids ordered; total no. of mast cell stabilizers ordered; total no. of refills permitted for mast cell stabilizers; total no. of units of mast cell stabilizers ordered; whether nebulizer was used; *total no. of nebulizer medications ordered; no. of nebulizer medication orders; total no. of units of nebulizer medications ordered; total no. of differing nebulizer medications ordered; total no. of refills permitted for nebulizer medications*; and whether spacer was used.
Features related to insurances	Whether the patient had any public insurance on the last day; whether the patient had any private insurance on the last day; and whether the patient was paid by oneself or a charity on the last day.
Features related to the patient's visit types	No. of ED visits; *no. of ED visits related to asthma*; the most recent ED visit's length of stay; the mean length of stay of an ED visit; no. of visits; no. of outpatient visits to the patient's PCP; no. of outpatient visits; no. of outpatient visits whose primary diagnosis is asthma; no. of hospitalizations; the hospitalizations' total length; the mean length of a hospitalization; no. of hospitalizations, ED visits, and outpatient visits; no. of intensive care admissions; the most emergent admission type of all of the visits; the most recent visit's admission type (elective, urgent, emergency, or trauma); no. of major visits for asthma; and *the most recent visit's type (ED visit, outpatient visit, or hospitalization)*. As in our prior paper [23], we defined a major visit for asthma as an ED visit having an asthma diagnosis code, a hospitalization having an asthma diagnosis code, or an outpatient visit having a primary diagnosis of asthma. An outpatient visit having only a secondary diagnosis of asthma was treated as a minor visit for asthma.
Features on visit status and appointment scheduling	The day of the week when the most recent ED visit began; no. of cancelled appointments; the most recent visit's discharge disposition location (home, left against medical advice, or other non-home location); no. of no shows; no. of times of leaving against medical advice; for the most recent visit, the time to the actual visit after making the request showing its urgency; across all of the visits, the shortest time to the actual visit after making the request; whether the most recent hospitalization came from the ED; no. of days from the most recent hospitalization; *no. of visits having same day appointments*; no. of days from the most recent ED visit; no. of days from the most recent outpatient visit; *no. of days from the most recent ED visit on asthma; and no. of days from the most recent outpatient visit on asthma*.
Features describing the patient's care continuity degree	No. of differing asthma medication prescribers; no. of differing EDs the patient went to; no. of differing medication prescribers; no. of differing providers the patient saw in outpatient visits; and no. of differing PCPs of the patient.
Features related to procedures	Mechanical ventilation reflected by ICD-10 and ICD-9 procedure codes; no. of ICD-10 and ICD-9 procedure codes; no. of HCPCS procedure codes of home oxygen therapy; no. of CPT procedure codes of the fractional exhaled nitric oxide test; no. of CPT/HCPCS procedure codes; no. of CPT procedure codes of pulmonary function tests; and no. of CPT/HCPCS procedure codes of influenza vaccination.
Allergy features	Indicator of drug or material allergy; *the greatest severity of the patient's drug or material allergies*; indicator of environmental allergy; *the greatest severity of the patient's environmental allergies*; indicator of food allergy; *the greatest severity of the patient's food allergies*; and no. of the patient's allergies.

Features related to pulmonary function tests	*The mean forced expiratory volume in 1 second (FEV1); and the lowest FEV1*.
Features related to social behavior history	*No. of fluid ounces of alcohol the patient consumed every week according to the most recent record; whether the patient was ever documented of consuming alcohol; whether the patient consumed alcohol according to the most recent record; the mean no. of fluid ounces of alcohol the patient consumed every week across all of the records; no. of alcohol drinks the patient consumed every week according to the most recent record; the mean no. of alcohol drinks the patient consumed every week across all of the records; no. of packs of cigarettes the patient consumed every day according to the most recent record; whether the patient was a smoker according to the most recent record; the mean no. of packs of cigarettes the patient consumed every day across all of the records; whether the patient was a former smoker according to the most recent record; no. of years the patient had smoked for according to the most recent record; no. of times the patient took illicit drugs every week according to the most recent record; whether the patient took any illicit drug according to the most recent record; the mean no. of times the patient took illicit drugs every week across all of the records; and whether the patient was ever documented of taking any illicit drug*.
Provider features	We defined the patient's PCP known at the most recent clinic visit as the patient's current PCP. We considered the following PCP features: *no. of years that the PCP had practiced at UWM for*; the PCP's age; whether the patient is of the same gender as the PCP; the PCP's primary specialty; whether the PCP is a resident; the PCP's type (physician, nurse, physician assistant, or other); the PCP's clinician title (doctor of medicine, registered nurse, physician assistant, or other); the percentage of the PCP's asthmatic patients in the pre-index year incurring asthma hospital encounters in the index year; and no. of asthmatic patients of the PCP.

Table 2. The features used in our final UWM predictive model and their importance values.

Rank	Feature	Importance calculated as the feature's apportioned contribution to the model
1	No. of ED visits related to asthma	0.2113
2	The mean length of stay of an ED visit	0.1088
3	No. of days from the most recent ED visit	0.1075
4	No. of primary or principal asthma diagnoses	0.0986
5	No. of days from the most recent diagnosis of asthma with (acute) exacerbation or status asthmaticus	0.0798
6	Whether the patient is black	0.0423
7	No. of ED visits	0.0368
8	No. of asthma diagnoses	0.0322
9	No. of years from the first encounter related to asthma in the data set	0.0233
10	No. of nebulizer medication orders	0.0174
11	Whether the patient is white	0.0148
12	The highest systolic blood pressure	0.0128
13	No. of CPT/HCPCS procedure codes	0.0120
14	The largest BMI	0.0107
15	The most recent ED visit's length of stay	0.0104
16	Whether the patient is married	0.0102
17	Total no. of units of medications ordered	0.0101
18	No. of asthma medication orders	0.0087
19	Whether nebulizer was used	0.0082
20	No. of no shows	0.0081
21	Total no. of differing asthma medications ordered	0.0078
22	The mean heart rate	0.0072
23	No. of diagnoses of asthma with (acute) exacerbation	0.0070
24	Whether the patient had any private insurance on the last day	0.0060
25	The mean respiratory rate	0.0058

26	No. of days from the most recent asthma diagnosis	0.0056
27	For the most recent visit, the time to the actual visit after making the request	0.0054
28	The mean systolic blood pressure	0.0053
29	Total no. of differing medications ordered	0.0046
30	Whether the patient has any drug or material allergy	0.0046
31	Whether the patient had any public insurance on the last day	0.0042
32	The lowest SpO_2	0.0040
33	No. of active problems	0.0038
34	Whether the most recent visit is an ED visit	0.0038
35	The highest temperature	0.0037
36	No. of laboratory tests	0.0034
37	No. of asthma controller orders	0.0029
38	No. of visits	0.0029
39	Total no. of short-acting beta-2 agonists ordered	0.0029
40	The largest blood eosinophil count	0.0028
41	No. of medication orders	0.0027
42	Total no. of asthma relievers ordered that are neither short-acting beta-2 agonists nor systemic corticosteroids	0.0025
43	The mean temperature	0.0022
44	Total no. of inhaled corticosteroids ordered	0.0020
45	No. of days from the most recent outpatient visit on asthma	0.0020
46	The severity of the most recent asthma diagnosis	0.0020
47	Total no. of refills permitted for short-acting beta-2 agonists	0.0019
48	No. of differing providers the patient saw in outpatient visits	0.0019
49	Age	0.0019
50	No. of outpatient visits to the patient's PCP	0.0019
51	No. of laboratory tests having abnormal results	0.0018
52	Total no. of systemic corticosteroids ordered	0.0016
53	Whether the most recent asthma diagnosis is a primary or principal one	0.0014
54	Whether the patient is single	0.0013
55	The day of the week when the most recent ED visit began	0.0013
56	The relative change of BMI	0.0013
57	The mean length of a hospitalization	0.0012
58	No. of days from the most recent ED visit on asthma	0.0011
59	No. of active problems of asthma	0.0010
60	Total no. of differing nebulizer medications ordered	0.0010
61	Total no. of differing asthma relievers ordered	0.0010
62	Total no. of refills permitted for asthma controllers	0.0010
63	Whether the patient has any mental disorder	0.0009
64	The relative change of weight	0.0008
65	Whether the most recent visit's admission type is emergency	0.0007
66	Whether the patient was a smoker according to the most recent record	0.0007
67	Whether the most recent hospitalization came from the ED	0.0006
68	The severity of the most severe asthma diagnosis	0.0005
69	No. of outpatient visits	0.0005
70	Total no. of medication refills permitted	0.0005
71	The highest diastolic blood pressure	0.0005

Abbreviations:

BMI: body mass index
CPT: Current Procedural Terminology
FEV1: forced expiratory volume in 1 second
FVC: forced vital capacity
HCPCS: Healthcare Common Procedure Coding System

IgE: immunoglobulin E
SpO_2: peripheral capillary oxygen saturation