

A Roadmap for Boosting Model Generalizability for Predicting Hospital Encounters for Asthma

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
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

In the United States, ~9% of people have asthma. Each year, asthma incurs high healthcare cost and many hospital encounters covering 1.8 million emergency room visits and 439,000 hospitalizations. A small percentage of patients with asthma use most healthcare resources. To improve outcomes and cut resource use, many healthcare systems use predictive models to prospectively find high-risk patients and enroll them in care management for preventive care. For maximal benefit from costly care management with limited service capacity, only patients at the highest risk should be enrolled. Yet, prior models built by others miss >50% of true highest-risk patients and mislabel many low-risk patients as high risk, leading to suboptimal care and wasted resources. To address this issue, we recently built three site-specific models to predict hospital encounters for asthma and gained up to 11%+ better performance. But, these models do not generalize well across sites and patient subgroups, creating two gaps before translating these models into clinical use. This paper points out these two gaps and outlines two corresponding solutions: a) a new machine learning technique to create cross-site generalizable predictive models to accurately find high-risk patients, and b) a new machine learning technique to automatically raise model performance for poorly performing subgroups while maintaining model performance on other subgroups. This gives a roadmap for future research.

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

Introduction

Asthma care management and our prior work on predictive modeling

In the United States, ~9% of people have asthma [1-3]. Each year, asthma incurs \$56 billion of healthcare cost [4] and many hospital encounters covering 1.8 million emergency room visits and 439,000 hospitalizations [1]. As is the case with many chronic diseases, a small percentage of patients with asthma use most healthcare resources [5,6]. The top 1% of patients spend 25% of healthcare costs. The top 20% spend 80% [5,7]. An effective approach is urgently in need to prospectively identify high-risk patients and intervene early to avoid health decline, improve outcomes, and cut resource use. Most major employers purchase and nearly all private health plans offer care management services for preventive care [8-10]. Care management is a collaborative process to assess, coordinate, plan, implement, evaluate, and monitor the services and options to meet the health and service needs of a patient [11]. A care management program employs care managers to call patients regularly to assess their status, arrange doctor appointments, and coordinate health-related services. Proper use of care management can cut down hospital encounters by up to 40% [10,12-17]; lower healthcare cost by up to 15% [13-18]; and improve patient satisfaction, quality of life, and adherence to treatment by 30-60% [12]. Care management can cost >\$5,000 per patient per year [13] and normally enrolls no more than 3% of patients [7] due to resource limits.

Correctly finding high-risk patients to enroll is crucial for effective care management. Currently, the best method to identify high-risk patients is to use models to predict each patient's risk [19]. Many health plans, such as those in 9 of 12 metropolitan communities [20], and many healthcare systems [21] use this method for care management. For patients predicted to have the highest risk, care managers manually review patients' medical records, consider factors like social dimensions, and make enrollment decisions. Yet, prior models built by others miss >50% of true highest-risk patients and mislabel many low-risk patients as high risk [5,12,22-36]. This makes enrollment align poorly with patients who would benefit most from care management [12], leading to suboptimal care and higher costs. As the patient population is large, a small boost in model performance will benefit many patients and produce a large positive impact. Of the top 1% asthma patients who would incur the highest costs, for every 1% more whom we could find and enroll, we could save up to \$21 million more in asthma care every year as well as improve outcomes [5,26,27].

To address the issue of low model performance, we recently built three site-specific models to predict whether a patient with asthma would incur any hospital encounter for asthma in the subsequent 12 months, one model for each of the three healthcare systems the University of Washington Medicine (UWM), Intermountain Healthcare (IH), and Kaiser Permanente Southern California (KPSC) [21,37,38]. Each prior model that others built for a comparable outcome [5,26-34] had an area under the receiver operating characteristic curve (AUC) that was ≤ 0.79 and a sensitivity that was $\leq 49\%$. Our models raised the AUC to 0.9 and the sensitivity to 70% on the UWM data [21], the AUC to 0.86 and the sensitivity to 54% on IH data [37], and the AUC to 0.82 and the sensitivity to 52% on KPSC data [38].

Our eventual goal is to translate our models into clinical use. Yet, despite major progress, our models do not generalize well across sites and patient subgroups, and two gaps remain.

Gap 1: The site-specific models have suboptimal generalizability when applied to the other sites

Each of our models was built for one site. As is typical in predictive modelling [39,40], when applied to the other sites, the site-specific model had AUC drops of up to 4.1% [38] potentially degrading care management enrollment decisions. One can do transfer learning using other source healthcare systems' raw data to boost model performance for the target healthcare system [41-45], but healthcare systems are seldom willing to share raw data. Research networks [46-48] mitigate the problem,

but do not solve it. Many healthcare systems are not in any network. Healthcare systems in the network share raw data of finite attributes. Our prior model-based transfer learning approach [49] requires no raw data from other healthcare systems. But, it does not control the number of features (independent variables) used in the final model for the target site, creating difficulty to build the final model for the target site for clinical use. Consequently, it is never implemented in computer code.

Gap 2: The models exhibit large performance gaps when applied to specific patient subgroups

Our models performed up to 8% worse on black patients. This is a typical barrier in machine learning, where many models exhibit large subgroup performance gaps, e.g., of up to 38% [50-57]. No existing tool for auditing model bias and fairness [58,59] has been applied to our models. Currently, it is unknown how our models perform on key patient subgroups defined by independent variables such as race, ethnicity, and insurance type. In other words, it is unknown how our models perform for different races, different ethnicities, and patients using different types of insurance. Large performance gaps among patient subgroups can lead to care inequity and should be avoided.

Many methods to improve fairness in machine learning exist [50-52]. These methods usually boost model performance on some subgroups at the price of lowering both model performance on others and the overall model performance [50-52]. Lowering the overall model performance is undesired [51,57]. Due to the large patient population, even a 1% drop in the overall model performance could potentially degrade many patients' outcomes. Chen et al. [57] cut model performance gaps among subgroups by collecting more training data and adding additional features, both of which are often difficult or infeasible to do. For classifying images via machine learning, Goel et al.'s method [55] raised the overall model performance and also cut model performance gaps among subgroups of a value of the dependent variable, not among subgroups defined by independent variables. The dependent variable is also known as the outcome or the prediction target. An example of the dependent variable is whether a patient with asthma will incur any hospital encounter for asthma in the subsequent 12 months. The independent variables are also known as features. Race, ethnicity, and insurance type are three examples of independent variables. Many machine learning techniques to handle imbalanced classes exist [60,61]. There, subgroups are defined by the dependent variable rather than by independent variables.

Contributions of this paper

To fill the two gaps on suboptimal model generalizability and let more high-risk patients obtain appropriate and equitable preventive care, the paper makes two contributions giving a roadmap for future research:

- 1) To address the first gap, we outline a new machine learning technique to create cross-site generalizable predictive models to accurately find high-risk patients. This is to cut model performance drop across sites.
- 2) To address the second gap, we outline a new machine learning technique to automatically raise model performance for poorly performing subgroups while maintaining model performance on other subgroups. This is to cut model performance gaps among patient subgroups and reduce care inequity.

In the following, we describe the main ideas of our proposed new machine learning techniques.

Machine Learning Technique for Creating Cross-site Generalizable Predictive Models to Accurately Find High-risk Patients

Our prior models

In our prior work [21,37,38], for each of the three healthcare systems (sites) KPSC, IH, and the UWM, we checked >200 candidate features and used the site's data to build a full site-specific extreme gradient boosting (XGBoost) model to predict hospital encounters for asthma. XGBoost [62] automatically chose from the candidate features the features to be used in the model, computed their importance values, and ranked them in descending order of these values. The top (~20) features with importance values $\geq 1\%$ have nearly all of the predictive power of all (on average ~140) features used in the model [21,37,38]. Although some lower-ranked features are unavailable at other sites, each top feature, such as the number of the patient's asthma-related emergency room visits in the prior 12 months, is computed using (e.g., diagnosis and encounter) attributes routinely collected by almost every American healthcare system that uses electronic medical records. Using the top features and the site's data, we built a simplified XGBoost model. It, but not the full model, can be applied to other sites. The simplified model performed similarly to the full model at the site. But, when applied to another site, even after being re-trained on its data, the simplified model performed up to 4.1% worse than the full model built specifically for it, as distinct sites have only partially overlapping top features [21,37,38].

Building cross-site generalizable models

To ensure that the same variable is called the same name at different sites and the variable's content is recorded in the same way across these sites, we convert the data sets at all source sites and the target site into the Observational Medical Outcomes

Partnership (OMOP) common data model [63] and its linked standardized terminologies [64]. If needed, the data model is extended to cover the variables that are not included in the original data model, but exist in the data sets.

Our goal is to build cross-site generalizable models fulfilling two conditions. First, the model uses a moderate number of features. Controlling the number of features used in the model would ease future clinical deployment of the model. Second, a separate component or copy of the model is initially built at each source site. When applied to the target site and possibly after being re-trained on its data, the model performs similarly to the full model built specifically for it. To reach our goal for the case of IH and the UWM being the source sites and KPSC being the target site, we proceed in two steps (Figure 1). In step 1, we combine the top features found at each source site. For each source site, we use the combined top features, its data, and the machine learning algorithm used to build its full model to build an expanded simplified model. Compared with the original simplified model built for the site, the expanded simplified model uses more features with predictive power and tends to generalize better across sites. In step 2, we do model-based transfer learning to further boost model performance. For each data instance of the target site, we apply each source site’s expanded simplified model to the data instance, compute a prediction result, and use it as a new feature. For the target site, we use its data, the combined top features found at the source sites, and the new features to build its final model.

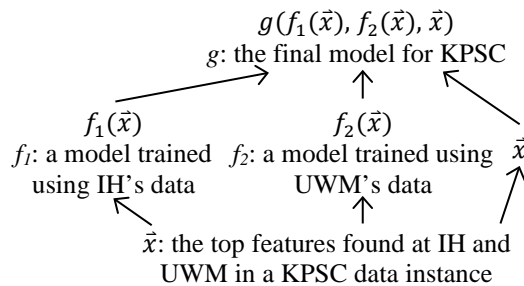


Figure 1. Our method to build cross-site generalizable models. IH: Intermountain Healthcare. KPSC: Kaiser Permanente Southern California. UWM: University of Washington Medicine.

To reach our goal for the case that IH or the UWM is the target site and KPSC is one of the source sites, we need to address the issue that the claim-based features used at KPSC [38] are unavailable at IH, the UWM, and many other healthcare systems with no claim data. At KPSC, we drop these features and use the other candidate features to build a site-specific model and re-compute the top features. This helps reach the effect that the top features found at each of KPSC, IH, and the UWM are available at all three sites and almost every other American healthcare system that uses electronic medical record systems. In the unlikely case that any re-computed top feature at KPSC violates this, we skip the feature when building cross-site generalizable models.

Our method to build cross-site generalizable models can handle all kinds of prediction targets, features, and models used at the source and target sites. Given a distinct prediction target, if some top features found at a source site are unavailable at many American healthcare systems using electronic medical record systems, we can use the drop→re-compute→skip approach shown above to handle these features. Also, at any source site, if the machine learning algorithm used to build the full site-specific model is like XGBoost [62] or random forest that automatically computes feature importance values, we can use the top features with the highest importance values. Otherwise, if the algorithm used to build the full model does not automatically compute feature importance values, we can use an automatic feature selection method [65] like the information gain method to choose the top features. Alternatively, we can use XGBoost or random forest to build a model, automatically compute feature importance values, and choose the top features with the highest importance values.

Our new model-based transfer learning approach waives the need for source sites’ raw data. Healthcare systems are more willing to share with others trained models than raw data. A model trained using the data of a source site contains much information that is useful for the prediction task at the target site. This information offers much value when the target site has insufficient data for model training. If the target site is large, this information can still be valuable. Distinct sites have differing data pattern distributions. A pattern that matches a small percentage of patients and is difficult to identify at the target site could match a larger percentage of patients and be easier to identify at one of the source sites. In this case, its expanded simplified model could incorporate the pattern through model training to better predict the outcomes of certain types of patients, which is difficult to do using only the information from the target site but no information from the source sites. Thus, we expect that compared with just re-training a source site’s expanded simplified model on the target site’s data, doing model-based transfer learning in step 2 could lead to a better performing final model for the target site.

When the target site goes beyond IH, the UWM, and KPSC, we can use IH, the UWM, and KPSC as the source sites and have more top features to combine. This would make our cross-site models generalize even better.

Machine Learning Technique for Automatically Raising Model Performance for Poorly Performing Patient Subgroups while Maintaining Model Performance on Other Subgroups to Reduce Care Inequity

We ask several clinical experts to identify several patient subgroups of great interest to clinicians (e.g., by race, ethnicity, or insurance type) through discussion. These subgroups are not necessarily mutually exclusive of each other. Each subgroup is defined by one or more attribute values. Given a predictive model built on a training set, we compute and show model performance on each subgroup on the test set [58,59]. Machine learning needs enough training data to work well. Often, the model performs much worse on a small subgroup than on a large subgroup [50,52]. After identifying one or more target subgroups where the model performs much worse than on other subgroups [51], we use a new dual-model approach to raise model performance on the target subgroups while maintaining model performance on other subgroups.

More specifically, given n target patient subgroups, we sort them as G_i ($1 \leq i \leq n$) in ascending order of size and oversample them based on n integers r_i ($1 \leq i \leq n$) satisfying $r_1 \geq r_2 \geq \dots \geq r_n > 1$. As Figure 2 shows, for each training instance in G_1 , we make r_1 copies of it including itself. For each training instance in $G_j - \bigcup_{i=1}^{j-1} G_i$ ($2 \leq j \leq n$), we make r_j copies of it including itself. Intuitively, the smaller the i ($1 \leq i \leq n$) and thus G_i , the more aggressive oversampling is needed on G_i for machine learning to work well on it. The sorting ensures that if a training instance appears in ≥ 2 target subgroups, copies are made for it based on the largest r_i of these subgroups. If needed, we could use one set of r_i 's for training instances with bad outcomes, and another set of r_i 's for training instances with good outcomes [66]. $G = \bigcup_{i=1}^n G_i$ is the union of the n target subgroups. Using the training instances outside G , the copies made for the training instances in G , and an automatic machine learning model selection method like our formerly developed one [67], we optimize the AUC on G , automatically select the values of r_i ($1 \leq i \leq n$), and train a second model. As is typical in using oversampling to improve fairness in machine learning, compared with the original model, the second model tends to perform better on G and worse on the patients outside G [51,66] because oversampling increases the percentage of training instances in G and decreases the percentage of training instances outside G . To avoid running into the case of having insufficient data for model training, no undersampling is performed on the training instances outside G . We use the original model to make predictions on the patients outside G , and the second model to make predictions on the patients in G . In this way, we can raise model performance on G without lowering either model performance on the patients outside G or the overall model performance. We use all patients' data instead of only the training instances in G to train the second model. Otherwise, the second model may perform poorly on G due to insufficient training data in G [51]. For a similar reason, we choose to not use decoupled classifiers, where a separate classifier is trained for each subgroup using only that subgroup's data [51], on the target subgroups [57].

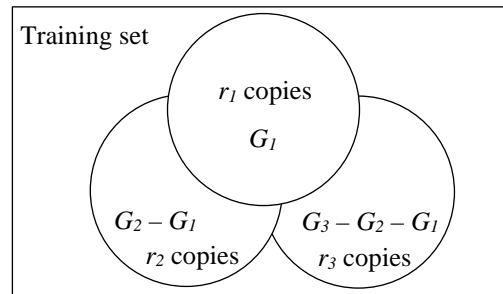


Figure 2. Oversampling for three target patient subgroups G_1 , G_2 , and G_3 .

The above discussion focuses on the case that the original model is built on one site's data without using any other site's information. When the original model is a cross-site generalizable model built for the target site using the method in the "Building cross-site generalizable models" section and models trained at the source sites, to raise model performance on the target patient subgroups, we change the way to build the second model for the target site by proceeding in two steps (Figure 3). In step 1, we combine the top features found at each source site. Recall that G is the union of the n target subgroups. For each source site, we oversample the target subgroups in the way mentioned above; optimize the AUC on G at the source site; and use its data both in and outside G , the combined top features, and the machine learning algorithm used to build its full model to build a second expanded simplified model. In step 2, we do model-based transfer learning to incorporate useful information from the source sites. For each data instance of the target site, we apply each source site's second expanded simplified model to

the data instance, compute a prediction result, and use it as a new feature. For the target site, we oversample the target subgroups in the way mentioned above; optimize the AUC on G at the target site; and use its data both in and outside G , the combined top features found at the source sites, and the new features to build the second model for it. For each i ($1 \leq i \leq n$), each of the source and target sites could use a distinct oversampling ratio r_i .

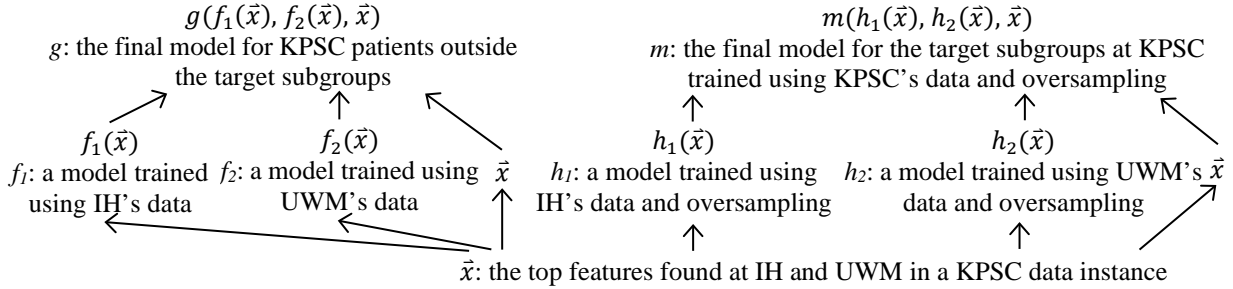


Figure 3. Our method to boost a cross-site generalizable model’s performance on the target patient subgroups. IH: Intermountain Healthcare. KPSC: Kaiser Permanente Southern California. UWM: University of Washington Medicine.

Discussion

Predictive models differ by diseases and other factors. Yet, our proposed machine learning techniques are general and depend on no specific disease, patient cohort, or healthcare system. Given a new data set with a differing prediction target, disease, patient cohort, set of healthcare systems, or set of variables, one can use our proposed machine learning techniques to improve model generalizability across sites, as well as to boost model performance on poorly performing patient subgroups while maintaining model performance on others. For instance, we can use our proposed machine learning techniques to improve model performance for predicting other outcomes such as adherence to treatment [68] and no show [69]. This will help target resources, such as interventions to improve adherence to treatment [68] and reminders by phone calls to reduce no shows [69]. Care management is widely adopted to manage patients with chronic obstructive pulmonary disease, patients with diabetes, and patients with heart disease [6], where our proposed machine learning techniques can also be used. Our proposed predictive models are based on the OMOP common data model [63] and its linked standardized terminologies [64], which standardize administrative and clinical variables from at least 10 large healthcare systems in the United States [47,70]. Our proposed predictive models apply to those healthcare systems and others using OMOP.

Conclusions

To better identify patients likely to benefit most from asthma care management, we recently built the most accurate models to date to predict hospital encounters for asthma. But, these models do not generalize well across sites and patient subgroups, creating two gaps before translating these models into clinical use. This paper points out these two gaps and outlines two corresponding solutions, giving a roadmap for future research. The principles of our proposed machine learning techniques generalize to many other clinical predictive modeling tasks.

Acknowledgments

We thank Flory L. Nkoy for useful 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.

Conflicts of interest

None declared.

Abbreviations

- AUC: area under the receiver operating characteristic curve
- IH: Intermountain Healthcare
- KPSC: Kaiser Permanente Southern California
- OMOP: Observational Medical Outcomes Partnership
- UWM: University of Washington Medicine
- XGBoost: extreme gradient boosting

References

1. Asthma. <http://www.cdc.gov/nchs/fastats/asthma.htm>, 2021.
2. Akinbami LJ, Moorman JE, Liu X. Asthma prevalence, health care use, and mortality: United States, 2005-2009. *Natl Health Stat Report* 2011;(32):1-14. PMID:21355352
3. Akinbami LJ, Moorman JE, Bailey C, Zahran HS, King M, Johnson CA, Liu X. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. *NCHS Data Brief* 2012;(94):1-8. PMID:22617340
4. Asthma in the US. <http://www.cdc.gov/vitalsigns/asthma>, 2021.
5. 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
6. Duncan I. *Healthcare Risk Adjustment and Predictive Modeling*, 2nd. Winsted, CT: ACTEX Publications Inc.; 2018. ISBN:1635884136
7. Axelrod RC, Vogel D. Predictive modeling in health plans. *Disease Manage Health Outcomes* 2003;11(12):779-787. doi:10.2165/00115677-200311120-00003
8. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, Blumenthal D. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. *J Gen Intern Med* 2007;22 Suppl 3:391-395. PMID:18026807
9. Nelson L. Lessons from Medicare's demonstration projects on disease management and care coordination. https://www.cbo.gov/sites/default/files/112th-congress-2011-2012/workingpaper/WP2012-01_Nelson_Medicare_DMCC_Demonstrations_1.pdf, 2012.
10. Caloyer 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
11. Commission for Case Manager Certification. Definition and philosophy of case management. <https://ccmcertification.org/about-ccmc/about-case-management/definition-and-philosophy-case-management>, 2021.
12. Levine SH, Adams J, Attaway K, Dorr DA, Leung M, Popescu P, Rich J. Predicting the financial risks of seriously ill patients. California HealthCare Foundation, 2011. <http://www.chcf.org/publications/2011/12/predictive-financial-risks>.
13. Rubin RJ, Dietrich KA, Hawk AD. Clinical and economic impact of implementing a comprehensive diabetes management program in managed care. *J Clin Endocrinol Metab* 1998;83(8):2635-2642. PMID:9709924
14. 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
15. Kelly CS, Morrow AL, Shults J, Nakas N, Strobe GL, Adelman RD. Outcomes evaluation of a comprehensive intervention program for asthmatic children enrolled in Medicaid. *Pediatrics* 2000;105(5):1029-1035. PMID:10790458
16. Axelrod RC, Zimbardo 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.
17. Dorr DA, Wilcox AB, Brunner CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. *J Am Geriatr Soc* 2008;56(12):2195-2202. PMID:19093919
18. Beaulieu N, Cutler DM, Ho K, Isham G, Lindquist T, Nelson A, O'Connor P. The business case for diabetes disease management for managed care organizations. *Forum Health Econ Policy* 2006;9(1):1-37. doi:10.2202/1558-9544.1072
19. Curry N, Billings J, Darin B, Dixon J, Williams M, Wennberg D. Predictive Risk Project Literature Review. London: King's Fund. http://www.kingsfund.org.uk/sites/files/kf/field/field_document/predictive-risk-literature-review-june2005.pdf, 2005.
20. 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
21. 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
22. Ash A, McCall N. Risk assessment of military populations to predict health care cost and utilization. http://www.rti.org/pubs/tricare_riskassessment_final_report_combined.pdf, 2005.
23. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analyzing health care utilization and costs. *Annu Rev Public Health* 1999;20:125-144. PMID:10352853
24. Iezzoni LI. *Risk adjustment for measuring health care outcomes*, 4th ed. Chicago, IL: Health Administration Press; 2012. ISBN:1567934374
25. Weir S, Aweh G, Clark RE. Case selection for a Medicaid chronic care management program. *Health Care Financ Rev* 2008;30(1):61-74. PMID:19040174

26. 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
27. 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
28. 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
29. 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
30. 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
31. 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
32. 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
33. 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
34. 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
35. 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
36. 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
37. 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
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 health care system: secondary analysis. *JMIR Med Inform*, 2020;8(11):e22689. PMID:33164906
39. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, Moons KG. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56(9):826-832. PMID:14505766
40. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015;68(1):25-34. PMID:25441703
41. Wiens J, Guttat J, Horvitz E. A study in transfer learning: leveraging data from multiple hospitals to enhance hospital-specific predictions. *J Am Med Inform Assoc* 2014;21(4):699-706. PMID:24481703
42. Gong JJ, Sundt TM, Rawn JD, Guttat JV. Instance weighting for patient-specific risk stratification models. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 10-13, 2015; Sydney, NSW, Australia p. 369-378. doi:10.1145/2783258.2783397
43. Lee G, Rubinfeld I, Syed Z. Adapting surgical models to individual hospitals using transfer learning. In: *Proceedings of IEEE International Conference on Data Mining Workshops*. 2012 Presented at: IEEE International Conference on Data Mining Workshops; December 10, 2012; Brussels, Belgium p. 57-63. doi:10.1109/ICDMW.2012.93
44. Pan S, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22(10):1345-1359. doi:10.1109/TKDE.2009.191
45. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *Journal of Big Data* 2016;3:9. doi:10.1186/s40537-016-0043-6
46. Jayanthi A. Down the rabbit hole at Epic: 9 key points from the Users Group Meeting. <http://www.beckershospitalreview.com/healthcare-information-technology/down-the-rabbit-hole-at-epic-8-key-points-from-the-users-group-meeting.html>, 2016.
47. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong ICK, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li Y, Stang PE, Madigan D, Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015;216:574-578. PMID:26262116

48. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014;21(4):578-582. PMID:24821743
49. Luo G, Sward K. A roadmap for optimizing asthma care management via computational approaches. *JMIR Med Inform* 2017;5(3):e32. PMID:28951380
50. Oakden-Rayner L, Dunnmon J, Carneiro G, Ré C. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. In: *Proceedings of ACM Conference on Health, Inference, and Learning*. 2020 Presented at: ACM Conference on Health, Inference, and Learning; April 2-4, 2020; Toronto, Ontario, Canada, p. 151-159. doi:10.1145/3368555.3384468
51. Caton S, Haas C. Fairness in machine learning: a survey. *CoRR abs/2010.04053*, 2020.
52. Barocas S, Hardt M, Narayanan A. Fairness and machine learning: limitations and opportunities. <https://fairmlbook.org>, 2021.
53. DeVries T, Misra I, Wang C, van der Maaten L. Does object recognition work for everyone? In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2019 Presented at: IEEE Conference on Computer Vision and Pattern Recognition Workshops; June 16-20, 2019; Long Beach, CA, p. 52-59.
54. Buolamwini J, Gebre T. Gender shades: intersectional accuracy disparities in commercial gender classification. In: *Proceedings of Conference on Fairness, Accountability and Transparency*. 2018 Presented at: Conference on Fairness, Accountability and Transparency; February 23-24, 2018; New York, NY, p. 77-91.
55. Goel K, Gu A, Li Y, Ré C. Model patching: closing the subgroup performance gap with data augmentation. In: *Proceedings of the 9th International Conference on Learning Representations*. 2021 Presented at: International Conference on Learning Representations; May 3-7, 2021; Vienna, Austria, p. 1-30.
56. Seyyed-Kalantari L, Liu G, McDermott M, Chen IY, Ghassemi M. CheXclusion: fairness gaps in deep chest X-ray classifiers. *Pac Symp Biocomput* 2021;26:232-243. PMID:33691020
57. Chen IY, Johansson FD, Sontag DA. Why is my classifier discriminatory? In: *Proceedings of Annual Conference on Neural Information Processing Systems*. 2018 Presented at: NeurIPS'18; December 3-8, 2018; Montréal, Canada p. 3543-3554.
58. Saleiro P, Kuester B, Stevens A, Anisfeld A, Hinkson L, London J, Ghani R. Aequitas: a bias and fairness audit toolkit. *CoRR abs/1811.05577*, 2018.
59. Panigutti C, Perotti A, Panisson A, Bajardi P, Pedreschi D. FairLens: auditing black-box clinical decision support systems. *Inf Process Manag* 2021;58(5):102657. doi:10.1016/j.ipm.2021.102657
60. Branco P, Torgo L, Ribeiro RP. A survey of predictive modeling on imbalanced domains. *ACM Comput Surv* 2016;49(2):31. doi:10.1145/2907070
61. Kaur H, Pannu HS, Malhi AK. A systematic review on imbalanced data challenges in machine learning: applications and solutions. *ACM Comput Surv* 2020;52(4):79. doi:10.1145/3343440
62. 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
63. Observational Health Data Sciences and Informatics data standardization homepage. <https://www.ohdsi.org/data-standardization>, 2021.
64. Observational Health Data Sciences and Informatics standardized vocabularies homepage. <https://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:sidebar>, 2021.
65. 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
66. Rancic S, Radovanovic S, Delibasic B. Investigating oversampling techniques for fair machine learning models. In: *Proceedings of the 7th International Conference on Decision Support System Technology*. 2021 Presented at: International Conference on Decision Support System Technology; May 26-28, 2021; Loughborough, UK, p. 110-123.
67. 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
68. Kumamaru H, Lee MP, Choudhry NK, Dong YH, Krumme AA, Khan N, Brill G, Kohsaka S, Miyata H, Schneeweiss S, Gagne JJ. Using previous medication adherence to predict future adherence. *J Manag Care Spec Pharm* 2018;24(11):1146-1155. PMID:30362915
69. Chariatte V, Berchtold A, Akre C, Michaud PA, Suris JC. Missed appointments in an outpatient clinic for adolescents, an approach to predict the risk of missing. *J Adolesc Health* 2008;43(1):38-45. PMID:18565436
70. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012;19(1):54-60. PMID:22037893