Error and Timeliness Analysis for Using Machine Learning to Predict Asthma Hospital Visits: Retrospective Cohort Study

Xiaoyi Zhang, MS; **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
xyzhang7@uw.edu, 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 98195, USA
Phone: 1-206-221-4596
Fax: 1-206-221-2671
Email: luogang@uw.edu

Abstract

Background: A significant burden on health care comes from asthma hospital visits including emergency department visits and inpatient stays. To leverage preventive care more effectively in managing asthma, we formerly employed machine learning and the University of Washington Medicine (UWM) data to build the world's most accurate model to forecast which asthma patients will encounter asthma hospital visits during the successive 12 months.

Objective: Currently, two questions remain regarding our model's performance. First, for a patient who will encounter asthma hospital visits in the future, how timely can our model identify the risk for the first time? Second, if our model erroneously predicts a patient to encounter asthma hospital visits at the UWM during the successive 12 months, how likely will the patient encounter ≥1 asthma hospital visit somewhere else or have ≥1 surrogate of a poor outcome? This work aims to answer these two questions.

Methods: The patient cohort covered every adult asthma patient who received care at the UWM during 2011-2018. Using the UWM data, our model made predictions on the asthma patients in 2018. For every such patient with ≥1 asthma hospital visit at the UWM in 2019, we computed the number of days of advanced warning that our model gave on the patient for the first time. For every such patient erroneously projected to encounter ≥1 asthma hospital visit at the UWM in 2019, we used PreManage and the UWM data to check whether the patient had ≥1 asthma hospital visit outside of the UWM in 2019 or any surrogate of a poor outcome. Surrogates of poor outcomes included order of systemic corticosteroids during the successive 12 months, any type of visit for asthma exacerbation during the successive 12 months, and asthma hospital visit between the successive 13-24 months.

Results: Among the 218 asthma patients in 2018 with asthma hospital visits at the UWM in 2019, 61.9% (135/218) were given risk warnings for such visits for the first time ≥3 months ahead by our model and 84.4% (184/218) were given risk warnings ≥1 day ahead. Among the 1,310 asthma patients in 2018 erroneously projected to encounter asthma hospital visits at the UWM in 2019, 29.01% (380/1,310) had asthma hospital visits outside of the UWM in 2019 or surrogates of poor outcomes.

Conclusions: Our model gave timely risk warnings for most asthma patients with poor outcomes. 29.01% (380/1,310) of asthma patients for whom our model gave false-positive predictions had asthma hospital visits somewhere else during the successive 12 months or surrogates of poor outcomes, and were reasonable candidates for preventive interventions. There is still significant room for improving our model to give more accurate and more timely risk warnings.

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

Keywords: Asthma; machine learning; clinical decision support; forecasting; patient care management

Introduction

Background

Over 262 million people in the world have asthma [1]. In the United States, around 7.8% of Americans have asthma, which leads to 1.6 million emergency department (ED) visits, 179 thousand inpatient stays [2], and an aggregate medical cost of U.S. $50.3 billion [3] annually. A main goal in asthma management is to curtail asthma hospital visits, i.e., ED visits and inpatient stays for asthma. To serve this purpose, a state-of-the-art approach is to implement a predictive model to find patients who are at significant risk of encountering asthma hospital visits in the future. If deemed high risk, the patient is considered for enrollment in a care management program to receive preventive interventions. There, a care manager regularly follows up with the patient to monitor the status of asthma control, to alter asthma medications as the need arises, and to help book relevant services. This approach is employed by lots of health care systems, such as Intermountain Healthcare, the University of Washington Medicine (UWM), and Kaiser Permanente Northern California [4], along with many health plans, such as the health plans in 9 of 12 urban communities [5]. When used properly, this approach can curtail asthma hospital visits by up to 40% [5-9].

A care management program typically takes no more than 3% of the patients due to capacity constraints [10]. To optimize the efficacy of such programs, we recently employed extreme gradient boosting (XGBoost) [11], a machine learning algorithm, and the UWM data to build the world's most accurate model to forecast which asthma patients will encounter asthma hospital visits during the successive 12 months [12]. Our model obtained an area under the receiver operating characteristic curve of 0.902, a specificity of 90.91% (13,115/14,426), a sensitivity of 70.2% (153/218), a positive predictive value of 10.45% (153/1,464), a negative predictive value of 99.51% (13,115/13,180), and an accuracy of 90.60% (13,268/14,644) [12]. Compared with every prior model for this prediction task [4,13-26], our model improved the area under the receiver operating characteristic curve by ≥10%.

Objective

Currently, two questions remain regarding our model's performance. First, for a patient who will encounter asthma hospital visits in the future, how timely can our model identify the risk for the first time? Since any preventive intervention requires sufficient time to take effect [27,28], a model should identify the risk as early as possible to provide preventive interventions in time to avoid the poor outcome. Second, if our model erroneously predicts a patient to encounter ≥1 asthma hospital visit at the UWM during the successive 12 months, how likely will the patient encounter ≥1 asthma hospital visit somewhere else outside of the UWM or have ≥1 surrogate of a poor outcome? As our model was trained on the UWM data, it can only predict future asthma hospital visits at the UWM. The goal of this work is to answer these two questions. Part of the analysis that we conducted to answer the second question was formerly published as an abstract in the 2022 American Academy of Allergy, Asthma & Immunology annual meeting [29].

Methods

Parts reused from our prior paper

The following parts were reused from our prior UWM model building paper [12]: the patient cohort, the features, the prediction target, the cutoff point for conducting binary classification, the training set, the test set, and the predictive model.

Ethics approval

The institutional review board of the UWM approved this retrospective cohort study.

Patient cohort

As the biggest academic health care system in Washington state, the UWM maintains an enterprise data warehouse storing clinical and administrative data from 12 clinics and 3 hospitals for adults. The patient cohort was composed of every adult asthma patient (age in years ≥ 18) who was given care at any of those 15 UWM facilities during 2011-2018. A patient was deemed asthmatic in a given year when the patient's visit billing data in that year included ≥1 asthma diagnosis code (International Classification of Diseases, Tenth Revision [ICD-10]: J45.x; International Classification of Diseases, Ninth Revision [ICD-9]: 493.1x, 493.0x, 493.9x, 493.8x) [13,30]. This asthma case finding method has been shown to strike the best balance between sensitivity and positive predictive value among several rule-based asthma case finding methods, does not require the patient to have >1 year of historical data, and suits to be used for population health management [30]. Patients who died during that year were excluded.

Data sets

Two data sets were used. The first data set was retrieved from the UWM's enterprise data warehouse. The data set held structured administrative and clinical data documented for the visits by the patient cohort to the 15 UWM facilities during 2011 to 2020. The second data set came from Collective Medical Technologies Inc.'s commercial product PreManage [31]. The data set contained structured visit and diagnosis data of the ED visits and inpatient stays that our patient cohort had at every hospital in Washington state and many other American hospitals outside of Washington state during 2019.

Overview of our predictive model

Prediction target, the training set, and the test set

For an asthma patient at a given time point, the prediction target was whether the patient would encounter ≥1 asthma hospital visit during the successive 12 months. The prediction was made using the patient's data up to this time point. An asthma hospital visit was defined as an ED visit or an inpatient stay with a principal diagnosis of asthma (ICD-10: J45.x; ICD-9: 493.1x, 493.0x, 493.9x, 493.8x). During model training and testing, for each patient who was asthmatic in a given year, we adopted the data of the patient by the end of the year to predict the outcome of the patient in the successive 12 months [12]. Since the prediction target came from the successive 12 months, the UWM data between 2011 and 2019 provided 8 years of effective data for model training and testing. The effective data between 2011 and 2017 were employed as the training set for training our predictive model. The effective data of 2018 were used as the test set for testing our model. To answer the two questions posed in the introduction, we focused on the asthma patients in the test set, i.e., the asthma patients in 2018, and examined the predictions that our model made on these patients. For the asthma patients in 2018 who were erroneously projected to encounter asthma hospital visits at the UWM in 2019, the UWM data of 2020 were used to compute one of the surrogates of poor outcomes.

Machine learning algorithm and features

Our predictive model was constructed using 71 features and the XGBoost classification algorithm [11]. These 71 features are presented in the online multimedia appendix of our prior UWM model building paper [12]. They were constructed using the attributes in our UWM data set, which cover diverse aspects such as diagnoses, patient demographics, vital signs, visits, laboratory tests, procedures, and medications. Two exemplary features are the number of days from the patient's most recent

ED visit and the count of asthma diagnoses that the patient received in the past 12 months. These 71 features were included in every data instance inputted to our predictive model.

Cutoff point for conducting binary classification

We put the cutoff point for conducting binary classification at the uppermost 10% of the risk scores computed by our model. Each patient with a risk score above this cutoff point was projected to encounter ≥1 asthma hospital visit during the successive 12 months.

Assessing the timeliness of the risk warnings given for the first time by our model

Given a predictive model and an asthma patient in 2018 whose first asthma hospital visit in 2019 happened on date T, we measured k, the number of days of advanced warning that the model gave on the patient for the first time. To compute k, we started from $T - 365$ and kept moving forward along the timeline to find the earliest date T' ($T - 365 \leq T' \leq T - 1$) such that by taking the feature values computed on the patient's historical data up to T' as inputs, the model would predict the patient to encounter ≥1 asthma hospital visit during the 12 months after T'. In this case, the model warned the patient's first asthma hospital visit after T' k ($1 \leq k \leq T - T'$) days beforehand, with $T' + k$ being the starting date of the patient's first asthma hospital visit after T' (see Figure 1). Otherwise, if the model still predicted no future asthma hospital visit when we reached $T - 1$, the model warned the patient's asthma hospital visit on T $k = 0$ day beforehand. The larger the k, the more timely the risk warning that the model gave on the patient for the first time. k reflected how early before a poor outcome occurred the care manager would be prompted for the first time to consider giving the patient preventive interventions. The value of k was not affected by any prediction made by the model when the feature values computed on the patient's historical data up to a given date after T' were taken as inputs.

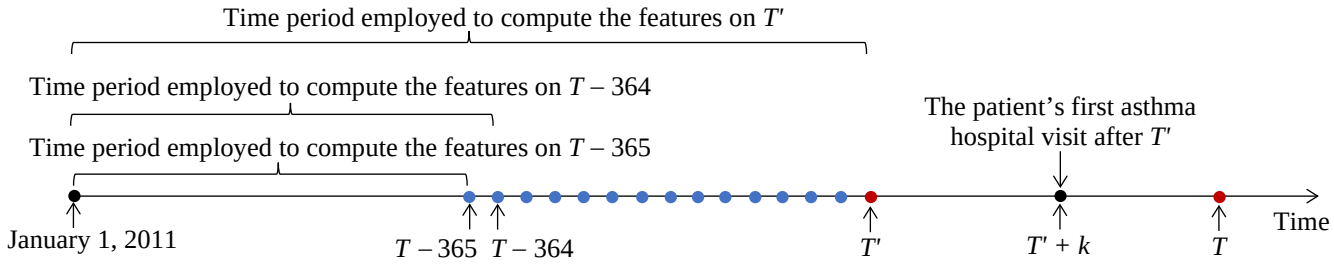

Figure 1. The way to compute k.

For our predictive model, we computed k on every asthma patient in 2018 who encountered ≥1 asthma hospital visit at the UWM in 2019 and presented the mean and the distribution of k.

Analyzing the false-positive predictions made by our model

For each asthma patient in 2018 whom our model erroneously projected to encounter ≥1 asthma hospital visit at the UWM in 2019, we used PreManage data to check whether the patient encountered ≥1 asthma hospital visit outside of the UWM in 2019. We used the UWM data to check whether the patient had any surrogate of a poor outcome. Surrogates of poor outcomes included order of systemic corticosteroids during the successive 12 months (i.e., in 2019), any type of visit with a primary or principal diagnosis of asthma exacerbation during the successive 12 months (i.e., in 2019), and asthma hospital visit between the successive 13-24 months (i.e., in 2020). Systemic corticosteroids are used to treat asthma exacerbations. In addition, if the patient had ≥1 order of systemic corticosteroids in 2019, we computed the number of systemic corticosteroids ordered for the patient in 2019 counting multiplicity. This number partially reflected the degree of the patient's poor asthma control. We presented the distribution of this number.

Results

The clinical characteristics and the demographics of our patient cohort

Tables 1 and 2 in the Appendix present the summary statistics of the clinical characteristics and the demographics of the UWM asthma patients during 2011-2017 and in 2018, respectively. There, every data instance links to a distinct (index year, patient) pair and is employed to project the outcome of the patient in the successive 12 months. Our prior paper [12] included detailed comparison results of the clinical characteristics and the demographics of the two sets of patients.

The timeliness of the risk warnings given for the first time by our model

218 or 1.49% (218/14,644) of the asthma patients in 2018 encountered asthma hospital visits at the UWM in 2019. Figure 2 plots the distribution of the number of days of advanced warning for asthma hospital visit that our model gave for the first time for every such patient. The mean and the standard deviation of the number of days of advanced warning is 190 and 150, respectively. For these 218 patients, our model could give risk warnings for the first time ≥12 months ahead for 30.7% (67/218) of them, ≥6 months ahead for 49.1% (107/218) of them, ≥3 months ahead for 61.9% (135/218) of them, ≥1 month ahead for 76.6% (167/218) of them, ≥2 weeks ahead for 83.0% (181/218) of them, and ≥1 day ahead for 84.4% (184/218) of them.

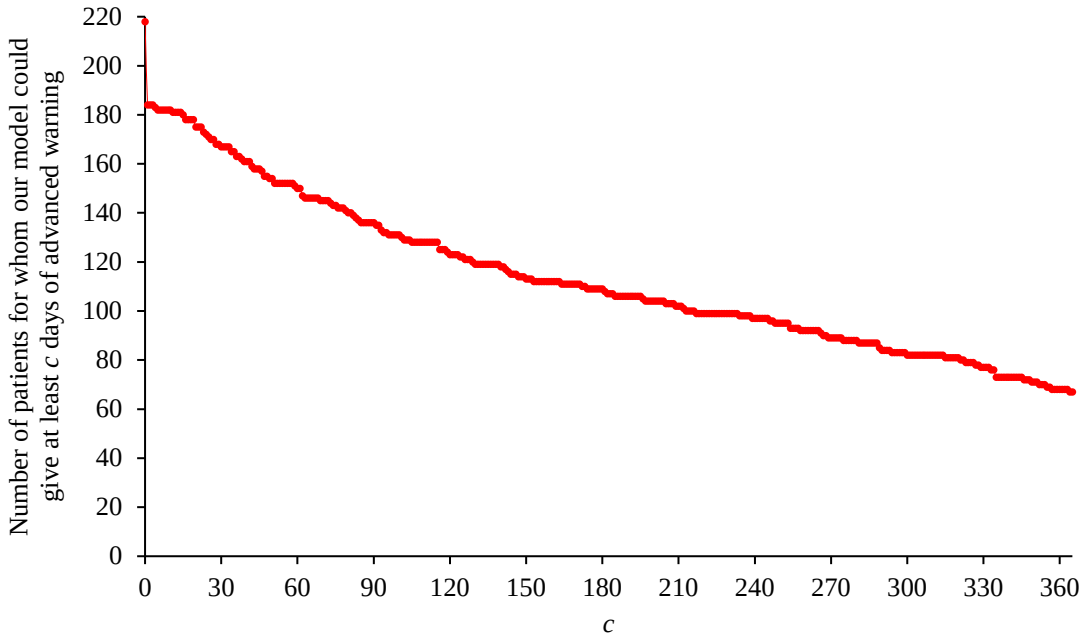

Figure 2. Among the 218 asthma patients in 2018 who encountered asthma hospital visits at the UWM in 2019, the number of patients for whom our model could give at least c days of advanced warning vs. c ($0 \leq c \leq 365$).

Breakdown of the false-positive predictions made by our model

1,310 asthma patients in 2018 were erroneously projected by our model to encounter asthma hospital visits at the UWM in 2019 [12]. Table 1 shows the number of these patients who had ≥1 asthma hospital visit outside of the UWM in 2019 or ≥1 surrogate of poor outcomes.

Table 1. Among the 1,310 asthma patients in 2018 whom our model erroneously projected to encounter asthma hospital visits at the UWM in 2019, the number of patients who had ≥1 asthma hospital visit outside of the UWM in 2019 or ≥1 surrogate of poor outcomes.

Outcome	Patients with this outcome (n=1,310), n (%)
(a) ≥1 order of systemic corticosteroids during the successive 12 months	316 (24.12)
(b) Any type of visit with a primary or principal diagnosis of asthma exacerbation during the successive 12 months	126 (9.62)
(c) Asthma hospital visit between the successive 13-24 months	18 (1.37)
(d) ≥1 asthma hospital visit outside of the UWM during the successive 12 months	39 (2.98)
Any of (a), (b), and (c)	358 (27.33)
Any of (a), (b), (c), and (d)	380 (29.01)

316 asthma patients in 2018 were erroneously predicted by our model to encounter ≥1 asthma hospital visit at the UWM in 2019 and had ≥1 order of systemic corticosteroids in 2019. Figure 3 plots the distribution of the number of systemic corticosteroids ordered for every such patient in 2019 counting multiplicity. The maximum value of this number is 118.

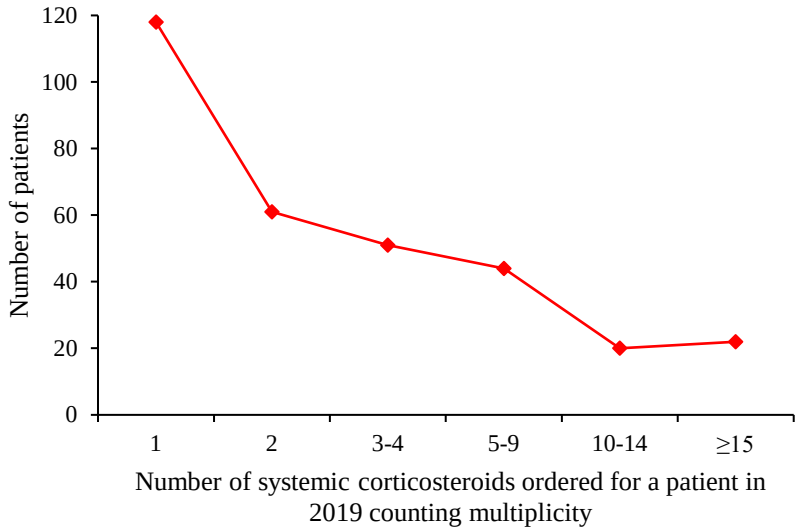

Figure 3. Among the 316 asthma patients in 2018 who were erroneously predicted by our model to encounter ≥1 asthma hospital visit at the UWM in 2019 and had ≥1 order of systemic corticosteroids in 2019, the distribution of the number of systemic corticosteroids ordered for every patient in 2019 counting multiplicity.

Discussion

Principal results

Among the 218 asthma patients in 2018 who encountered asthma hospital visits at the UWM in 2019, the number of patients for whom our model could give at least c days of advanced warning decreases roughly linearly with c ($0 \leq c \leq 365$) at a fast pace. Our model gave timely risk warnings (e.g., ≥3 months ahead) for a large proportion of these 218 asthma patients. Nevertheless, for another large proportion of these 218 asthma patients, our model could not give timely risk warnings. The model either gave risk warnings at most a few days ahead or could not foresee a patient's risk even on the day immediately before an asthma hospital visit.

Among the 1,310 asthma patients in 2018 whom our model erroneously projected to encounter asthma hospital visits at the UWM in 2019, 29.01% (380/1,310) had asthma hospital visits outside of the UWM in 2019 or surrogates of poor outcomes, and hence were reasonable candidates for preventive interventions. Among the 316 of these patients who had ≥1 order of systemic corticosteroids in 2019, a large proportion had rather poor asthma control, as reflected by a non-trivial number of systemic corticosteroids being ordered for each patient in 2019.

Are the risk warnings given for the first time by our model timely enough?

A predictive model should identify the risk for having future asthma hospital visits as early as possible to give the patient preventive interventions in time to avoid the poor outcome. The time needed for a preventive intervention to take effect varies by interventions. To the best of our knowledge, there is no consensus on the amount of time needed for a particular preventive intervention or a particular combination of preventive interventions to take effect for averting future asthma hospital visits. Consequently, in this study, we could not compute the exact percentage of patients with future asthma hospital visits for whom our model could give timely risk warnings. Nevertheless, we can shed some light on the rough range of this percentage. In a prior study [27,28], several clinicians gave the opinion that up to 3 months could be needed for any intervention to take effect for averting inpatient stays for an ambulatory care sensitive, chronic condition such as asthma. For 61.9% (135/218) of the 218 asthma patients in 2018 who encountered asthma hospital visits at the UWM in 2019, our model could give risk warnings for the first time ≥3 months ahead. Accordingly, we would expect the percentage of patients with future asthma hospital visits for whom our model could give timely risk warnings to be at least 61.9%, which is large. On the other hand, for 15.6% (34/218) of the 218 asthma patients in 2018 who encountered asthma hospital visits at the UWM in 2019, our model could not foresee a patient's risk even on the day immediately before an asthma hospital visit. Thus, the percentage of patients with future asthma hospital visits for whom our model could not give timely risk warnings is at least 15.6%, which is also large. Combining these two parts, we estimate the percentage of patients with future asthma hospital visits for whom our model could give timely risk warnings to be somewhere between 61.9% and 84.4%. There is still significant room for improving our model to give more timely risk warnings.

Potential impact of the false-positive predictions made by our model

We formerly invented an automated method to supply rule-style explanations for the predictions that an arbitrary machine learning model makes on tabular data and to suggest tailored interventions [32,33]. Whenever our model gave a risk warning for a patient, we could use this method to help clinicians decide whether the patient should be enrolled in a care management program, receive other preventive interventions that are less expensive than care management, or obtain no preventive intervention. For 134 (87.6%) of the 153 asthma patients in 2018 whom our model accurately projected to encounter asthma hospital visits at the UWM in 2019, our method supplied rule-style explanations for the predictions made by our model [32]. Each such explanation included ≥1 modifiable risk factor and linked to ≥1 intervention [32], whereas the situation could change for another prediction target or another health care system.

29.01% (380/1,310) of the asthma patients in 2018 whom our model erroneously projected to encounter asthma hospital visits at the UWM in 2019 had asthma hospital visits outside of the UWM in 2019 or surrogates of poor outcomes. These patients could benefit from the information provided by our automated explanation method. For the other 70.99% (930/1,310) of the asthma patients in 2018 whom our model erroneously projected to encounter asthma hospital visits at the UWM in 2019, our model's predictions could be truly inaccurate, leaving significant room for improving our model's accuracy. To know how many of these predictions would mislead clinicians to make incorrect intervention decisions, we would need to perform a user study with clinicians. This is left as an area of interest for future work.

Related work

To the best of our knowledge, no prior study has used either any surrogate of a poor outcome or future asthma hospital visits somewhere else to analyze the false-positive predictions made by a predictive model for asthma hospital visits. Also, no prior study has assessed the timeliness of the risk warnings given for the first time by such a model. For predicting Clostridium difficile infection during an inpatient stay, Wiens *et al*. [34] measured the number of days of advanced warning that a model gives on the patient. For predicting the total amount of donations that a fundraiser can obtain on a medical crowdfunding platform, Wang *et al*. [35] measured the prediction timeliness by the number of days of input data that a model needs in order to produce predictions within a certain percentage error rate and with a given level of confidence. For predicting the onset of sepsis, Guan *et al*. [36] and Lauritsen *et al*. [37] showed how model accuracy varies by the amount of time from when the model makes predictions to when sepsis occurs. Sepsis is an acute condition, whereas asthma is a chronic condition.

Limitations

This study has five limitations:

1) This study was performed for one health care system. In the future, we plan to use other health care systems' data to perform similar error and timeliness analyses on predicting asthma hospital visits [38,39].
2) This study shows that many false-positive predictions made by our model could be truly inaccurate. While this study does not examine the factors that could cause our model to make incorrect predictions, future work may investigate these factors to help improve model performance.
3) The Premanage data set covers every hospital in Washington state and many other American hospitals outside of Washington state, but not every hospital in the United States. Consequently, our computational results on asthma hospital visits outside of the UWM in 2019 could miss a small number of asthma patients in 2018 who encountered asthma hospital visits in 2019 at other American hospitals that are outside of the UWM and whose data are unavailable in Premanage.
4) The surrogates of poor outcomes were computed on the UWM data. Consequently, our computational results on the three surrogates of poor outcomes miss the asthma patients in 2018 who had surrogates of poor outcomes outside of the UWM.
5) This study computed the number of days of advanced warning for asthma hospital visit that our model gave on a patient for the first time. This number reflected how early before a poor outcome occurred the care manager would be prompted for the first time to consider giving the patient preventive interventions. It is currently unknown how likely the care manager would take actions after receiving such a warning. This is worth studying in future work.

Conclusions

This study analyzed the errors and the timeliness of the risk warnings given by our model for predicting asthma hospital visits. Our results show that our model gave timely risk warnings for most asthma patients with poor outcomes. 29.01% (380/1,310) of asthma patients for whom our model gave false-positive predictions had asthma hospital visits somewhere else during the successive 12 months or surrogates of poor outcomes, and hence were reasonable candidates for preventive interventions. There is still significant room for improving our model to give more accurate and more timely risk warnings, e.g., by using predictive and comprehensible temporal features semi-automatically extracted from longitudinal medical data [35,40,41].

Acknowledgments

We thank Brian Kelly for helpful discussions. GL was 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

XZ took part in the study design and the literature review, did the computer coding and the experiments, and wrote the first draft of the paper. GL conceptualized and designed the study, performed the literature review, and rewrote the entire paper. Both authors read and approved the final manuscript.

Conflicts of interest

None declared.

Abbreviations

ED: emergency department
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10: International Classification of Diseases, Tenth Revision
UWM: University of Washington Medicine
XGBoost: extreme gradient boosting

Appendix

Table 1. The summary statistics of the clinical characteristics and the demographics of the UWM patients with asthma during 2011-2017.

Variable	Data instances tied to asthma hospital visits at the UWM in the successive 12 months (*N*=1,184), *n* (%)	Data instances tied to no asthma hospital visit at the UWM in the successive 12 months (*N*=67,060), *n* (%)	Data instances (*N*=68,244), *n* (%)
Age (years)			
<40	466 (39.36)	22,993 (34.29)	23,459 (34.38)
40 to 65	583 (49.24)	33,306 (49.67)	33,889 (49.66)
65+	135 (11.40)	10,761 (16.05)	10,896 (15.97)
Gender			
Male	551 (46.54)	23,647 (35.26)	24,198 (35.46)
Female	633 (53.46)	43,413 (64.74)	44,046 (64.54)
Race			
White	507 (42.82)	47,240 (70.44)	47,747 (69.97)
Black or African American	520 (43.92)	7,900 (11.78)	8,420 (12.34)
Asian	96 (8.11)	5,625 (8.39)	5,721 (8.38)
American Indian or Alaska native	32 (2.70)	1,326 (1.98)	1,358 (1.99)
Native Hawaiian or other Pacific islander	14 (1.18)	659 (0.98)	673 (0.99)
Unknown or not reported	15 (1.27)	4,310 (6.43)	4,325 (6.34)
Ethnicity			
Non-Hispanic	1,062 (89.70)	55,247 (82.38)	56,309 (82.51)
Hispanic	82 (6.93)	3,444 (5.14)	3,526 (5.17)
Unknown or not reported	40 (3.38)	8,369 (12.48)	8,409 (12.32)
Insurance			
Private	424 (35.81)	39,585 (59.03)	40,009 (58.63)
Public	756 (63.85)	28,031 (41.80)	28,787 (42.18)
Self-paid or charity	65 (5.49)	1,301 (1.94)	1,366 (2.00)
Number of years since the first asthma-related visit in the UWM data set			
≤3	986 (83.28)	59,887 (89.30)	60,873 (89.20)
>3	198 (16.72)	7,173 (10.70)	7,371 (10.80)
Asthma medication prescription			

Short-acting inhaled beta-2 agonist	1,010 (85.30)	46,798 (69.79)	47,808 (70.05)
Inhaled corticosteroid	626 (52.88)	28,263 (42.15)	28,889 (42.33)
Long-acting beta-2 agonist and inhaled corticosteroid combination	499 (42.15)	21,516 (32.08)	22,015 (32.26)
Systemic corticosteroid	614 (51.86)	18,085 (26.97)	18,699 (27.40)
Long-acting beta-2 agonist	374 (31.59)	11,919 (17.77)	12,293 (18.01)
Leukotriene modifier	201 (16.98)	7,970 (11.88)	8,171 (11.97)
Mast cell stabilizer	4 (0.34)	43 (0.06)	47 (0.07)
Comorbidity			
Anxiety or depression	372 (31.42)	19,513 (29.10)	19,885 (29.14)
Gastroesophageal reflux	238 (20.10)	12,053 (17.97)	12,291 (18.01)
Allergic rhinitis	172 (14.53)	11,277 (16.82)	11,449 (16.78)
Obesity	177 (14.95)	7,668 (11.43)	7,845 (11.50)
Sinusitis	89 (7.52)	7,172 (10.69)	7,261 (10.64)
Sleep apnea	88 (7.43)	4,468 (6.66)	4,556 (6.68)
Eczema	66 (5.57)	3,825 (5.70)	3,891 (5.70)
Chronic obstructive pulmonary disease	133 (11.23)	3,693 (5.51)	3,826 (5.61)
Cystic fibrosis	1 (0.08)	60 (0.09)	61 (0.09)
Bronchopulmonary dysplasia	0 (0.00)	1 (0.00)	1 (0.00)
Smoking status			
Former smoker	221 (18.67)	15,309 (22.83)	15,530 (22.76)
Current smoker	255 (21.54)	13,826 (20.62)	14,081 (20.63)
Never smoker or unknown	708 (59.80)	37,925 (56.55)	38,633 (56.61)

Table 2. The summary statistics of the clinical characteristics and the demographics of the UWM patients with asthma in 2018.

Variable	Data instances tied to asthma hospital visits at the UWM in the successive 12 months (*N*=218), *n* (%)	Data instances tied to no asthma hospital visit at the UWM in the successive 12 months (*N*=14,426), *n* (%)	Data instances (*N*=14,644), *n* (%)
Age (years)			
<40	77 (35.3)	4,746 (32.90)	4,823 (32.94)
40 to 65	111 (50.9)	6,683 (46.33)	6,794 (46.39)
65+	30 (13.8)	2,997 (20.78)	3,027 (20.67)
Gender			
Male	100 (45.9)	5,138 (35.62)	5,238 (35.77)
Female	118 (54.1)	9,288 (64.38)	9,406 (64.23)
Race			
White	110 (50.5)	10,103 (70.03)	10,213 (69.74)
Black or African American	79 (36.2)	1,491 (10.34)	1,570 (10.72)
Asian	18 (8.3)	1,307 (9.06)	1,325 (9.05)
American Indian or Alaska native	8 (3.7)	273 (1.89)	281 (1.92)
Native Hawaiian or other Pacific islander	2 (0.9)	129 (0.89)	131 (0.89)
Unknown or not reported	1 (0.5)	1,123 (7.78)	1,124 (7.68)
Ethnicity			
Non-Hispanic	196 (89.9)	12,370 (85.75)	12,566 (85.81)
Hispanic	20 (9.2)	830 (5.75)	850 (5.80)
Unknown or not reported	2 (0.9)	1,226 (8.50)	1,228 (8.39)
Insurance			
Private	108 (49.5)	10,692 (74.12)	10,800 (73.75)
Public	182 (83.5)	7,841 (54.35)	8,023 (54.79)
Self-paid or charity	25 (11.5)	459 (3.18)	484 (3.31)

Number of years since the first asthma-related visit in the UWM data set			
≤3	124 (56.9)	10,442 (72.38)	10,566 (72.15)
>3	94 (43.1)	3,984 (27.62)	4,078 (27.85)
Asthma medication prescription			
Short-acting inhaled beta-2 agonist	164 (75.2)	9,540 (66.13)	9,704 (66.27)
Inhaled corticosteroid	108 (49.5)	6,069 (42.07)	6,177 (42.18)
Long-acting beta-2 agonist and inhaled corticosteroid combination	83 (38.1)	4,425 (30.67)	4,508 (30.78)
Systemic corticosteroid	120 (55.1)	4,043 (28.03)	4,163 (28.43)
Long-acting beta-2 agonist	62 (28.4)	2,456 (17.02)	2,518 (17.19)
Leukotriene modifier	46 (21.1)	2,130 (14.77)	2,176 (14.86)
Mast cell stabilizer	1 (0.5)	13 (0.09)	14 (0.10)
Comorbidity			
Anxiety or depression	62 (28.4)	4,284 (29.70)	4,346 (29.68)
Gastroesophageal reflux	46 (21.1)	2,611 (18.10)	2,657 (18.14)
Allergic rhinitis	26 (11.9)	2,069 (14.34)	2,095 (14.31)
Obesity	25 (11.5)	1,579 (10.95)	1,604 (10.95)
Sinusitis	15 (6.9)	1,357 (9.41)	1,372 (9.37)
Sleep apnea	24 (11.0)	1,475 (10.22)	1,499 (10.24)
Eczema	11 (5.1)	732 (5.07)	743 (5.07)
Chronic obstructive pulmonary disease	30 (13.8)	902 (6.25)	932 (6.36)
Cystic fibrosis	0 (0.0)	17 (0.12)	17 (0.12)
Bronchopulmonary dysplasia	0 (0.0)	4 (0.03)	4 (0.03)
Smoking status			
Former smoker	41 (18.8)	3,453 (23.94)	3,494 (23.86)
Current smoker	49 (22.5)	3,193 (22.13)	3,242 (22.14)
Never smoker or unknown	128 (58.7)	7,780 (53.93)	7,908 (54.00)

References

1. Chronic respiratory diseases: asthma. World Health Organization. 2021. https://www.who.int/news-room/q-a-detail/chronic-respiratory-diseases-asthma [accessed March 22, 2022].
2. Most recent national asthma data. Centers for Disease Control and Prevention. 2021. https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm [accessed March 22, 2022].
3. Nurmagambetov T, Kuwahara R, Garbe P. The economic burden of asthma in the United States, 2008-2013. Ann Am Thorac Soc 2018;15(3):348-356. PMID:29323930
4. 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):1173-1180. PMID:9563736
5. Mays GP, Claxton G, White J. Managed care rebound? Recent changes in health plans' cost containment strategies. Health Aff (Millwood). 2004;Suppl Web Exclusives:W4-427-436. PMID:15451964
6. 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):124-131. PMID:24395944
7. 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):436-440. PMID:10069877
8. 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):1029-1035. PMID:10790458
9. 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.
10. Axelrod RC, Vogel D. Predictive modeling in health plans. Dis Manag Health Outcomes 2003;11(12):779-787. doi:10.2165/00115677-200311120-00003
11. 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. 785-794. doi:10.1145/2939672.2939785

12. Tong Y, Messinger AI, Wilcox AB, Mooney SD, Davidson GH, Suri P, Luo G. Forecasting future asthma hospital encounters of patients with asthma in an academic health care system: predictive model development and secondary analysis study. J Med Internet Res 2021;23(4):e22796. PMID:33861206
13. 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):538-547. PMID:12921231
14. 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):113-119. PMID:9161058
15. 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):838-846. PMID:27044486
16. 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
17. 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):677-682. PMID:19728204
18. 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):1151-1161. PMID:17573515
19. 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):1145-1155. PMID:16870656
20. 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):918-924. PMID:16608939
21. 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):321-328. PMID:15152702
22. 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):1942-1952.e15. PMID:29454163
23. 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
24. 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
25. 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):1156-1165. PMID:20472862
26. 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
27. Longman JM, Passey ME, Ewald DP, Rix E, Morgan GG. Admissions for chronic ambulatory care sensitive conditions - a useful measure of potentially preventable admission? BMC Health Serv Res 2015;15:472. PMID:26475293
28. Johnston JJ, Longman JM, Ewald DP, Rolfe MI, Diez Alvarez S, Gilliland AHB, Chung SC, Das SK, King JM, Passey ME. Validity of a tool designed to assess the preventability of potentially preventable hospitalizations for chronic conditions. Fam Pract 2020;37(3):390-394. PMID:31848589
29. Zhang X, Luo G. Error analysis of machine learning predictions on asthma hospital encounters. J Allergy Clin Immunol 2022;149(2): supplement, AB47. doi:10.1016/j.jaci.2021.12.184
30. Howell D, Rogers L, Kasarskis A, Twyman K. Comparison and validation of algorithms for asthma diagnosis in an electronic medical record system. Ann Allergy Asthma Immunol 2022;128(6):677-681.e7. PMID:35367347
31. Collective Medical Technologies Inc. Collective Medical and Consonus Healthcare announce partnership to improve postacute transitions of care. 2018. https://collectivemedical.com/resources/press-release/collective-medical-and-consonus-healthcare-announce-partnership-to-improve-post-acute-transitions-of-care [accessed March 22, 2022].
32. 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
33. Zhang X, Luo G. Ranking rule-based automatic explanations for machine learning predictions on asthma hospital encounters in patients with asthma: retrospective cohort study. JMIR Med Inform 2021;9(8):e28287. PMID:34383673
34. Wiens J, Guttag JV, Horvitz E. Patient risk stratification with time-varying parameters: a multitask learning approach. J Mach Learn Res 2016;17(79):1-23.

35. Wang T, Jin F, Hu Y, Cheng Y. Early predictions for medical crowdfunding: a deep learning approach using diverse inputs. 2019. https://arxiv.org/abs/1911.05702 [accessed March 22, 2022].
36. Guan Y, Wang X, Chen X, Yi D, Chen L, Jiang X. Assessment of the timeliness and robustness for predicting adult sepsis. iScience 2021;24(2):102106. PMID:33659874
37. Lauritsen SM, Kalør ME, Kongsgaard EL, Lauritsen KM, Jørgensen MJ, Lange J, Thiesson B. Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif Intell Med 2020;104:101820. PMID:32498999
38. 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 healthcare system: secondary analysis. JMIR Med Inform 2020;8(11):e22689. PMID:33164906
39. 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
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. 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