APPLICATION OF MACHINE LEARNING AND CAUSAL INFERENCE APPROACHES TO BRONCHIOLITIS AND ASTHMA RESEARCH

by

Yoshihiko Raita, MD, MPH

A dissertation submitted to the faculty of Harvard Medical School
in Partial Fulfillment of
the Requirement for the Degree of Master of Medical Sciences in Clinical Investigation
(MMSCI)

Harvard University

Boston, Massachusetts

April, 2020

Area of Concentration: Respiratory disease/asthma/bronchiolitis/causal inference/machine learning

Project Advisor: Dr. Carlos A. Camargo, Jr.; Dr. Ajay K. Singh; Dr. Finnian R. McCausland; Dr. Amil M. Shah; Dr. Stuart N. Harris and Dr. Kohei Hasegawa

Primary Mentor: Dr. Kohei Hasegawa

Revised April 22nd 2020

TABLE OF CONTENTS……2

ACKNOWLEDGEMENTS……4

OVERVIEW……5

Paper 1: Machine learning-based prediction of acute severity in infants hospitalized for bronchiolitis: A multicenter prospective study

ABSTRACT……10
INTRODUCTION……11
METHODS……14
RESULTS……20
DISCUSSION……23
FIGURES……27
Figure 1. Prediction ability of the reference and machine learning models for positive pressure ventilation outcome in the overall cross-validation dataset……27
Figure 2. Prediction ability of the reference and machine learning models for intensive treatment outcome in the overall cross-validated dataset……29
TABLES……31
Table 1. Patient characteristics and clinical outcomes in 1,016 infants hospitalized for bronchiolitis……31
Table 2. Prediction performance of the reference, and machine learning models in infants hospitalized for bronchiolitis……33
Table 3. The number of actual and predicted outcomes of prediction models, according to the score of the reference model……34
SUPPLEMENTARY MATERIAL……35
Table E1. Principal investigators at the 17 participating sites in MARC-35……36
Table E2. Predictors and scores of the reference models in 1,016 infants hospitalized for bronchiolitis……37
Table E3. Proportion of missing data in 1,016 infants hospitalized for bronchiolitis……38
Figure E1. Variable importance plot of random forest model……39
Figure E2. Variable importance plot of gradient boosted decision tree model……41
REFERENCES……43

Paper 2: Risk of acute myocardial infarction and ischemic stroke in patients with asthma exacerbation: A population-based, self-controlled case series study

ABSTRACT……54
INTRODUCTION……56
METHODS……57
RESULTS……61
DISCUSSION……63
FIGURES……68

Figure 1. Example self-controlled case series timeline……68
Figure 2. Adjusted incidence rate ratio for acute myocardial infarction and ischemic stroke after asthma exacerbation……69
TABLES……70
Table 1. Baseline characteristics of 4,607 patients hospitalized with asthma exacerbation……70
Table 2. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke in patients with asthma exacerbation (n=4,607)……71
SUPPLEMENTARY MATERIAL……72
Figure E1. Example self-controlled case series timeline, excluding the pre-exposure (0-14 days before asthma exacerbation) period from the reference period……73
Table E1. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke in patients with asthma exacerbation, stratified by age (n=4,607)……74
Table E2. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by sex (n=4,607)……75
Table E3 Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by race/ethnicity (n=4,607)……76
Table E4 Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by primary health insurance (n=4607)……78
Table E5 Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by household income quartile (n=4,607)……79
Table E6 Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by state (n=4,607)……81
Table E7. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke in patients with asthma exacerbation, excluding patients with concurrent chronic obstructive pulmonary disease (n=4,030) ……83
Table E8. Incidence rate and adjusted incidence rate ratio (IRR) for acute myocardial infarction and ischemic stroke, excluding the pre-exposure period from the reference period (n=4,542)……84
Table E9. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke, excluding patients who died within 30 days of outcome (n=4,570)……85
Table E10. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke, excluding patients with prolonged hospitalization for asthma exacerbation (n=4,398)……86
REFERENCES……87
SUMMARY OF PAPER 1 AND PAPER 2 CONCLUSION……93
OVERALL DISCUSSION AND PERSPECTIVES……95

ACKNOWLEDGEMENT

First, I would like to express highest gratitude to my mentor Dr. Hasegawa for his invaluable and constructive mentoring through the development of the work. I am grateful for Dr. Hasegawa's persistent support, guidance and trust on me.

I would also like to extend my deep gratitude to Dr. Camargo for his constructive feedback, enthusiastic encouragement. This work would not have been possible without their support. I owe what I have achieved as an early-stage researcher to Drs. Singh and McCausland and faculty members of this program, who have provided this great opportunity and supported my educational goals. I would also like to acknowledge Ms. Cacioppo and O'Connor for arranging courses and associated activities during the program.

I would like to express my appreciation to the colleagues of the Emergency Medicine Network for their assistance. I thank the MARC-35 study hospitals and research personnel for their ongoing dedication to bronchiolitis and asthma research. I also thank Ms. Sullivan, Espinola (Massachusetts General Hospital, Boston, MA) as well as Dr. Togias (National Institute of Allergy and Infectious Diseases) for their helpful contributions to MARC-35 study. Finally, I would express a deep sense of gratitude to my wife Miwa for her constant love, encouragement and support.

Thank you.

Yoshihiko Raita

OVERVIEW

During the curriculum of Master of Medical Sciences Program at Harvard Medical School, I have been focusing on two inferential goals through respiratory disease research—1) prediction and 2) causal inference—which are central pillars of research questions as clinical investigators.(1) Bronchiolitis in infants and asthma in adults are two of the major respiratory diseases in industrialized nations. Indeed, bronchiolitis is the leading cause of infant hospitalization in the US, accounting for 107,000 infant hospitalizations each year.(2) Besides, previous literature suggests that bronchiolitis during infancy is one of the major risk factors for subsequent asthma.(3) Likewise, approximately 26 million Americans have asthma.(4) Asthma exacerbations account for a substantial proportion of the personal and societal burden, leading to approximately 340,000 hospitalizations annually.(5)

Paper 1: The literature has demonstrated a high variability in bronchiolitis management across the nation, which is, at least partially, attributable to the variable clinical course in this population.(6) Despite the development of prediction scoring models (e.g., logistic regression models),(1,4–6) identifying the subgroup of infants with bronchiolitis who require higher acuity care (e.g., positive pressure ventilation, intensive care unit [ICU] admission) remains an important challenge. The difficulty and uncertainty of predicting acute severity—and, consequently, the appropriate level of care for infants with bronchiolitis—are reflected by the well-documented variability in inpatient management across the nation.(3,7–9) Recently, machine learning models have gained increasing attention because of their advantages, such as the ability to incorporate high-order, nonlinear interactions between predictors and to yield more accurate and stable

predictions. While machine learning approaches may enhance the predictive ability,(10–12) little is known about their utility to predict acute severity outcomes in infants with bronchiolitis. We developed advanced machine learning models using the data from a well-characterized prospective cohort study.

Paper 2: Previous research has demonstrated that patients with *chronic* asthma have a high incidence of cardiovascular events.(13–15) Besides, the emerging evidence suggests that acute inflammatory processes (e.g., bacteremia,(16) pneumonia,(17) influenza virus infection,(18) Streptococcal infection,(19) and acute exacerbation of COPD(20)) have been linked to acute cardiovascular outcomes. However, despite the clinical and research importance of asthma exacerbations, little is known about its acute effect on cardiovascular outcomes. Herein, using large population-based data of four diverse states in the US, I conducted a self-controlled case series study to investigate the acute causal effect of asthma exacerbation on cardiovascular events.

Paper 1

Machine learning-based prediction of acute severity in infants hospitalized for bronchiolitis: A multicenter prospective study

Authors: Yoshihiko Raita, MD, MPH[1]; Carlos A. Camargo, Jr. MD, DrPH[1]; Charles G. Macias, MD, MPH[2]; Jonathan M. Mansbach, MD, MPH[3]; Pedro A. Piedra, MD[4]; Stephen C. Porter, MD, MPH, MSc[5,6]; Stephen J. Teach, MD, MPH[7]; and Kohei Hasegawa, MD, MPH[1]

Affiliations:

1. Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
2. Department of Pediatric Emergency Medicine, Rainbow Babies and Children's Hospital, OH, USA
3. Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
4. Departments of Molecular Virology and Microbiology and Pediatrics, Baylor College of Medicine, Houston, TX, USA
5. Department of Pediatrics, University of Cincinnati, College of Medicine, Cincinnati, OH
6. Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH
7. Division of Emergency Medicine and Department of Pediatrics, Children's National Health System, Washington, DC, USA

Corresponding Author: Dr. Yoshihiko Raita, Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, 125 Nashua Street, Suite 920, Boston, MA 02114-1101. Tel: 617-726-5276 E-mail: Yoshihiko_Raita@hms.harvard.edu

Acknowledgement: We thank the MARC-35 study hospitals and research personnel for their ongoing dedication to bronchiolitis and asthma research (see **Table E1** in **Additional file 1**). We also thank Ashley Sullivan, MPH, MS and Janice Espinola, MPH (Massachusetts General Hospital, Boston, MA) as well as Alkis Togias, MD (National Institute of Allergy and Infectious Diseases) for their helpful contributions.

Authors' contributions: YR contributed to the design of the study, data analysis, and interpretation of results, and writing the manuscript. CAC organized the study cohort and interpretation of results, and critically reviewed the manuscript. CGM, JMM, PAP, SCP, and SJT contributed to the acquisition and maintaining of the data, and critically reviewed the manuscript. KH contributed to the analysis of data, design of the study, and interpretation of result, and critically reviewed the manuscript. All authors finally approved the final version of manuscript.

Financial/non-financial disclosures: The authors have no financial relationships relevant to this article to disclose.

A declaration of all sources of funding: This study was supported by grants from the National Institutes of Health (Bethesda, MD): U01 AI-087881, R01 AI-114552, R01 AI-

108588, R01 AI-134940, and UG3/UH3 OD-023253. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding organizations were not involved in the collection, management, or analysis of the data; preparation or approval of the manuscript; or decision to submit the manuscript for publication.

ABSTRACT

BACKGROUND: We aimed to develop machine learning models to accurately predict bronchiolitis severity, and to compare their predictive performance with a conventional scoring (reference) model.

METHODS: In a 17-center prospective study of infants (aged <1 year) hospitalized for bronchiolitis, by using routinely-available pre-hospitalization data as predictors, we developed four machine learning models: Lasso regression, elastic net regression, random forest, and gradient boosted decision tree. We compared their predictive performance—e.g., area-under-the-curve (AUC), sensitivity, specificity, and net benefit (decision curves)—using a cross-validation method, with that of the reference model. The outcomes were positive pressure ventilation use and intensive treatment (admission to intensive care unit and/or positive pressure ventilation use).

RESULTS: Of 1,016 infants, 5.4% underwent positive pressure ventilation and 16.0% had intensive treatment. For the positive pressure ventilation outcome, machine learning models outperformed reference model (e.g., AUC 0.88 [95%CI 0.84-0.93] in gradient boosted decision tree vs 0.62 [95%CI 0.53-0.70] in reference model), with higher sensitivity (0.89 [95%CI 0.80-0.96] vs. 0.62 [95%CI 0.49-0.75]) and specificity (0.77 [95%CI 0.75-0.80] vs. 0.57 [95%CI 0.54-0.60]). The machine learning models also achieved a greater net benefit over ranges of clinical thresholds.

CONCLUSIONS: Machine learning models consistently demonstrated a superior ability to predict acute severity and achieved greater net benefit.

KEY WORDS

Machine learning, bronchiolitis, MARC-35

ABBREVIATIONS

MARC-35, the 35th Multicenter Airway Research Collaboration study; AUC, area-under-the-curve; CI, confidential interval; ICU, intensive care unit; ED, emergency department; EMNet, Emergency Medicine Network; RSV, respiratory syncytial virus; IQR, interquartile range; NRI, net reclassification improvement; PPV, positive predictive value; NPV, negative predictive value

INTRODUCTION

Bronchiolitis is the leading cause of infant hospitalization in the US, accounting for 107,000 infant hospitalizations each year with direct cost of 734 million US dollars.(2) Even among hospitalized infants, the severity of bronchiolitis can range from moderate severity (which requires observation and supportive therapies, such as supplemental oxygen, fluid, and nutrition) to near-fatal and fatal infections. Previous studies have identified individual risk factors for higher severity of bronchiolitis (e.g., young age, prematurity, viral etiology)(24–27) and developed prediction scoring models (e.g., logistic regression models).(7–10) However, identifying the subgroup of infants with bronchiolitis who require higher acuity care (e.g., positive pressure ventilation, intensive care unit [ICU] admission) remains an important challenge. The difficulty and uncertainty of predicting acute severity—and, consequently, the appropriate level of care for infants with bronchiolitis—are reflected by the well-documented variability in inpatient management across the nation.(2,6,11,12)

Machine learning models have gained increasing attention because of their advantages, such as the ability to incorporate high-order, nonlinear interactions between predictors and to yield more accurate and stable predictions. Indeed, recent studies have reported that the use of machine learning models provide a high predictive ability in various conditions and settings—e.g., sepsis,(28,29) asthma exacerbation,(14) emergency department (ED) triage,(13,30) and unplanned transfers to ICU.(15) Despite the clinical and research promise, no study has yet examined the utility of modern machine learning models in predicting outcomes in infants hospitalized for bronchiolitis—a large population with high morbidity and health resource use.

In this context, we aimed to develop machine learning models that accurately predict acute severity in infants hospitalized with bronchiolitis, and compare their predictive performance with that of conventional scoring approaches.(7)

METHODS

Study Design, Setting and Participants

The current study aimed to develop machine learning models that accurately predict acute severity in infants with bronchiolitis, by using the data from a multicenter prospective cohort study of 1,016 infants hospitalized for bronchiolitis—the 35th Multicenter Airway Research Collaboration (MARC-35) study.(31,32) MARC-35 is coordinated by the Emergency Medicine Network (EMNet, www.emnet-usa.org),(33) an international research collaboration with 246 participating hospitals. Briefly, at 17 sites across 14 U.S. states (**Table E1** in **Additional file 1**), MARC-35 enrolled infants (aged <1 year) who were hospitalized with an attending physician diagnosis of bronchiolitis during three consecutive bronchiolitis seasons (November 1 to April 30) during 2011 to 2014. The diagnosis of bronchiolitis was made according to the American Academy of Pediatrics bronchiolitis guidelines,(24) defined as acute respiratory illness with a combination of rhinitis, cough, tachypnea, wheezing, crackles, and retractions. We excluded infants who were transferred to a participating hospital >24 hours after initial hospitalization or with a preexisting heart and lung disease, immunodeficiency, immunosuppression or gestational age of <32 weeks.

We followed the Standards for Reporting Diagnostic Accuracy statement guideline for the reporting of prediction models.(34) The institutional review board of the 17 participating hospitals (Alfred I. duPont Hospital for Children, Arnold Palmer Hospital for Children, Boston Children's Hospital, Children's Hospital of Los Angeles, Children's Hospital of Philadelphia, Children's Hospital of Pittsburgh, The Children's Hospital at St. Francis, The Children's Mercy Hospital & Clinics, Children's National Medical Center, Cincinnati

Children's Hospital and Medical Center, Connecticut Children's Medical Center, Dell Children's Medical Center of Central Texas, Norton Children's Hospital, Massachusetts General Hospital, Phoenix Children's Hospital, Seattle Children's Hospital, Texas Children's Hospital) approved the study. Written informed consent was obtained from the parent or guardian.

Predictors

For predictors in the machine learning models, we selected variables based on clinical plausibility and *a priori* knowledge.(7–10,25,35–37) These predictors—which are available in most prehospitalization settings—included demographics (age, sex, and race/ethnicity), medical history (prenatal maternal smoking, gestational age, birth weight, postnatal ICU admission, history of hospital and ICU admission, history of breathing problems, and history of eczema), parent-reporting symptoms (poor feeding, cyanosis, apnea, and duration of symptoms), ED presentation (vital signs [temperature, pulse rate, respiratory rate, oxygen saturation], interaction between oxygen saturation and supplemental oxygen use, wheezing, retractions, apnea, and dehydration), and detection of respiratory syncytial virus (RSV) by PCR.(31) These clinical data were obtained through a structured interview and medical record review by trained physicians and investigators using a standardized protocol.(32) All data were reviewed at the EMNet Coordinating Center at Massachusetts General Hospital (Boston, MA), and site investigators were queried about missing data and discrepancies identified by manual data checks.

Outcomes

The primary outcome was the use of positive pressure ventilation – continuous positive airway pressure ventilation and/or intubation during inpatient stay.(38) The secondary outcome was intensive treatment defined as a composite of ICU admission and/or the use of positive pressure ventilation during the inpatient stay.(25,37) In this observational study, patients were managed at the discretion of treating physicians. These two outcomes have been employed for outcomes in the MARC-35 study.

Statistical Analysis

In the training sets (80% randomly-selected samples) in 5-fold cross-validation, we developed five models: the reference model(7) and four machine learning models for each outcome. As the reference model,(7)(22)(17) we fit logistic regression models using the predictors of a previously-established clinical prediction score that was derived using an ED sample (**Table E2**).(7) We selected this prediction score as the reference model since it was recently developed in a large sample and focused on similar clinical outcomes reflecting acute severity of bronchiolitis.(7,39) The predictors included age, poor feeding, oxygen saturation, retractions, apnea, and dehydration, excluding nasal flaring/grunting, based on the availability of data in the current study.

Next, using the prehospitalization predictors, we developed four machine learning models: 1) logistic regression with Lasso regularization (Lasso regression), 2) logistic regression with elastic net regularization (elastic net regression), 3) random forest, and 4) gradient boosted decision tree models. First, Lasso regression is an extension of regression-based models that has an ability to shrink (or regularize) the predictor coefficients toward zero, thereby effectively selecting important predictors and improving

interpretability of the model.(40) Lasso regression computes the optimal regularization parameter (lambda) that minimizes the sum of least square plus L1-shrinkage penalty using a cross-validation method.(41) Second, elastic net regression is another regression-based model incorporating both Lasso-regularization and Ridge-regularization.(40,42) Elastic net regression calculates the optimal regularization parameter that minimizes the sum of least square plus weighted L1-shrinkage penalty and weighted L2-shrinkage penalty. We used R *glmnet* and *caret* packages for Lasso regression and elastic net regression models.(43,44) Third, random forest is an ensemble of decision trees generated by bootstrapped training samples with random predictor selection in tree induction.(40,45) We created a hyperparameter tuning grid to identify the best set of parameters using cross-validation methods. We used *randomForest* and *caret* packages to construct random forest models.(44,46) Lastly, gradient boosted decision tree is another ensemble method which constructs new simple tree models predicting the errors and residuals of the previous model. When adding a new tree, this model uses a gradient descent algorithm minimizes a loss function.(47) We performed hyperparameter tuning sequentially using a 5-fold cross-validation method. We used R *xgboost* and *caret* packages to construct gradient boosted decision tree.(44,48) To minimize potential overfitting, we utilized several methods – e.g., regularizations (or penalizations) in Lasso and elastic net regression models, out-of-bag estimation in random forest models, and cross-validation in all models.

As for the predictor engineering methods of the machine learning models, we preprocessed predictors sequentially. First, we investigated non-linear relationships between the continuous predictors and outcomes and created quadric terms of age,

respiratory rate, and temperature. These quadratic terms were used only for regression-based machine learning models (i.e., logistic regression models with Lasso regularization and those with elastic net regularization). Second, we also chose either of highly-correlated predictors (e.g., age and weight at hospitalization). Third, we imputed predictors with missing values (**Table E3**) using bagged tree imputation. Fourth, we converted continuous predictors into normalized scales using Yeo-Johnson transformation. Categorical predictors were coded as dummy variables while birth weight, gestational age, previous breathing problem, and degree of retraction were coded as ordinal variables. Fifth, to incorporate clinically evident interaction between oxygen saturation level and use of supplemental oxygen, we created an interaction term between oxygen saturation and use of supplemental oxygen. Lastly, we removed predictors that are highly sparse in the dataset. We applied these preprocessing methods independently to the training sets and the test sets to avoid carrying the information from the training sets to the test sets. We used R *recipe* package for these predictor preprocessing.(49)

To examine the variable importance in the random forest, we used permutation-based variable importance – normalized average values of difference between the prediction accuracy of out-of-bag estimation and that of the same measure after permutating each predictor. In the gradient boosted model, we also computed the variable importance that is summed over iterations.(45) We graphically presented the rank of variable importance using unscaled values.

To measure the test performance of each model, we computed the overall cross-validation performance from the test sets (the remaining randomly-selected 20% samples). As the predictive performance, we used 1) the area under the receiver-operating-characteristic

curve (AUC), 2) net reclassification improvement, 3) confusion matrix results (i.e., sensitivity, specificity, positive predictive value, and negative predictive value), and 4) net benefit from decision curve analysis. To compare the AUC between the models, we used Delong's test.(50) To compute AUC and its confidential interval, we used *pROC* package.(51) We also used the net reclassification improvement to quantify whether a new model provides clinically relevant improvements in prediction when compared to the reference model.(52) To compute the net reclassification improvement, we used *PredictABEL* package.(53) To address the class imbalance in the both outcomes, we employed the value with the shortest distance to the top-left part of the AUC plot as the threshold for the confusion matrix.(45) The decision curve analysis incorporates the information on both the benefit of correctly predicting the outcome (true-positives) and the relative harm of incorrectly labelling patients as if they would have the outcome (false-positives) – i.e., the net benefit.(54–58) We made a graphical presentation of the net benefit for each model over a range of threshold probabilities (or clinical preferences) of the outcome as decision curves. We used decision curve analysis R source code from Memorial Sloan Kettering Cancer Center(59) and plotted the graphs using *ggplot2* package.(60) We performed all analysis with R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria).(61)

RESULTS

During 2011-2014, 1,016 infants with bronchiolitis were enrolled into a 17-center prospective cohort study. The median age at the enrolment was 3.2 months (IQR 1.6-6.0), 40% were female, and 42% were non-Hispanic white. The length-of-hospital stay varied widely from 0 day to 60 days (median, 2 days; IQR 1-3 days) (**Table 1**). Clinical data had a small proportion of missingness; most had <1% missingness (e.g., missingness on oxygen saturation with the use of supplemental oxygen, 0.1%) while the maximum proportion of missing was 4.8% (**Table E3** in **Additional file 1**). Overall, 55 infants (5.4%) underwent positive pressure ventilation and 163 infants (16.0%) had intensive treatment outcome.

Predicting Positive Pressure Ventilation Outcome

In the prediction of positive pressure ventilation outcome, the discriminatory abilities of all models are summarized in **Figure 1A** and **Table 2**. All four machine learning models demonstrated significantly superior AUCs (all P<0.001). For example, compared with the reference model (AUC 0.62 [95%CI 0.53-0.70]), the AUC was higher in the elastic net regression (AUC 0.89 [95%CI 0.85-0.92]) and gradient boosted decision tree (AUC 0.88 [95%CI 0.84-0.93]) models. Similarly, compared with the reference model, all machine learning models also achieved a significant net reclassification improvement (all P<0.001).

Additionally, compared with the reference model, all machine learning models also demonstrated a higher sensitivity (e.g., 0.62 [95%CI 0.49-0.75] in the reference model vs. 0.89 [95%CI 0.80-0.96] in the elastic net regression; **Table 2**) and specificity (e.g., 0.57 [95%CI 0.54-0.60] in the reference model vs. 0.79 [95%CI 0.77-0.82] in the Lasso regression model). More specifically, all machine learning models correctly predicted a

larger number of infants who underwent positive pressure ventilation (true-positives) with a fewer number of predicted outcomes (**Table 3**). For example, the reference scoring system categorized most infants (n=629, 62%) into the prediction score groups of 2-3. The reference model correctly identified 16 out of 25 infants who underwent positive pressure ventilation, while predicting that 265 infants would have undergone positive pressure ventilation. In contrast, the gradient boosted decision tree model correctly identified 23 (of 25) patients, while predicting that 135 infants would have undergone positive pressure ventilation in the same patient groups. Considering the low prevalence of the positive pressure ventilation outcome, all models had a high negative predictive value (e.g., 0.96 [95%CI 0.95-0.97] in the reference model vs. 0.99 [95%CI 0.99-0.99] in the Lasso regression model; **Table 2**).

Likewise, in the decision curve analysis (**Figure 1B**), all four machine learning models outperformed the reference model, demonstrating a greater net benefit throughout the range of clinical thresholds, indicating that the machine learning prediction would more accurately identify high-risk infants (true-positives) while taking the trade-off with false-positives into consideration.

Predicting Intensive Treatment Outcome

In the prediction of intensive treatment outcome, the discriminatory abilities of all models are shown in **Figure 2A** and **Table 2**. All four machine learning models demonstrated a significantly higher AUC (all P<0.001). For example, compared with the reference model (AUC 0.62 [95%CI 0.57-0.67]), the AUC was higher in the elastic net regression (AUC 0.80 [95%CI 0.76-0.83]) and random forest (AUC 0.79 [95%CI 0.75-0.84])

models. Similarly, compared with the reference model, all machine learning models also achieved significant net reclassification improvement (all P<0.001).

Additionally, all machine learning models demonstrated a higher sensitivity (e.g., 0.58 [95%CI 0.49-0.75] in the reference model vs. 0.75 [95%CI 0.69-0.82] in the Lasso regression; **Table 2**) and specificity (e.g., 0.58 [95%CI 0.50-0.66] in the reference model vs. 0.78 [95%CI 0.76-0.81] in the random forest model). For example, among the infants categorized into the reference score groups of 2-3 (62% of cohort infants), the reference model correctly identified 39 out of 80 infants who had intensive treatment, while predicting that 275 infants would have had intensive treatment (**Table 3**). In contrast, the gradient boosted decision tree correctly identified 52 (out of 80) infants with the outcome, while predicting that 162 infants would have had intensive treatment. Likewise, in the decision curve analysis (**Figure 2B**), all four machine learning models outperformed the reference model, demonstrating a greater net benefit throughout the range of clinical thresholds.

Variable Importance

To yield insights into the relevance of each predictor, **Figures 1E** and **2** (**Additional file 1)** summarized the 15 most important predictors of random forest and gradient boosted decision tree models for each outcome. In the prediction of positive pressure ventilation outcome, age, oxygen saturation level with the use of supplemental oxygen, and other vital signs [at the presentation] were the most important predictors in both models (**Figures 1EA** and **2EA**). Likewise, in the prediction of intensive treatment outcome, similar predictors were considered important in the both models (**Figures 1EB** and **2EB**).

DISCUSSION

In this analysis of multicenter prospective cohort data from 1,016 infants, we applied four modern machine learning approaches (i.e., Lasso regression, elastic net regression, random forest, and gradient boosted decision tree) to the prediction of acute severity outcomes of bronchiolitis. Compared to the reference model that was derived in an ED sample,(7) these machine learning models consistently demonstrated a superior performance in predicting positive pressure ventilation and intensive treatment outcomes, including AUC and net reclassification. Additionally, the machine learning models achieved a higher sensitivity and specificity for the two outcomes, in both the overall cohort and the majority of cohort infants that were categorized into the reference score groups of 2-3. Furthermore, the decision curve analysis also demonstrated the net benefit of machine learning models was also greater—i.e., a larger number of true-positives considering a trade-off with false-positives—across a range of clinical thresholds. To the best of our knowledge, this is the first study that has investigated the performance of modern machine learning models in predicting severity in infants with bronchiolitis.

One of the main objectives in the risk stratification of infants with bronchiolitis is to promptly identify infants at risk for higher severity and efficiently utilize finite healthcare resources. The American Academy of Pediatrics bronchiolitis guideline(24) highlights the importance of assessing the risk in infants with bronchiolitis. However, optimal risk stratification and prediction remains a challenge as the clinical course in this population (even in infants hospitalized for bronchiolitis) is highly variable.(6,11,12) Previous studies, by using conventional modeling (e.g., logistic regression models), have reported a moderate ability to predict severity outcomes (e.g., ED-to-hospital admission, hospital

length-of-stay, ICU admission, positive pressure ventilation use) of infants with bronchiolitis.(7–10,62) Although the use of an expanded set of predictors—e.g., repeated examinations and invasive monitoring during hospital course—may yield better predictive performance, it is often impractical in the real-world acute care settings with an aim to promptly risk-stratify these infants. Alternatively, the use of advanced machine learning models may improve the clinician's decision-making ability. Indeed, machine learning models have recently been applied to the prediction of various disease conditions and clinical settings, such as early identification of mortality risk in patients with sepsis,(28) rehospitalization in patients with heart failure,(63) intensive treatment outcomes in patients with asthma exacerbation,(14) unplanned transfer to ICU,(15) and escalated care at pediatric ED triage.(30) Our multicenter study builds on these earlier reports, and extends them by demonstrating that the modern machine learning models outperform conventional approaches in predicting higher severity of infants with bronchiolitis. While external validation is warranted, these machine learning models using routinely-available predictors can be implemented to clinical practice (e.g., online risk calculators or build-in risk assessment systems)—similar to existent clinical scoring rules.

Clinical prediction systems strive for an appropriate balance between sensitivity and specificity because of the trade-off relationship between these two factors in the context of prevalence of clinical outcomes. In the present study, we observed that the reference score model did not effectively categorize most infants (i.e., 62% of cohort were categorized into the two score groups) or appropriately predicted infants who developed the outcomes. By contrast, the machine learning models correctly identified a larger number of true-positives (i.e., higher sensitivity). This finding supports the utility of these models in the

target population, for which the one of the major priorities is to reduce "missed" high-risk cases. Additionally, the machine learning models also had a fewer number of false-positives (i.e., higher specificity) in predicting both outcomes while they were imperfect in the setting of relatively-smaller prevalence of outcome (5.4% for positive pressure ventilation use). This may mitigate excessive resource use in this large population. These findings are further supported by the decision curve analysis that demonstrated a greater net benefit of the machine learning models incorporating the trade-offs between true-positives and false-positives across the wide ranges of clinical thresholds.

There are several potential explanations for the observed gains in the predictive abilities of machine learning models. For example, machine learning models incorporate high-order interactions between predictors and nonlinear relationships with outcomes. Additionally, machine learning models are able to mitigate potential overfitting by adopting several methods, such as regularization, out-of-bagging estimation, and cross-validation. Furthermore, the use of large multicenter data with rigorous quality assurance might have contributed to low bias and variance in the machine models. Although the machine learning models achieved superior predictive ability, their performance remained imperfect. This may be explained, at least partially, by the limited set of predictors, subjectivity of some data elements (e.g., parent-reported symptoms at home), variable clinical factors after prehospitalization assessment (e.g., ED management and patient responses), difference in clinician's practice patterns, and availability of intensive care resources. Notwithstanding the complexity and challenges of clinical prediction in infants with bronchiolitis, machine learning models have scalable advantages in the era of health information technology, such as automated sophistication of models through the sequential extraction of electronic

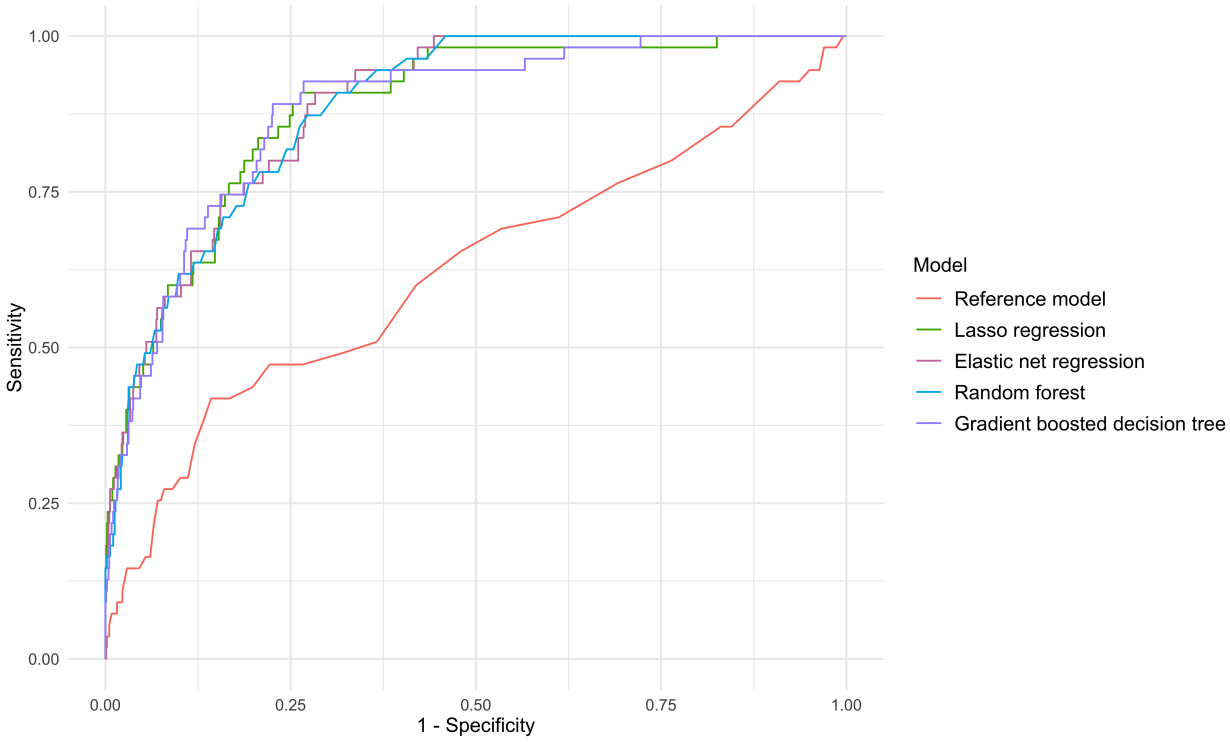
health records, continuous non-invasive physiological monitoring, natural language processing, and reinforcement learning.(64–67) In the past, this scalability had not been attainable with the use of conventional approaches. Taken together, our findings and recent developments support cautious optimism that modern machine learning may enhance the clinician's ability as an assistive technology.

Our study has several potential limitations. First, the data may be subject to measurement bias and missingness. However, the study was conducted by trained investigators using a standardized protocol, which led to the low proportion of missingness in the predictors (**Table E3** in **Additional file 1**). Second, the clinical thresholds for these outcomes may depend on local resources and vary between clinicians and hospitals (e.g., different criteria for admission to the ICU). Yet, the decision curve analysis demonstrated the greater benefit of the machine learning models across the wide range of clinical thresholds. Third, we used the reference model with a limited number of predictors (i.e., six predictors). However, the reference model was recently developed in a large sample and focused on similar clinical outcomes reflecting acute severity of bronchiolitis. Besides, considering the number of the positive ventilation outcomes (n=55), the number of predictors used for the conventional logistic regression model (reference model) is statistically justified. Lastly, the study cohort consisted of a racially/ethnically- and geographically-diverse US sample of infants hospitalized with bronchiolitis. While the severity of this population was highly variable and the model used pre-hospitalization data, our models might not be generalizable to infants in ambulatory settings. External validation of the models in different populations and settings is necessary. Nonetheless, our data remain highly relevant for the 107,000 infants hospitalized yearly in the US.(2)

FIGURES

Figure 1. Prediction ability of the reference and machine learning models for positive pressure ventilation outcome in the overall cross-validation dataset

A) Receiver-operating-characteristics (ROC) curves. The corresponding value of the area under the receiver-operating-characteristics curve (AUC) for each model are presented in **Table 2**.



B) Decision curve analysis. X-axis indicates the threshold probability for positive pressure ventilation outcome; Y-axis indicates the net benefit. Compared to the reference model, the net benefit of all machine learning models was larger over the range of clinical threshold.

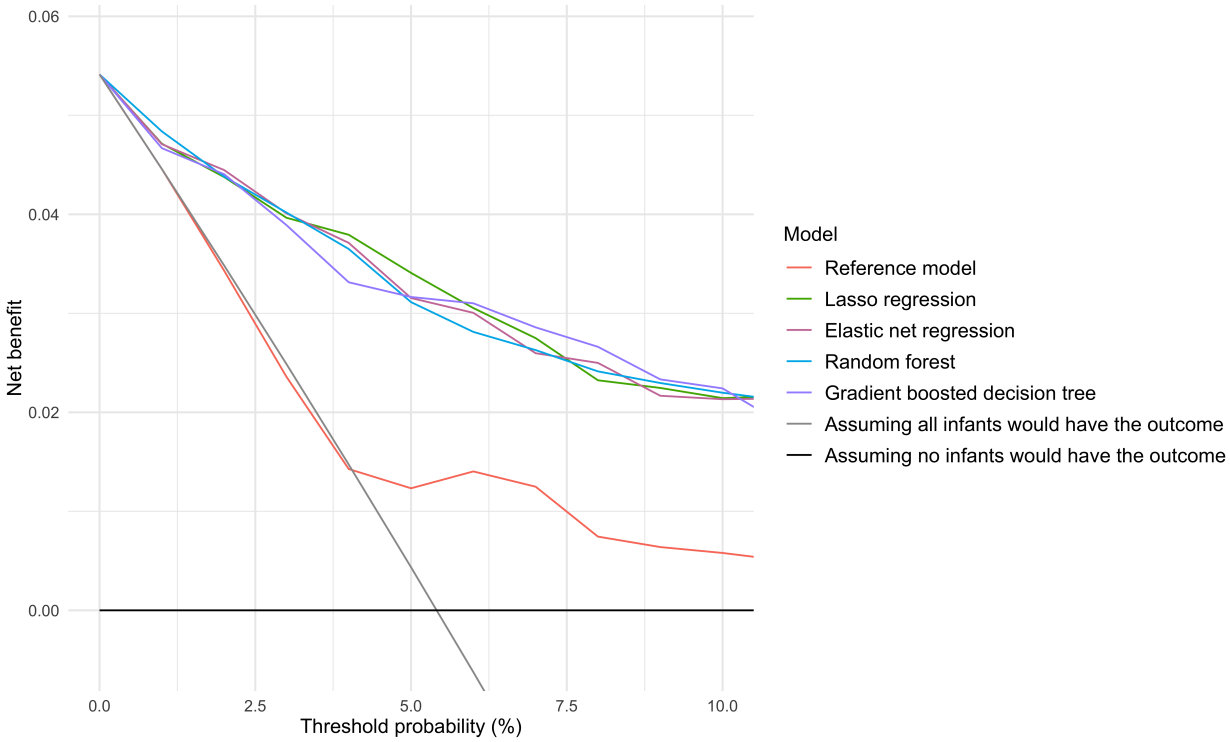
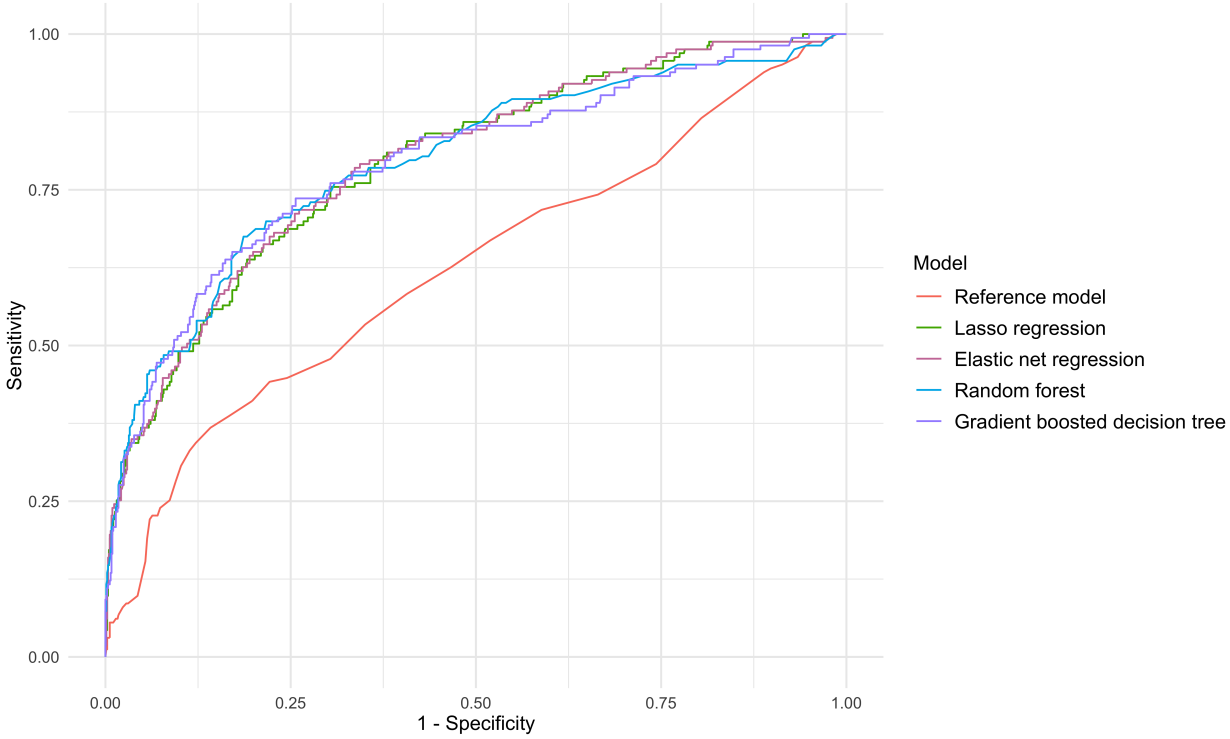
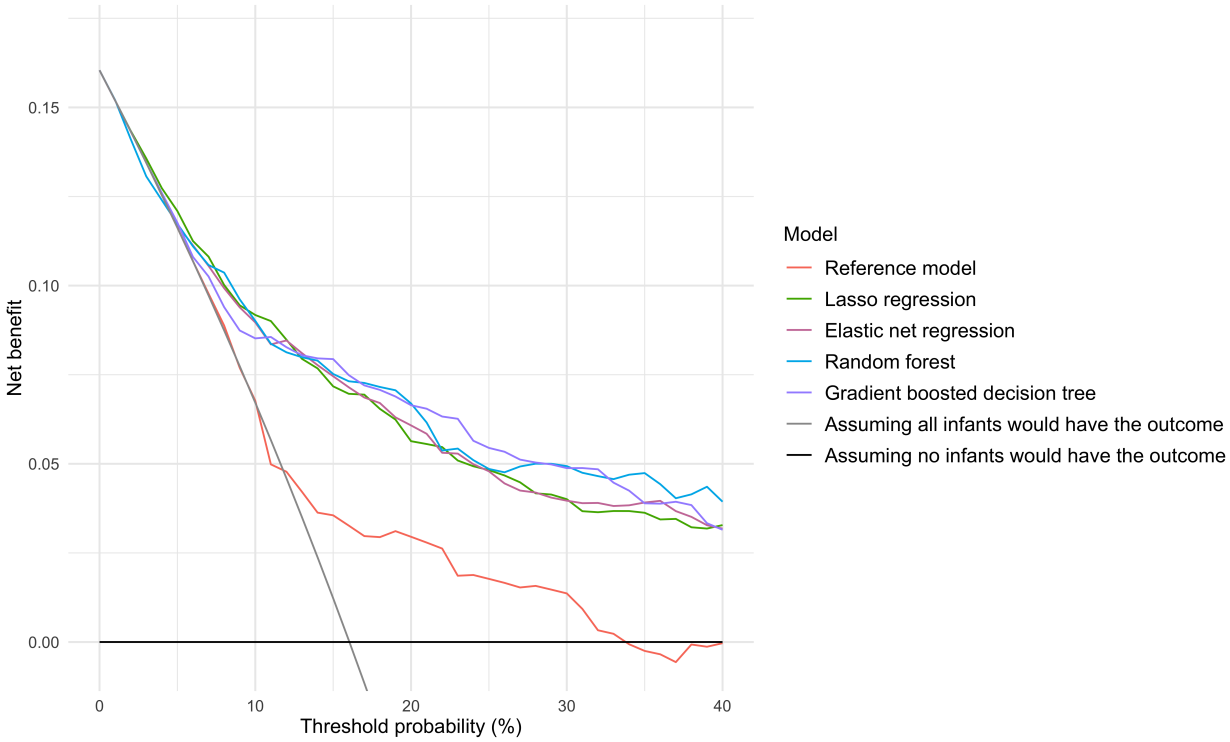

0.06
0.04
0.02
0.00
Net benefit
0.0
2.5
5.0
7.5
10.0
Threshold probability (%)
Model
Reference model
Lasso regression
Elastic net regression
Random forest
Gradient boosted decision tree
Assuming all infants would have the outcome
Assuming no infants would have the outcome

Figure 2. Prediction ability of the reference and machine learning models for intensive treatment outcome in the overall cross-validated dataset.

A) Receiver-operating-characteristics (ROC) curves. The corresponding values of the area under the receiver-operating-characteristics curve (AUC) for each model are presented in **Table 2**.



B) Decision curve analysis. X-axis indicates the threshold probability for intensive treatment outcome; Y-axis indicates the net benefit. Compared to the reference model, the net benefit of all machine learning models was larger over the range of clinical threshold.


0.15
0.10
0.05
0.00
Net benefit
0
10
20
30
40
Threshold probability (%)
Model
Reference model
Lasso regression
Elastic net regression
Random forest
Gradient boosted decision tree
Assuming all infants would have the outcome
Assuming no infants would have the outcome

TABLES

Table 1. Patient characteristics and clinical outcomes in 1,016 infants hospitalized for bronchiolitis

Variables	n=1,016	
Demographics:		
Age (month), median (IQR)	3.2	(1.6-6.0)
Female sex	406	(40.0)
Race/ethnicity		
Non-Hispanic white	430	(42.0)
Non-Hispanic black	239	(23.5)
Others	347	(34.2)
Medical history:		
Prenatal maternal smoking	147	(14.7)
Gestational age (week)		
32-33	35	(3.4)
34-36	151	(14.9)
37-39	417	(41.0)
40-41	391	(38.5)
≥42	22	(2.2)
Birth weight (kg)		
0-1.3	3	(0.3)
1.4-2.2	61	(6.0)
2.3-3.1	343	(33.9)
≥3.2	604	(59.7)
Postnatal ICU admission	167	(16.4)
Previous hospital admission	162	(16.0)
Previous ICU admission	17	(1.7)
Previous breathing problems (count)	32	(3.2)
0	810	(79.7)
1	160	(15.7)
2	46	(4.5)
History of eczema	149	(14.7)
Parent-reported symptoms at home		
Poor feeding	32	(3.2)
Cyanosis within 24 hours	92	(9.1)
Apnea	131	(12.9)
Apnea within 24 hours	86	(8.5)
Duration of symptom (≤24 hours)	53	(5.2)
Signs and symptom at ED		
Vital signs at presentation		
Temperature (F), median (IQR)	99.4	(98.8-100)
Pulse rate (bpm), median (IQR)	162	(150-176)
Respiratory rate (per min), median (IQR)	48	(40-60)
Use of supplemental oxygen (%)	51	(5)
Oxygen saturation level (%) at room air (IQR)	96	(94-98)
Oxygen saturation level (%) with the use of supplemental oxygen	98	(95-100)
Wheeze	602	(62.3)
Severity of retraction		
None	192	(19.6)
Mild	431	(43.9)
Moderate/severe	358	(36.5)
Apnea	56	(5.5)
Dehydration	392	(39.5)
Virology		
RSV	821	(80.8)
Length of hospital stay (days), range	0-60	
Clinical outcomes		
Positive pressure ventilation use[a]	55	(5.4)
Intensive treatment use[b]	163	(16.0)

Abbreviations: bpm, beats per minute; IQR, interquartile range; ICU, intensive care unit; RSV,
Data are no. (%) of infants unless otherwise indicated. Percentages may not equal 100, because of

[a] Infants with bronchiolitis who underwent continuous positive airway ventilation and/or
[b] Infants with bronchiolitis who were admitted to ICU and/or who underwent positive pressure

Table 2. Prediction performance of the reference, and machine learning models in infants hospitalized for bronchiolitis

Outcomes and models	AUC	P-value[a]	NRI[b]	P-value[b]	Sensitivity	Specificity	PPV	NPV
Positive pressure ventilation outcome								
Reference model	0.62 (0.53-0.70)	Reference	Reference	Reference	0.62 (0.49-0.75)	0.57 (0.54-0.60)	0.075 (0.054-0.097)	0.96 (0.95-0.97)
Logistic regression with Lasso regularization	0.88 (0.84-0.93)	<0.001	1.09 (0.87 - 1.32)	<0.001	0.84 (0.73-0.93)	0.79 (0.77-0.82)	0.19 (0.14-0.24)	0.99 (0.99-0.99)
Logistic regression with elastic net regularization	0.89 (0.85-0.92)	<0.001	1.05 (0.82 - 1.28)	<0.001	0.89 (0.80-0.96)	0.73 (0.70-0.75)	0.15 (0.11-0.18)	0.99 (0.99-0.99)
Random forest	0.89 (0.85-0.92)	<0.001	1.17 (0.96 - 1.38)	<0.001	0.85 (0.75-0.95)	0.74 (0.71-0.76)	0.15 (0.12-0.21)	0.99 (0.99-0.99)
Gradient boosted decision tree	0.88 (0.84-0.93)	<0.001	1.08 (0.84 - 1.33)	<0.001	0.89 (0.80-0.96)	0.77 (0.75-0.80)	0.17 (0.08-0.21)	0.99 (0.99-0.99)
Intensive treatment outcome								
Reference model	0.62 (0.57-0.67)	Reference	Reference	Reference	0.58 (0.55-0.62)	0.58 (0.50-0.66)	0.21 (0.18-0.24)	0.88 (0.86-0.89)
Logistic regression with Lasso regularization	0.79 (0.76-0.83)	<0.001	0.68 (0.52 - 0.84)	<0.001	0.75 (0.69-0.82)	0.70 (0.66-0.73)	0.31 (0.26-0.38)	0.94 (0.93-0.94)
Logistic regression with elastic net regularization	0.80 (0.76-0.83)	<0.001	0.58 (0.42- 0.74)	<0.001	0.72 (0.64-0.79)	0.74 (0.71-0.77)	0.33 (0.28-0.41)	0.93 (0.92-0.94)
Random forest	0.79 (0.75-0.84)	<0.001	0.70 (0.55 - 0.86)	<0.001	0.70 (0.63-0.77)	0.78 (0.76-0.81)	0.37 (0.29-0.45)	0.93 (0.92-0.94)
Gradient boosted decision tree	0.79 (0.75-0.84)	<0.001	0.72 (0.57 - 0.87)	<0.001	0.74 (0.67-0.80)	0.74 (0.71-0.77)	0.33 (0.26-0.42)	0.93 (0.92-0.94)

Abbreviations: AUC, area under the receiver-operating-characteristic curve; NRI, net reclassification improvement; PPV, positive predictive value; NPV, negative predictive value;

[a] P-value was calculated to compare area-under-the-curve of the reference model with that of each machine model

[b] We used continuous NRI and its P-value

Table 3. The number of actual and predicted outcomes of prediction models, according to the score of the reference model

Reference model (score)	Positive pressure ventilation use		Reference model		Lasso regression		Elastic net regression		Random forest		Gradient boosted tree	
			Correctly identified outcome	Predicted outcome	Correctly identified outcome	Predicted outcome	Correctly identified outcome	Predicted outcome	Correctly identified outcome	Predicted outcome	Correctly identified outcome	Predicted outcome
	n (%)		n (%)	n	n (%)	n	n (%)	n	n (%)	n	n (%)	n
0: (n=41)	1	(2.4)	0	6	0	7	0	8	1	9	0	8
1: (n=64)	3	(4.7)	1	15	3	11	2	14	2	11	2	11
2: (n=359)	13	(3.6)	9	156	12	80	11	106	11	78	12	79
3: (n=270)	12	(4.4)	7	109	10	52	12	65	10	58	11	56
4: (n=122)	3	(0.8)	2	46	1	20	1	30	2	41	1	24
5: (n=58)	8	(13.8)	3	22	7	21	8	24	7	31	8	24
6: (n=15)	0	(0.0)	0	3	0	6	0	6	0	8	0	6
7: (n=41)	5	(12.5)	3	22	3	22	5	30	4	28	5	29
8: (n=24)	4	(16.7)	2	8	4	12	4	14	4	17	4	15
9: (n=11)	0	(0.0)	0	2	0	4	0	4	0	7	0	5
10: (n=8)	5	(62.5)	3	3	5	7	5	7	5	8	5	7
11: (n=3)	1	(33.3)	0	2	1	2	1	3	1	3	1	2
12: (n=0)	0	(0)										
Overall (n=1,016)	55	5.4	30 (55)	394	46 (84)	244	49 (89)	311	47 (85)	299	49 (89)	266

Reference model (Score)	Intensive treatment		Reference model		Lasso regression		Elastic net regression		Random forest		Gradient boosted tree	
			Correctly identified outcome	Predicted outcome	Correctly identified outcome	Predicted outcome	Correctly identified outcome	Predicted outcome	Correctly identified outcome	Predicted outcome	Correctly identified outcome	Predicted outcome
	n (%)		n (%)	n	n (%)	n	n (%)	n	n (%)	n	n (%)	n
0: (n=41)	2	(4.9)	1	10	1	8	0	5	0	4	0	4
1: (n=64)	8	(12.5)	1	11	5	17	5	11	5	12	5	9
2: (n=359)	44	(12.2)	21	157	27	133	23	111	26	77	24	87
3: (n=270)	36	(13.3)	18	118	28	75	28	71	23	64	28	75
4: (n=122)	17	(13.8)	8	53	9	36	9	33	7	40	10	53
5: (n=58)	17	(29.3)	6	26	14	30	14	30	15	28	15	29
6: (n=15)	3	(20.0)	0	3	3	7	2	7	3	9	3	12
7: (n=41)	19	(47.5)	15	28	19	38	19	36	18	32	18	35
8: (n=24)	7	(29.2)	4	8	7	18	7	17	7	16	7	17
9: (n=11)	1	(9.1)	0	3	1	10	1	9	1	6	1	7
10: (n=8)	7	(87.5)	3	4	7	7	7	7	7	8	7	8
11: (n=3)	2	(66.7)	0	1	2	3	2	3	2	3	2	3
12: (n=0)	0	(0)										
Overall (n=1,016)	163	16.0	77 (47)	422	123 (75)	382	117 (72)	340	114 (70)	299	120 (74)	339

SUPPLEMENTARY MATERIAL

Machine learning-based prediction of acute severity in infants hospitalized for bronchiolitis: A multicenter prospective study

Authors: Yoshihiko Raita, MD, MPH[1]; Carlos A. Camargo, Jr. MD, DrPH[1]; Charles G. Macias, MD, MPH[2]; Jonathan M. Mansbach, MD, MPH[3]; Pedro A. Piedra, MD[4]; Stephen C. Porter, MD, MPH, MSc[5,6]; Stephen J. Teach, MD, MPH[7]; and Kohei Hasegawa, MD, MPH[1]

Table E1. Principal investigators at the 17 participating sites in MARC-35

Amy D. Thompson, MD	Alfred I. duPont Hospital for Children, Wilmington, DE
Federico R. Laham, MD, MS	Arnold Palmer Hospital for Children, Orlando, FL
Jonathan M. Mansbach, MD, MPH	Boston Children's Hospital, Boston, MA
Vincent J. Wang, MD, MHA and Susan Wu, MD	Children's Hospital of Los Angeles, Los Angeles, CA
Michelle B. Dunn, MD and Jonathan M. Spergel, MD, PhD	Children's Hospital of Philadelphia, Philadelphia, PA
Juan C. Celedón, MD, DrPH	Children's Hospital of Pittsburgh, Pittsburgh, PA
Michael R. Gomez, MD, MS-HCA and Nancy Inhofe, MD	The Children's Hospital at St. Francis, Tulsa, OK
Brian M. Pate, MD and Henry T. Puls, MD	The Children's Mercy Hospital & Clinics, Kansas City, MO
Stephen J. Teach, MD, MPH	Children's National Medical Center, Washington, D.C.
Richard T. Strait, MD and Stephen C. Porter, MD, MSc, MPH	Cincinnati Children's Hospital and Medical Center, Cincinnati, OH
Ilana Y. Waynik, MD	Connecticut Children's Medical Center, Hartford, CT
Sujit Iyer, MD	Dell Children's Medical Center of Central Texas, Austin, TX
Michelle D. Stevenson, MD, MS	Norton Children's Hospital, Louisville, KY
Wayne G. Shreffler, MD, PhD and Ari R. Cohen, MD	Massachusetts General Hospital, Boston, MA
Anne K. Beasley, MD and Cindy S. Bauer, MD	Phoenix Children's Hospital, Phoenix, AZ
Thida Ong, MD and Markus Boos, MD, PhD	Seattle Children's Hospital, Seattle, WA
Charles G. Macias, MD, MPH	Texas Children's Hospital, Houston, TX

Table E2. Predictors and scores of the reference models in 1,016 infants hospitalized for bronchiolitis

	Scores and predictors	Overall n=1,016	
Reference model[a]	Overall score, median(IQR)	3	(2-4)
	Age ≤2 month	311	(30.6)
	Poor feeding (reported on history)	32	(3.2)
	Oxygen saturation <90%	85	(9.0)
	Apnea (history or observed in the ED)	153	(15.1)
	Dehydration (observed in ED)	392	(39.5)
	Retraction	789	(80.4)

Abbreviations: ED, emergency department; IQR, interquartile range

Data are no. (%) of infants unless otherwise indicated. Percentages may not equal 100, because of rounding and missingness.

[a] Original model developed by *Freire et al.*[1] includes nasal flaring and/or grunting variable that is not available in the MARC-35 data.

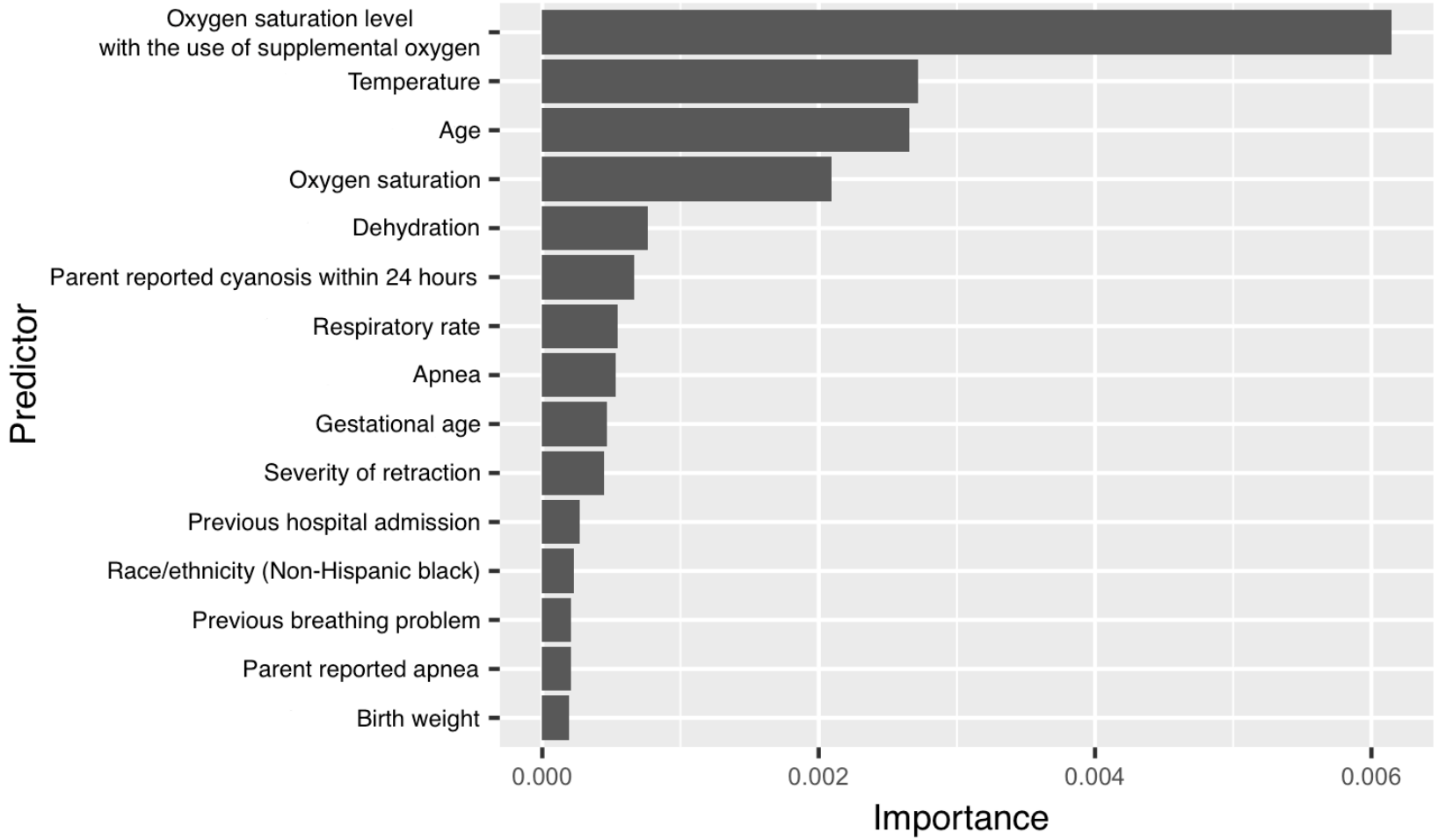
Table E3. Proportion of missing data in 1,016 infants hospitalized for bronchiolitis

Variables	Missing n (%)
Demographics:	
Age	0 (0)
Female sex	0 (0)
Race/ethnicity	0 (0)
Medical history:	
Prenatal maternal smoking	18 (1.8)
Gestational age	0 (0)
Birth weight	5 (0.5)
Postnatal ICU admission	0 (0)
Previous hospital admission	2 (0.2)
Previous ICU admission	0 (0)
Previous breathing problems	0 (0)
History of eczema	1 (0.1)
Parent-reporting symptoms at home	
Poor feeding	1 (0.1)
Cyanosis within 24 hours	0 (0)
Apnea	0 (0)
Apnea within 24 hours	0 (0)
Duration of symptom	0 (0)
Signs and symptom at ED	
Vital signs at presentation	
Temperature	17 (1.7)
Pulse rate	19 (1.9)
Respiratory rate	19 (1.9)
Use of supplemental oxygen	16 (1.6)
Oxygen saturation	21 (2.1)
Oxygen saturation with use of supplemental oxygen	1 (0.1)
Wheeze	49 (4.8)
Severity of retraction	35 (3.4)
Apnea	0 (0)
Dehydration	23 (2.3)
Virology	0 (0)
Clinical outcomes	
Positive pressure ventilation use[a]	0 (0)
Intensive treatment use[b]	0 (0)

Abbreviations: ED, emergency department; ICU, intensive care unit; RSV, respiratory

[a] Infants with bronchiolitis who underwent continuous positive airway ventilation and/or mechanical ventilation.
[b] Infants with bronchiolitis who were admitted to ICU and/or who underwent positive pressure ventilation.

Figure E1. Variable importance plot of random forest model
A) Positive pressure ventilation outcome



B) Intensive treatment outcome

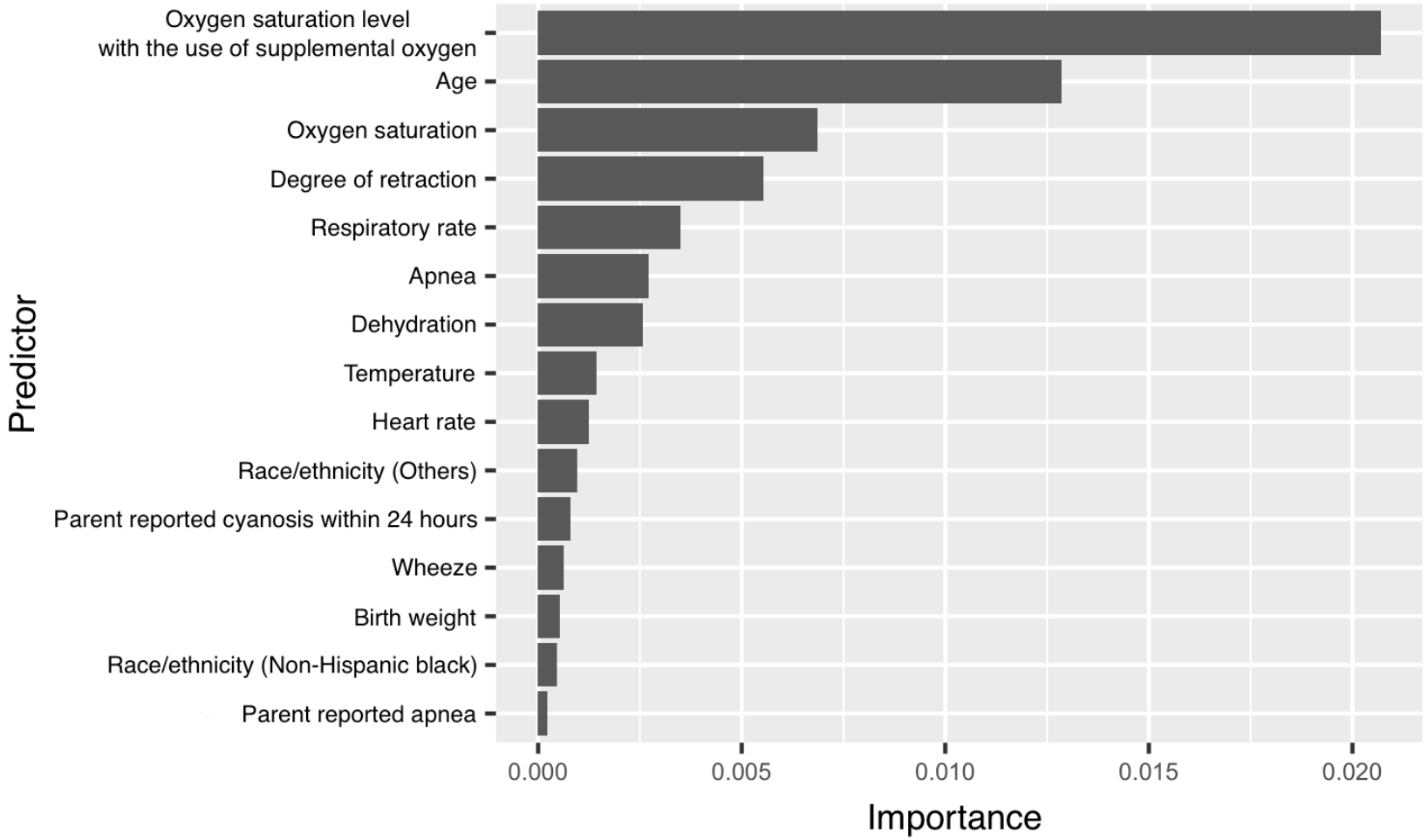
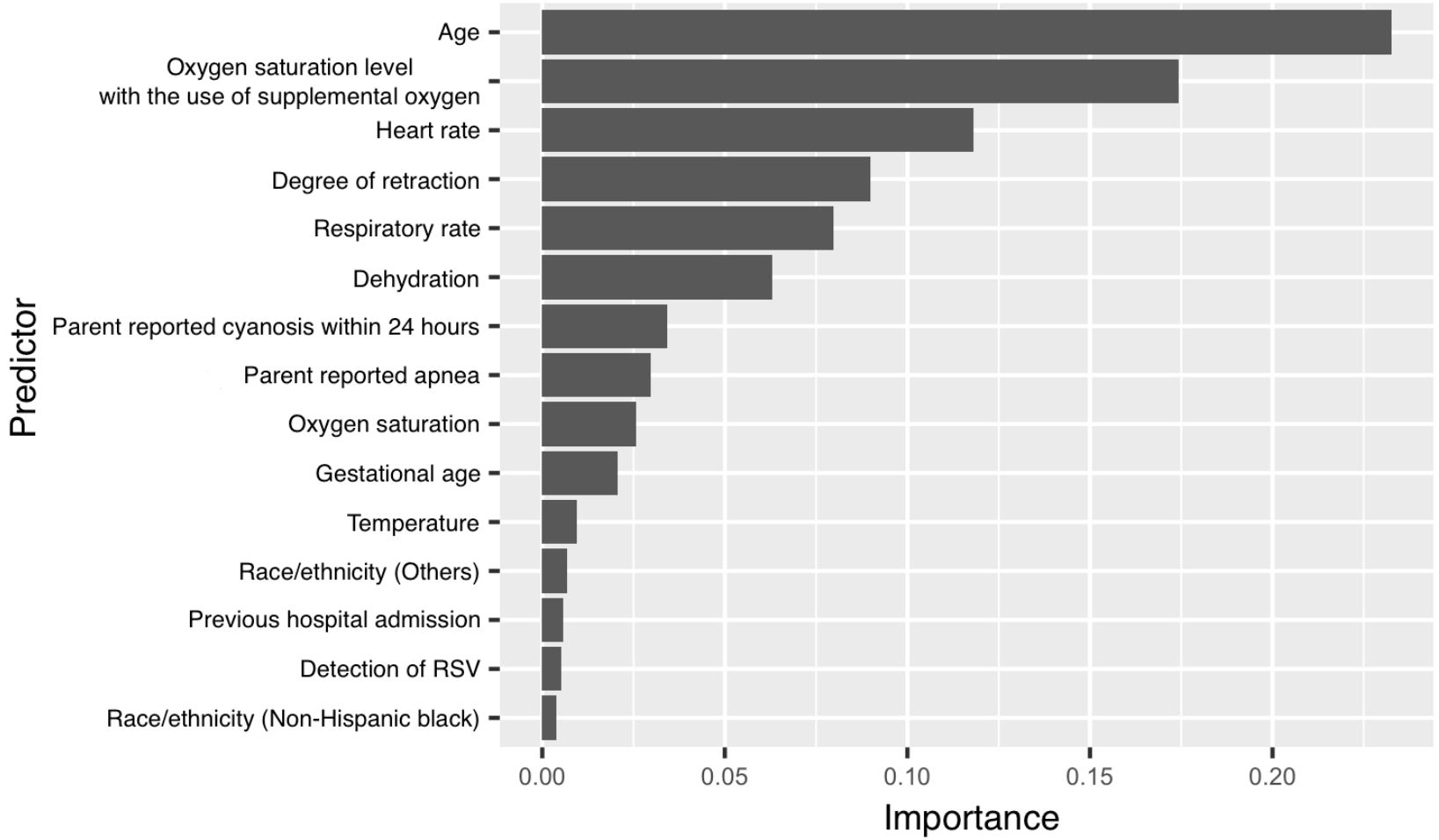
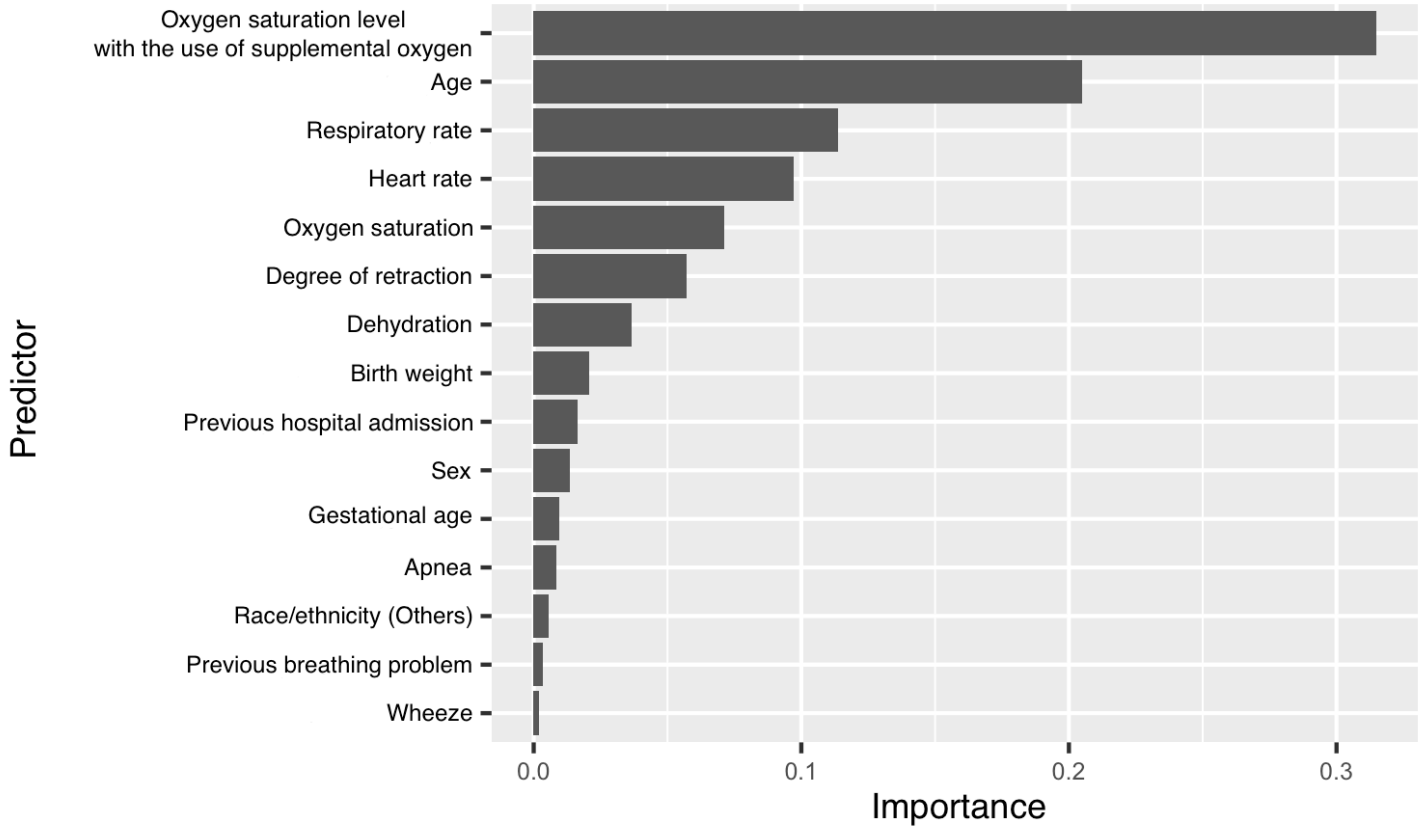


Figure E2. Variable importance plot of gradient boosted decision tree model
A) Positive pressure ventilation outcome



B) Intensive treatment outcome



REFERENCES

1. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: A classification of data science tasks. CHANCE. 2019 Jan 2;32(1):42–9.
2. Fujiogi M, Goto T, Yasunaga H, Fujishiro J, Mansbach JM, Camargo CA, et al. Trends in bronchiolitis hospitalizations in the United States: 2000-2016. Pediatrics. 2019;144(6).
3. Abreo A, Gebretsadik T, Stone CA, Hartert TV. The impact of modifiable risk factor reduction on childhood asthma development. Clin Transl Med. 2018 Jun 11;7(1):15.
4. Most Recent National Asthma Data | CDC [Internet]. 2019 [cited 2019 Apr 20]. Available from: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm
5. HCUPnet [Internet]. HCUPnet. [cited 2018 Nov 27]. Available from: https://hcupnet.ahrq.gov
6. Christakis DA, Cowan CA, Garrison MM, Molteni R, Marcuse E, Zerr DM. Variation in inpatient diagnostic testing and management of bronchiolitis. Pediatrics. 2005 Apr;115(4):878–84.
7. Freire G, Kuppermann N, Zemek R, Plint AC, Babl FE, Dalziel SR, et al. Predicting escalated care in infants with bronchiolitis. Pediatrics. 2018 Sep;142(3).
8. Walsh P, Rothenberg SJ, O'Doherty S, Hoey H, Healy R. A validated clinical model to predict the need for admission and length of stay in children with acute bronchiolitis. Eur J Emerg Med Off J Eur Soc Emerg Med. 2004 Oct;11(5):265–72.
9. Voets S, van Berlaer G, Hachimi-Idrissi S. Clinical predictors of the severity of bronchiolitis. Eur J Emerg Med Off J Eur Soc Emerg Med. 2006 Jun;13(3):134–8.

10. Damore D, Mansbach JM, Clark S, Ramundo M, Camargo CA. Prospective multicenter bronchiolitis study: predicting intensive care unit admissions. Acad Emerg Med Off J Soc Acad Emerg Med. 2008 Oct;15(10):887–94.
11. Macias CG, Mansbach JM, Fisher ES, Riederer M, Piedra PA, Sullivan AF, et al. Variability in inpatient management of children hospitalized with bronchiolitis. Acad Pediatr. 2015 Feb;15(1):69–76.
12. Mansbach JM, Clark S, Piedra PA, Macias CG, Schroeder AR, Pate BM, et al. Hospital course and discharge criteria for children hospitalized with bronchiolitis. J Hosp Med. 2015 Apr;10(4):205–11.
13. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care Lond Engl. 2019 Feb 22;23(1):64.
14. Goto T, Camargo CAJ, Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. Am J Emerg Med. 2018 Sep;36(9):1650–4.
15. Wellner B, Grand J, Canzone E, Coarr M, Brady PW, Simmons J, et al. Predicting unplanned transfers to the intensive care unit: A machine learning approach leveraging diverse clinical elements. JMIR Med Inform. 2017 Nov 22;5(4):e45.
16. Schanen J, Iribarren C, Shahar E, Punjabi N, Rich S, Sorlie P, et al. Asthma and incident cardiovascular disease: the Atherosclerosis Risk in Communities Study. Thorax. 2005 Aug;60(8):633–8.

17. Onufrak SJ, Abramson JL, Austin HD, Holguin F, McClellan WM, Vaccarino LV. Relation of adult-onset asthma to coronary heart disease and stroke. Am J Cardiol. 2008 May 1;101(9):1247–52.

18. Tattersall MC, Guo M, Korcarz CE, Gepner AD, Kaufman JD, Liu KJ, et al. Asthma predicts cardiovascular disease events: the multi-ethnic study of atherosclerosis. Arterioscler Thromb Vasc Biol. 2015 Jun;35(6):1520–5.

19. Dalager-Pedersen M, Søgaard M, Schønheyder HC, Nielsen H, Thomsen RW. Risk for myocardial infarction and stroke after community-acquired bacteremia: a 20-year population-based cohort study. Circulation. 2014 Apr 1;129(13):1387–96.

20. Musher DM, Abers MS, Corrales-Medina VF. Acute infection and myocardial infarction. N Engl J Med. 2019 10;380(2):171–6.

21. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, et al. Acute myocardial infarction after laboratory-confirmed influenza infection. N Engl J Med. 2018 Jan 25;378(4):345–53.

22. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. Eur Respir J. 2018;51(3).

23. Goto T, Shimada YJ, Faridi MK, Camargo CA, Hasegawa K. Incidence of acute cardiovascular event after acute exacerbation of COPD. J Gen Intern Med. 2018 Sep;33(9):1461–8.

24. Ralston SL, Lieberthal AS, Meissner HC, Alverson BK, Baley JE, Gadomski AM, et al. Clinical practice guideline: the diagnosis, management, and prevention of bronchiolitis. Pediatrics. 2014 Nov;134(5):e1474-1502.

25. Mansbach JM, Piedra PA, Stevenson MD, Sullivan AF, Forgey TF, Clark S, et al. Prospective multicenter study of children with bronchiolitis requiring mechanical ventilation. Pediatrics. 2012 Sep;130(3):e492-500.

26. Hasegawa K, Mansbach JM, Camargo CA. Infectious pathogens and bronchiolitis outcomes. Expert Rev Anti Infect Ther. 2014 Jul;12(7):817–28.

27. Mansbach JM, Piedra PA, Teach SJ, Sullivan AF, Forgey T, Clark S, et al. Prospective multicenter study of viral etiology and hospital length of stay in children with severe bronchiolitis. Arch Pediatr Adolesc Med. 2012 Aug;166(8):700–6.

28. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data-driven, machine learning approach. Acad Emerg Med Off J Soc Acad Emerg Med. 2016 Mar;23(3):269–78.

29. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med. 2018 Apr;46(4):547–53.

30. Goto T, Camargo CA, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. JAMA Netw Open. 2019 Jan 4;2(1):e186937.

31. Mansbach JM, Hasegawa K, Henke DM, Ajami NJ, Petrosino JF, Shaw CA, et al. Respiratory syncytial virus and rhinovirus severe bronchiolitis are associated with

distinct nasopharyngeal microbiota. J Allergy Clin Immunol. 2016 Jun;137(6):1909-1913.e4.

32. Hasegawa K, Mansbach JM, Ajami NJ, Espinola JA, Henke DM, Petrosino JF, et al. Association of nasopharyngeal microbiota profiles with bronchiolitis severity in infants hospitalised for bronchiolitis. Eur Respir J. 2016;48(5):1329–39.

33. Emergency Medicine Network [Internet]. [cited 2019 Feb 3]. Available from: http://www.emnet-usa.org/

34. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015 Oct 28;351:h5527.

35. Corneli HM, Zorc JJ, Holubkov R, Bregstein JS, Brown KM, Mahajan P, et al. Bronchiolitis: clinical characteristics associated with hospitalization and length of stay. Pediatr Emerg Care. 2012 Feb;28(2):99–103.

36. Yusuf S, Caviness AC, Adekunle-Ojo AO. Risk factors for admission in children with bronchiolitis from pediatric emergency department observation unit. Pediatr Emerg Care. 2012 Nov;28(11):1132–5.

37. Hasegawa K, Pate BM, Mansbach JM, Macias CG, Fisher ES, Piedra PA, et al. Risk factors for requiring intensive care among children admitted to ward with bronchiolitis. Acad Pediatr. 2015 Feb;15(1):77–81.

38. Stewart CJ, Mansbach JM, Wong MC, Ajami NJ, Petrosino JF, Camargo CA, et al. Associations of nasopharyngeal metabolome and microbiome with severity among infants with bronchiolitis. A multiomic analysis. Am J Respir Crit Care Med. 2017 01;196(7):882–91.

39. Luo G, Nkoy FL, Gesteland PH, Glasgow TS, Stone BL. A systematic review of predictive modeling for bronchiolitis. Int J Med Inf. 2014 Oct;83(10):691–714.

40. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R [Internet]. New York: Springer-Verlag; 2013 [cited 2018 Dec 1]. (Springer Texts in Statistics). Available from: //www.springer.com/la/book/9781461471370

41. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B Methodol. 1996;58(1):267–88.

42. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005 Apr;67(2):301–20.

43. Friedman J, Hastie T, Tibshirani R, Simon N, Narasimhan B, Qian J. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models [Internet]. 2018 [cited 2018 Dec 1]. Available from: https://CRAN.R-project.org/package=glmnet

44. Kuhn M. The caret Package [Internet]. [cited 2018 Dec 1]. Available from: http://topepo.github.io/caret/index.html

45. Kuhn M, Johnson K. Applied Predictive Modeling [Internet]. New York: Springer-Verlag; 2013 [cited 2018 Dec 1]. Available from: //www.springer.com/us/book/9781461468486

46. Cutler F original by LB and A, Wiener R port by AL and M. randomForest: Breiman and Cutler's Random Forests for Classification and Regression [Internet]. 2018 [cited 2019 Feb 3]. Available from: https://CRAN.R-project.org/package=randomForest

47. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobotics [Internet]. 2013 Dec 4 [cited 2018 Dec 1];7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3885826/

48. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. xgboost: Extreme Gradient Boosting [Internet]. 2018 [cited 2018 Dec 1]. Available from: https://CRAN.R-project.org/package=xgboost

49. Kuhn M, Wickham H, RStudio. recipes: Preprocessing Tools to Create Design Matrices [Internet]. 2018 [cited 2019 Jan 19]. Available from: https://CRAN.R-project.org/package=recipes

50. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988 Sep;44(3):837–45.

51. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: Display and Analyze ROC Curves [Internet]. 2020 [cited 2020 Feb 29]. Available from: https://CRAN.R-project.org/package=pROC

52. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk-prediction instruments: A critical review. Epidemiol Camb Mass. 2014 Jan;25(1):114–21.

53. Kundu S, Aulchenko YS, Janssens ACJW. PredictABEL: Assessment of Risk Prediction Models [Internet]. 2020 [cited 2020 Feb 29]. Available from: https://CRAN.R-project.org/package=PredictABEL

54. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565–574.

55. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. JAMA. 2015;313(4):409–10.

56. Rousson V, Zumbrunn T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. BMC Med Inform Decis Mak. 2011 Jun 22;11:45.

57. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: A guide for investigators. Eur Urol. 2018 Dec;74(6):796–804.

58. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016 Jan 25;352:i6.

59. Biostatistics: Decision Curve Analysis | Memorial Sloan Kettering Cancer Center [Internet]. [cited 2020 Feb 29]. Available from: https://www.mskcc.org/departments/epidemiology-biostatistics/biostatistics/decision-curve-analysis

60. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics [Internet]. 2019 [cited 2020 Feb 29]. Available from: https://CRAN.R-project.org/package=ggplot2

61. R: The R Project for Statistical Computing [Internet]. [cited 2020 Mar 2]. Available from: https://www.r-project.org/

62. Mansbach JM, Clark S, Christopher NC, LoVecchio F, Kunz S, Acholonu U, et al. Prospective multicenter study of bronchiolitis: predicting safe discharges from the emergency department. Pediatrics. 2008 Apr;121(4):680–8.

63. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li S-X, et al. Analysis of machine learning techniques for heart failure readmissions. Circ Cardiovasc Qual Outcomes. 2016;9(6):629–40.

64. Priesol AJ, Cao M, Brodley CE, Lewis RF. Clinical vestibular testing assessed with machine-learning algorithms. JAMA Otolaryngol-- Head Neck Surg. 2015 Apr;141(4):364–72.

65. Bacchi S, Oakden-Rayner L, Zerner T, Kleinig T, Patel S, Jannes J. Deep learning natural language processing successfully predicts the cerebrovascular cause of transient ischemic attack-like presentations. Stroke. 2019 Jan 17;STROKEAHA118024124.

66. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of emergency department hospital admission based on natural language processing and neural networks. Methods Inf Med. 2017 Oct 26;56(5):377–89.

67. Kolachalama VB, Singh P, Lin CQ, Mun D, Belghasem ME, Henderson JM, et al. Association of pathological fibrosis with renal survival using deep neural networks. Kidney Int Rep. 2018 Mar;3(2):464–75.

.

Paper 2

Risk of acute myocardial infarction and ischemic stroke in patients with asthma exacerbation: A population-based, self-controlled case series study

Authors: Yoshihiko Raita, MD, MPH[1]; Carlos A. Camargo, Jr. MD, DrPH[1]; Mohammad Kamal Faridi, MPH[1]; David F. M. Brown, MD[1]; Yuichi J. Shimada, MD, MPH[2]; and Kohei Hasegawa, MD, MPH[1]

Affiliations:

8. Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
9. Division of Cardiology, Department of Medicine, Columbia University Medical Center、New York, NY, USA

Corresponding Author: Dr. Yoshihiko Raita, Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, 125 Nashua Street, Suite 920, Boston, MA 02114-1101. Tel: 617-726-5276 E-mail:

Author contributions: Y.R. takes responsibility for the paper as a whole. Y.R., C.A.C., and K.H. conceived the study. C.A.C. obtained research funding. K.H. and D.F.M.B. supervised the conduct of the study. K. H. and Y.J.S. provided statistical advice. Y.R. and M.K.F. analyzed the data. Y.R. drafted the manuscript, and all authors contributed substantially to its revision.

Financial/non-financial disclosures: Dr. Hasegawa has received grants for asthma-related research from Novartis and Teva. Dr. Camargo has provided asthma-related consulting services to AstraZeneca and GlaxoSmithKline. The other authors have no financial relationships relevant to this article to disclose.

A declaration of all sources of funding: None

ABSTRACT

BACKGROUND: Patients with asthma have a high incidence of acute myocardial infarction and ischemic stroke.

OBJECTIVE: We aim to investigate the acute effect of asthma exacerbation on these cardiovascular events.

METHODS: Using population-based inpatient data of three geographically-diverse U.S. states (Florida, Nebraska, and New York) during 2011-2014, we conducted a self-controlled case series study of adults (aged ≥40 years) hospitalized with asthma exacerbation. The primary outcome was a composite of acute myocardial infarction and ischemic stroke. We used conditional Poisson regression to compare each patient's incidence rate of the outcome during three sequential risk periods (1-7, 8-14, and 15-28 days after asthma exacerbation) with that of reference period (i.e., summed period before and after the three risk periods).

RESULTS: We identified 4,607 adults hospitalized for asthma exacerbation who had a first episode of acute myocardial infarction or ischemic stroke. During the reference period, the incidence rate of acute myocardial infarction or ischemic stroke was 25.0/100 person-years. Compared with the reference period, the incidence rate significantly increased during the first risk period (129.1/100 person-years), with a corresponding adjusted incidence rate ratio (aIRR) of 5.04 (95%CI, 4.29-5.88; P<0.001). In the two subsequent risk periods, the incidence rate declined but remained high—50.1/100 person-years (aIRR 1.96; 95%CI, 1.51-2.48; P<0.001) and 38.0/100 person-years (aIRR, 1.48; 95%CI, 1.20-1.81; P<0.001), respectively. The findings were similar when the two outcomes were examined separately.

CONCLUSIONS: In this population-based study of adults with asthma, the risk of acute myocardial infarction and ischemic stroke increased significantly after asthma exacerbation.

KEY WORDS

Asthma exacerbation; myocardial infarction; ischemic stroke; self-controlled case series design

ABBREVIATIONS

aIRR, adjusted incidence rate ratio; COPD, chronic obstructive pulmonary disease; HCUP, the Healthcare Cost and Utilization Project; SID, State Inpatient Database; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; IQR, interquartile range; ARIC, Atherosclerosis Risk in Communities Study; MESA, Multi-Ethnic Study of Atherosclerosis

INTRODUCTION

Asthma and cardiovascular diseases are major public health problems in the U.S.(1,2) Approximately 26 million Americans have asthma.(1) Asthma exacerbations account for a substantial proportion of the personal and societal burden, leading to approximately 340,000 hospitalizations annually.(3) In parallel, cardiovascular disorders contribute to 610,000 acute myocardial infarction hospitalizations and 480,000 ischemic stroke hospitalizations in 2014.(3)

Studies have shown that patients with asthma—along with chronic obstructive pulmonary disease (COPD) and interstitial lung diseases—have a higher incidence of cardiovascular diseases (e.g., acute myocardial infarction, ischemic stroke)(4–8) through chronic activation of proinflammatory cytokines with resultant systemic and vascular inflammation.(5,9) Additionally, *acute* inflammatory processes (e.g., bacteremia,(10) pneumonia,(11) influenza virus infection,(12) Streptococcal infection,(13) and acute exacerbation of COPD(14)) have been linked to acute cardiovascular outcomes. However, despite the clinical and research importance of asthma exacerbations, little is known about its acute effect on cardiovascular outcomes.

To address the knowledge gap, using a large population-based database of adults with asthma from three diverse U.S. states, we investigated the association of asthma exacerbation with the incidence of acute myocardial infarction and ischemic stroke. We hypothesized that asthma exacerbations would be associated with an acute increase in cardiovascular risk. A better understanding of the role of asthma exacerbation in the development of cardiovascular events could inform preventive strategies during hospital discharge.

METHODS

Design and Setting

We performed a self-controlled case series study using the Healthcare Cost and Utilization Project (HCUP) State Inpatient Database (SID). We compared each patient's incidence of cardiovascular outcomes during sequential periods after asthma exacerbation. The self-controlled case series design is particularly suitable to estimate the effect of transient exposures (e.g., asthma exacerbation) on acute outcomes (e.g., cardiovascular events) within the same subject.(15,16) By performing intra-person comparisons in subjects with both the exposure and outcome of interest, each subject functions as her/his own control, and hence no separate controls are required. The major advantage of this design is that time-invariant confounders—regardless of measured or unmeasured—are implicitly controlled, thereby minimizing unmeasured confounding, which is an inherent problem in conventional cohort and case-control studies.

We used the SID from three geographically-diverse states—Florida, Nebraska, and New York—from January 2011 through December 2014. The SID includes all inpatient discharges—regardless of payer or disposition—from short-term, acute-care, nonfederal, general, and other specialty hospitals within the study states.(17) These three states were chosen for their high data quality in addition to geographical diversity and chiefly because their data included unique encrypted patient identifiers that enable longitudinal follow-up of specific patients across years. The details of SID may be found elsewhere.(17) The institutional review board of Massachusetts General Hospital waived review of this study.

Study Sample

Within these states, we first identified all adults aged ≥40 years with at least one hospitalization for asthma exacerbation by using *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* diagnosis code for asthma (493.00, 493.01,493.02,493.10, 493.11, 493.12, 493.20, 493.21, 493.22, 493.81, 493.82, 493.90, 493.91, and 493.92) in the primary diagnosis field, or those with a primary diagnosis of respiratory failure (518.81, 518.82, 518.84, and 799.1) plus a secondary diagnosis of asthma.(14,18,19) We used this case definition to focus on adults with poorly-controlled asthma with high morbidity rather than well-controlled asthma. Then, among these patients, we further identified patients with a first event of acute myocardial infarction (410, 410.01, 410.1, 410.11, 410.2, 410.21, 410.3, 410.31, 410.4, 410.41, 410.5, 410.51, 410.6, 410.61, 410.7, 410.71, 410.8, 410.81, 410.9, and 410.91) or ischemic stroke (433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434.01, 434.11, 434.91, and 436).(14,20) We used only the first episode because the study design assumes that recurrent events within each patient are independent.(13) We excluded out-of-state residents (300 patients), those with myocardial infarction or ischemic stroke as known comorbidities (244 patients) to ensure that patients did not have an outcome before the study period, and patients who died within 28 days from the first hospitalization for asthma exacerbation (47 patients).

Measurements

The SID databases contain information on the patients' characteristics, including demographics (age, sex, and race/ethnicity), primary insurance type (e.g., Medicare, Medicaid, private sources), quartiles for estimated household income, *ICD-9-CM* diagnosis and procedures, comorbidity measures, hospital length-of-stay, and disposition (including

inhospital deaths).(17) The baseline patient characteristics were determined at the time of first hospitalization for asthma exacerbation. Race/ethnicity was categorized into non-Hispanic white, non-Hispanic black, Hispanic, and other. Primary insurance types were categorized into Medicare, Medicaid, private sources, and other. We also included the baseline characteristics of quartile classifications of the estimated median household income of residents in the patient's ZIP code.

Outcome Measures

The primary outcome was a composite of acute myocardial infarction or ischemic stroke. In secondary analyses, we separately examined the acute myocardial infarction and ischemic stroke outcomes.

Statistical Analysis

We used conditional Poisson regression to compare each patient's incidence rate of outcomes occurring in the three consecutive risk periods (1-7, 8-14, and 15-28 days) after hospitalization for asthma exacerbation, compared with the non-risk period—the reference period (**Figure 1**).(13) We have chosen these risk periods based on previous studies and the number of outcome events.(12,13) Because each patient was matched to her/his risk and reference periods, the incidence rate ratios from the conditional regression were equivalent to person fixed effects. Therefore, the estimates are adjusted incidence rate ratios (aIRRs). For each patient with multiple hospitalizations for asthma exacerbation within 28 days, we employed the earliest date of hospitalization as the exposure.(13)

To examine the robustness of our inference, we performed several sensitivity analyses. First, we stratified the analysis by age (<65 vs ≥65 years), sex, race/ethnicity, primary insurance, household income quartile, and state of hospitals. We assumed that the age effect of each patient is time-invariant since we followed the maximum duration of four years. Second, we repeated the analysis excluding the patients with concurrent COPD (*ICD-9-CM* code of 491.21, 491.22 491.8, 491.9, 492.8, 493.20, 493.21, 493.22, or 496) as an ancillary diagnosis even though *ICD-9-CM* codes for asthma have a high specificity.[15] Third, we excluded the pre-exposure (0-14 days before asthma exacerbation) period from the reference period because developing cardiovascular events during this pre-exposure period may affect the subsequent probability of asthma exacerbation (**Figure 1E**)(15) (e.g., the use of β-blocker may trigger asthma exacerbation, more-intense care may increase or decrease the probability of subsequent hospitalization for asthma exacerbation), which could violate one of the assumptions in the self-controlled case series design. Fourth, we also excluded the patients who died within 30 days of outcome because the study design assumes that the outcome should not censor the study follow-up.(15) Lastly, we also repeated the analysis excluding the patients with a prolonged hospitalization (hospital length-of-stay >14 days) for asthma exacerbation which may affect the probability of developing an outcome during the risk periods. All P-values were 2-sided, and a level of <0.05 was considered significant. Analyses were performed using Stata 15 (Stata Corp, College Station, TX) and R 3.5.1.

RESULTS

Using the population-based data from three U.S. states, we identified 4,607 patients hospitalized for asthma exacerbation who also had first acute myocardial infarction or ischemic stroke during 2011-2014. The baseline patients' characteristics are summarized in **Table 1**. Overall, the median age was 70 years (IQR 60-80); most patients were female (68%) and non-Hispanic white (51%). Additionally, most patients had risk factors for cardiovascular diseases, such as hypertension (80%) and diabetes (46%).

The risk of developing an acute myocardial infarction or ischemic stroke (composite outcome) during the reference period and the consecutive three risk periods is shown in **Figure 2** and **Table 2**. During the reference period, the incidence rate was 25.0/100 person-years. In the subsequent risk period of 1-7 days after asthma exacerbation, the incidence rate significantly increased to 129.1/100 person-years with a corresponding aIRR of 5.04 (95% CI, 4.29-5.88). In the subsequent risk periods, the incidence rate decreased but remained high—50.1/100 person-years with a corresponding aIRR of 1.96 (95% CI, 1.51-2.48) during 8-14 days after asthma exacerbation, and 38.0/100 person-years with a corresponding aIRR of 1.48 (95% CI, 1.20-1.81) during 15-28 days after asthma exacerbation.

Looking at the two outcomes separately, we observed a similar temporal pattern (**Figure 2** and **Table 2**). Compared to the reference period, the incidence rate of acute myocardial infarction significantly increased during the first risk period (1-7 days after asthma exacerbation)—aIRR of 5.75 (95% CI, 4.75-6.90), followed by aIRR of 1.98 (95% CI, 1.43-2.66) and 1.38 (95% CI, 1.05-1.78) in the two subsequent risk periods. Likewise, the incidence rate of ischemic stroke also increased after asthma exacerbation—aIRR of 3.82

(95% CI, 2.80-5.07) in the first risk period, followed by aIRR of 1.91 (95% CI, 1.23-2.81) and 1.67 (95% CI, 1.19-2.25) in the two subsequent risk periods.

The sensitivity analyses demonstrated consistent results. Indeed, similar temporal patterns, particularly the high incidence rates in the first 7-day period, were found in the stratified analyses by age (<65 and ≥65 years; **Table E1**), sex (**Table E2**), race/ethnicity (**Table E3**), primary health insurance (**Table E4**), household income quartile (**Table E5**), and state of hospitals (**Table E6**), despite the limited statistical power for these stratified analyses and ischemic stroke outcome. Likewise, the subgroup analysis excluding the patients with concurrent COPD also demonstrated a consistent temporal pattern (**Table E7**). Furthermore, we yielded similar inferences across the sensitivity analyses that address possible violations of the analytic assumptions—i.e., the exclusion of the pre-exposure (0-14 days before asthma exacerbation) period from the reference period (**Table E8**), exclusion of the patients who died within 30 days of outcome (**Table E9**), and exclusion of the patients with prolonged hospitalization for asthma exacerbation (**Table E10**).

DISCUSSION

By applying the self-controlled case series design to the population-based data of 4,607 patients with asthma, we found that asthma exacerbation was associated with a significantly increased incidence of acute myocardial infarction and ischemic stroke—particularly in the first 1-week period. The observed patterns were consistent across several different patient subgroups—e.g., younger age group and patients without concurrent COPD—and with different statistical assumptions. This is the first study that has investigated the *acute* effect of asthma exacerbation on the risk of cardiovascular events.

The results are consistent with previous cohort studies that demonstrated that patients with chronic asthma have a higher incidence of cardiovascular events. For example, in the Atherosclerosis Risk in Communities (ARIC) Study, compared to adult women (age 45–64 years) without asthma, women with adult-onset asthma had a 1.8-fold higher hazard for incident coronary heart disease.(6) In a different analysis of the ARIC data, patients with a self-reported history of asthma had a 1.5-fold higher hazard of stroke.(4) Additional studies have also reported that the severity of chronic asthma is proportional to the risk of developing cardiovascular events.(22,23) In the Multi-Ethnic Study of Atherosclerosis (MESA), compared to adults without asthma and those with intermittent asthma, patients with persistent asthma had a higher C-reactive protein level and higher hazard for developing cardiovascular diseases.(22) Interestingly, previous studies also reported that the association between chronic asthma and cardiovascular outcomes is more evident in women compared to men (6,24) while no obvious between-sex differences were found in the current study except for the apparently smaller effect size in men for the ischemic

stroke outcome. The apparent discrepancies may be attributed to the differences in study population (i.e., patients with asthma exacerbation), settings (i.e., the use of population-based data), design (i.e., self-controlled case series design that accounts for all time-invariant confounders), outcomes, or any combination of these factors. In addition, stratified analyses in the subgroup of age<65 (**Table E1**), men (**Table E2**), and highest household income quartile (**Table E5**) demonstrated *non-significantly higher* aIRRs for the ischemic stroke outcome during the 1-7 days period. While potential heterogeneity of the effect is possible, these non-significant findings are attributable, at least partially, to the limited statistical power in these strata particularity with a fewer number of stroke outcomes. Regardless, the current study builds on these earlier reports, and extends them by demonstrating the effect of asthma exacerbation— acute inflammation on chronic disease processes—on cardiovascular events in a large population-based sample of patients with asthma.

The validity of our findings is buttressed by the application of self-controlled case series design. In this design, each patient functions as the control for herself/himself. Therefore, any time-invariant confounders—regardless of measured or unmeasured—are cancelled out, thereby yielding robust causal inference. This design also is more efficient than other study designs because inter-person variations are removed, and hence the estimates of effects can be more precise. Furthermore, the current study with the use of asthma exacerbation as the (transient) exposure and cardiovascular events as the (acute) outcomes meets the required assumptions of self-controlled case series design.

There are several potential mechanisms linking asthma exacerbation to the increased incidence of cardiovascular events. The most common cause of asthma exacerbation is

acute respiratory infections,(25,26) leading to an activation of inflammatory pathways and cytokines (e.g., CXCL-10, interleukin-5, interleukin-6, tumor necrosis factor-α)(27,28) and resultant systemic vascular inflammation with platelet activation, inhibition of fibrinolysis, and elevation of C-reactive protein levels.(9,29–31) High-sensitivity C-reactive protein—an established biomarker for cardiovascular outcomes—upregulates other inflammatory regulators, leading to the adhesion of leukocyte to arterial endothelium.(32) Platelet activation and endothelial dysfunction cause arterial thrombosis. Specifically, in acute inflammation, thin-cap atheroma in coronary arteries ruptures and releases inflammatory cells, causing acute accumulation of platelets, neutrophils, and fibrin as well as trapping of red blood cells, which are characteristics in type 1 myocardial infarction.(11) This potential mechanism is supported by the finding from an analysis of the Nurses' Health Study, which showed that the patients with persistent asthma who use inhaled corticosteroids had a lower incidence of cardiovascular disease compared to those who did not use.(33) This finding suggests that a potential role of anti-inflammatory effects (and that of improved asthma control) in cardiovascular event risks among patients with asthma. Another potential mechanism is the mismatch of oxygen demand and supply in myocardium, specifically characterized by type 2 myocardial infarction. During the process of acute inflammation, interleukin-1, tumor necrosis factor-α, and catecholamines are released, thereby increasing the body core temperature, heart rate, afterload, and oxygen demand. Additionally, patients with asthma exacerbation often experience hypoxemia, resulting in insufficient oxygen supply to myocardium. Consequently, the metabolic demand of myocardium exceeds the supply of oxygen from coronary arteries. In addition, vasospasm from vascular smooth muscle cell contraction and endothelial dysfunction could also

contribute to acute myocardial infarction in patients with asthma. Both asthma and coronary vasospasm share common pathobiologies (e.g., autonomic dysfunction, histamine provocation).(34) Furthermore, management for asthma exacerbation (e.g., excessive use of β_2-agonists,(35) discontinuation of β-blockers,(36) discontinuation of aspirin in patients with aspirin-exacerbated respiratory disease) may have attributed to the subsequent cardiovascular event risks. Moreover, these mechanisms are not mutually exclusive. Notwithstanding the complexity, the identification of the asthma exacerbation-cardiovascular event link is an important finding. Our finding should advance the research into the development of preventive strategies for cardiovascular morbidities (e.g., targeting systemic inflammation among patients with high inflammatory markers(37)) in patients with asthma.

This study has several potential limitations. First, our population-based study is potentially limited by not being a random sample of the nationwide asthma populations. However, we used the population-based data from the selected racially/ethnically- and geographically-diverse states which comprise 13% of U.S. population.(38) Second, we identified the study sample from administrative data with possible misclassification of hospitalizations for asthma exacerbation. Yet, HCUP data are accurate, rigorously tested, and widely used to estimate diagnoses(18,19,39) and the *ICD-9-CM* codes for asthma are known to have high specificity (98%).(21) Additionally, our sensitivity analysis demonstrated consistent results even after excluding the patients with concurrent COPD. Third, misclassifications of outcomes are also possible. However, the literature has shown that the *ICD 9-CM* coding of these outcomes are accurate (specificity of 89%-99% for acute myocardial infarction and 95% for ischemic stroke).(20,40–42) Fourth, we were unable to differentiate the subtypes

of myocardial infarction (i.e., type 1 and type 2 myocardial infarction) using *ICD-9-CM* codes. However, both types of myocardial infarction can be explained by activated cytokine response and other pathobiology caused by asthma exacerbation. In addition, the increased risk of ischemic stroke also supports the possibility of thromboembolic pathobiology rather than demand ischemia. Fifth, SID do not have some helpful clinical data. For example, the differentiation between type 1 and type 2 myocardial infarction may be helpful for understanding the exact mechanisms linking asthma exacerbation to the outcome. Sixth, we excluded patients who died within 28 days from the first hospitalization for asthma exacerbation as these patients did not have a risk period and an adjusted incidence rate ratio could not be computed. Additionally, due to the lack of out-of-hospital mortality data in SID, we were not able to account for the competing risk of mortality. However, these mechanisms should have biased the inferences toward the null through excluding patients who might have had an outcome during the risk periods and shortening the reference period. Seventh, we did not account for time-varying confounding in this study. For example, medication such as β-blockers might be a time-varying confounder. However, we have followed up patients for the short period of time (i.e., 4 years) and as the reference period, we used summed period before and after the three risk periods. Therefore, we have mitigated the effect of time-varying confounding within a short duration of time. Lastly, even with our racially/ethnically- and geographically-diverse population-based sample, we must generalize the inferences cautiously beyond patients with more severe asthma exacerbations. Still, the data remain highly relevant for 340,000 hospitalized adults each year.(3) This population with large healthcare use and morbidity burden is the one for which targeted prevention strategies are most urgently required.

FIGURES

Figure 1. Example self-controlled case series timeline

This timeline shows the risk period divided into three periods: 1–7, 8–14, and 15–28 days after hospitalization for asthma exacerbation. The orange solid line represents the first risk period; the green solid line represents the second risk period; and the blue solid line represents the third risk period. The grey solid and dashed lines indicate the reference periods. The incidence rate ratio for the first acute myocardial infarction and ischemic stroke occurring within each risk period, compared with the reference period, was calculated conditioning on each individual. Patients may have multiple index hospitalizations for asthma exacerbation.

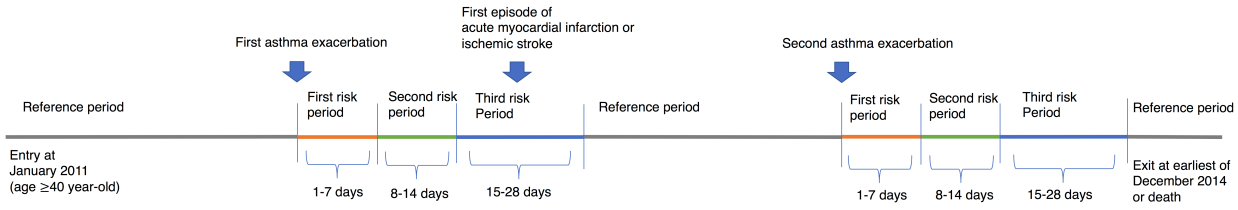
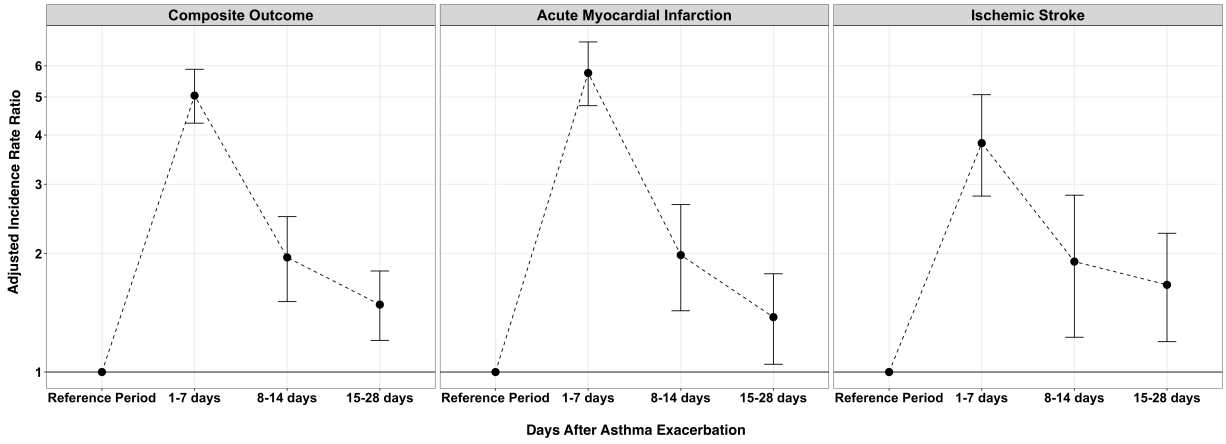


Figure 2. Adjusted incidence rate ratio for acute myocardial infarction and ischemic stroke after asthma exacerbation

The adjusted incidence rate ratios are demonstrated in a log-scale. The error bars indicate 95% confidence intervals.



TABLES

Table 1. Baseline characteristics of 4,607 patients hospitalized with asthma exacerbation

Characteristics	n	(%)
Age, median (IQR), years	70	(60-80)
Age ≥65 years	2,964	(64.3)
Female sex	3,126	(67.9)
Race/ethnicity		
Non-Hispanic white	2,369	(51.4)
Non-Hispanic black	988	(21.4)
Hispanic	869	(18.9)
Others	315	(6.9)
Missing[a]	66	(1.4)
Primary health insurance		
Medicare	3,311	(71.9)
Medicaid	639	(13.9)
Private	457	(9.9)
Self-pay	111	(2.4)
Others	90	(2.0)
Quartiles for median household income		
1 (lowest)	1,681	(36.5)
2	1,046	(22.7)
3	998	(21.7)
4 (highest)	697	(15.1)
Missing	185	(4.0)
Selected comorbidities[b]		
Hypertension	3,701	(80.3)
Diabetes	2,117	(45.9)
Congestive heart failure	1,546	(33.6)
Obesity	978	(21.2)
Renal disease	908	(19.7)
Peripheral vascular diseases	349	(7.6)
Anemia	170	(3.7)
Liver disease	88	(1.9)
State		
Florida	2,454	(53.3)
Nebraska	58	(1.3)
New York	2,095	(45.5)

Data are expressed as number (%), unless otherwise
Percentages may not equal 100, because of rounding
[a] Race/ethnicity data were not available in Nebraska.
[b] Selected from 29 Elixhauser comorbidity measures

Table 2. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke in patients with asthma exacerbation (n=4,607)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Reference period	25.0 (24.3-25.8)	Reference		25.0 (24.1-26.0)	Reference		25.0 (23.9-26.3)	Reference	
Time after asthma exacerbation: 1-7 days	129.1 (110.8-150.4)	5.04 (4.29-5.88)	<0.001	147.4 (123.1-176.4)	5.75 (4.75-6.90)	<0.001	97.7 (73.2-130.5)	3.82 (2.80-5.07)	<0.001
8-14 days	50.1 (39.2-64.0)	1.96 (1.51-2.48)	<0.001	50.8 (37.4-69.0)	1.98 (1.43-2.66)	<0.001	48.9 (32.5-73.5)	1.91 (1.23-2.81)	0.002
15-28 days	38.0 (31.1-46.3)	1.48 (1.2-01.81)	<0.001	35.3 (27.2-45.8)	1.38 (1.05-1.78)	0.02	42.5 (31.2-58.0)	1.67 (1.19-2.25)	0.002

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence interval

[a] Incidence rate per 100 person-years, [b] n=2,847, [c] n=1,760

SUPPLEMENTARY MATERIAL

Risk of acute myocardial infarction and ischemic stroke in patients with asthma exacerbation: A population-based, self-controlled case series study

Authors: Yoshihiko Raita, MD, MPH; Carlos A. Camargo, Jr. MD, DrPH; Mohammad Kamal Faridi, MPH; David F. M. Brown, MD; Yuichi J. Shimada, MD, MPH; and Kohei Hasegawa, MD, MPH

Figure E1. Example self-controlled case series timeline, excluding the pre-exposure (0-14 days before asthma exacerbation) period from the reference period

This timeline shows the risk period divided into three periods: 1–7, 8–14, and15–28 days after hospitalization for asthma exacerbation. The orange solid line represents the first risk period; the green solid line represents the second risk period, the blue solid line represents the third risk period, and the blue dashed line represents excluded time. The grey solid and dashed lines indicate the reference periods. The incidence ratio for the first acute myocardial infarction and ischemic stroke occurring within each risk period compared with baseline time was calculated for each individual. Patients may have multiple index hospitalizations for asthma exacerbation.

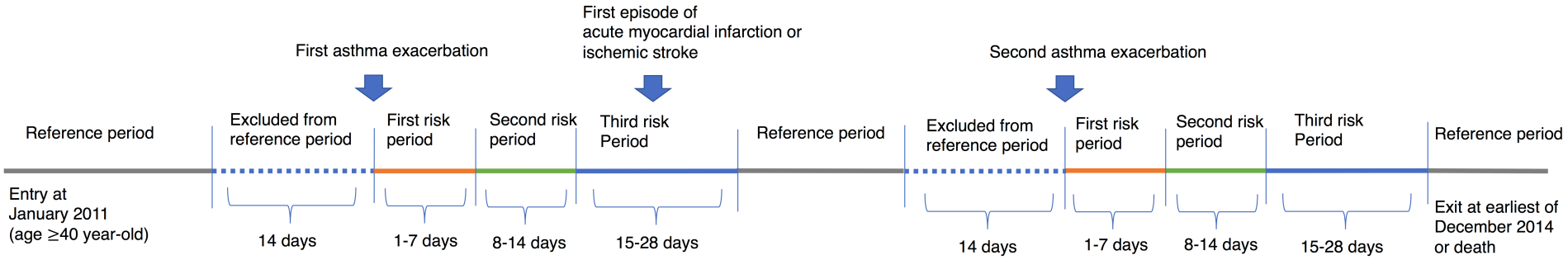


Table E1. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke in patients with asthma exacerbation, stratified by age (n=4,607)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Age <65 years (n=1,643)									
Reference period	24.8 (26.1-26.1)	Reference		24.7 (23.2-26.3)	Reference		25.1 (23.1-27.2)	Reference	
Time after exacerbation: 1-7 days	110.3 (143.0-171.3)	4.48 (3.39-5.80)	<0.001	148.5 (112.2-196.4)	6.10 (4.49-8.07)	<0.001	42.8 (21.4-85.6)	1.70 (0.77-3.21)	0.14
8-14 days	42.6 (64.6-74.7)	1.73 (1.1-2.57)	0.02	39.4 (22.9-67.8)	1.62 (0.88-2.68)	0.09	48.2 (25.1-92.6)	1.92 (0.91-3.50)	0.06
15-28 days	28.1 (40.4-56.7)	1.14 (0.77-1.62)	0.32	25.8 (16.0-41.4)	1.06 (0.63-1.66)	0.82	32.1 (18.2-56.6)	1.28 (0.68-2.17)	0.41
Age ≥65 years (n=2,964)									
Reference period	25.2 (24.1-26.0)	Reference		25.3 (24.1-26.5)	Reference		25.0 (23.6-26.6)	Reference	
Time after exacerbation: 1-7 days	141.9 (124.5-176.8)	5.38 (4.40-6.50)	<0.001	146.6 (116-185.3)	5.53 (4.31-6.99)	<0.001	133.9 (97.4-184)	5.12 (3.63-6.99)	<0.001
8-14 days	55.2 (36.8-67.4)	2.09 (1.52-2.80)	<0.001	58.7 (40.5-85.0)	2.21 (1.48-3.15)	<0.001	49.3 (29.2-83.3)	1.89 (1.06-3.07)	0.02
15-28 days	44.7 (27.6-45.9)	1.70 (1.32-2.14)	<0.001	41.9 (30.7-57.1)	1.58 (1.13-2.14)	0.004	49.3 (34.1-71.5)	1.90 (1.27-2.71)	<0.001

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,847, [c] n=1,760

Table E2. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by sex (n=4,607)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Men (n=1,481)									
Reference period	25.1 (23.8-26.5)	Reference		24.9 (23.3-26.6)	Reference		25.6 (23.4-28.0)	Reference	
Time after exacerbation: 1-7 days	113.4 (84.7-151.9)	4.41 (3.22-5.88)	<0.001	145.0 (106.0-198.5)	5.71 (4.07-7.79)	<0.001	46.9 (21.1-104.4)	1.77 (0.70-3.64)	0.17
8-14 days	58.0 (38.5-87.2)	2.26 (1.45-3.33)	<0.001	66.9 (42.2-106.3)	2.64 (1.59-4.08)	<0.001	39.1 (16.3-93.9)	1.48 (0.53-3.21)	0.39
15-28 days	41.6 (29.6-58.5)	1.63 (1.13-2.26)	0.006	44.6 (29.9-66.6)	1.76 (1.14-2.58)	0.006	35.2 (18.3-67.7)	1.35 (0.64-2.47)	0.38
Women (n=3,126)									
Reference period	25.0 (24.1-25.9)	Reference		25.1 (24.0-26.3)	Reference		24.8 (23.5-26.3)	Reference	
Time after exacerbation: 1-7 days	136.2 (113.9-162.9)	5.32 (4.40-6.38)	<0.001	148.6 (119.3-185.0)	5.77 (4.56-7.19)	<0.001	116.7 (85.6-159.1)	4.61 (3.3-6.25)	<0.001
8-14 days	46.5 (34.3-63.2)	1.82 (1.31-2.44)	<0.001	42.7 (28.4-64.3)	1.66 (1.07-2.44)	0.017	52.5 (33.1-83.4)	2.08 (1.25-3.21)	0.002
15-28 days	36.3 (28.4-46.4)	1.42 (1.10-1.81)	0.005	30.6 (21.8-43.1)	1.19 (0.82-1.65)	0.33	45.2 (31.8-64.3)	1.79 (1.22-2.51)	0.002

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,847, [c] n=1,760

Table E3 Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by race/ethnicity (n=4,607)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Non-Hispanic white (n=2,369)									
Reference period	25.1 (24.1-26.2)	Reference		25.2 (23.9-26.6)	Reference		24.9 (23.3-26.7)	Reference	
Time after exacerbation: 1-7 days	133.4 (106.8-166.5)	5.13 (4.05-6.4)	<0.001	150.7 (115.9-195.8)	5.70 (4.30-7.39)	<0.001	103.3 (68.0-156.8)	4.10 (2.60-6.11)	<0.001
8-14 days	49.6 (34.5-71.4)	1.91 (1.29-2.70)	<0.001	48.4 (30.5-76.9)	1.83 (1.11-2.83)	0.01	51.6 (28.6-93.2)	2.05 (1.06-3.53)	0.02
15-28 days	37.6 (28.-50.6)	1.45 (1.06-1.93)	0.02	33.6 (22.7-49.8)	1.27 (0.83-1.85)	0.23	44.6 (28.4-69.9)	1.77 (1.08-2.71)	0.01
Non-Hispanic black (n=988)									
Reference period	24.8 (23.2-26.5)	Reference		25.0 (22.9-27.3)	Reference		24.6 (22.3-27.1)	Reference	
Time after exacerbation: 1-7 days	120.5 (88.4-164.3)	4.88 (3.48-6.65)	<0.001	123.9 (83-184.9)	5.01 (3.21-7.44)	<0.001	115.7 (70.9-188.9)	4.69 (2.71-7.54)	<0.001
8-14 days	48.2 (29.5-78.7)	1.95 (1.14-3.10)	0.008	41.3 (20.7-82.6)	1.67 (0.76-3.15)	0.15	57.9 (28.9-115.7)	2.35 (1.06-4.44)	0.02
15-28 days	39.2 (26.7-57.5)	1.59 (1.04-2.30)	0.02	33.6 (19.5-57.8)	1.36 (0.74-2.27)	0.28	47.0 (27.3-81.0)	1.91 (1.03-3.20)	0.02
Hispanic (n=869)									
Reference period	25.2 (23.5-27.0)	Reference		24.9 (22.8-27.1)	Reference		25.8 (23.0-29.0)	Reference	
Time after exacerbation: 1-7 days	143.1 (103.7-197.5)	5.47 (3.85-7.54)	<0.001	180.6 (127-256.8)	7.14 (4.84-10.15)	<0.001	69.1 (31-153.7)	2.47 (1.09-5.59)	0.03
8-14 days	46.4 (26.4-81.7)	1.78 (0.94-3.00)	0.05	52.4 (27.3-100.8)	2.07 (0.99-3.78)	0.03	34.5 (11.1-107.1)	1.24 (0.39-3.88)	0.72
15-28 days	31.0 (19.0-50.5)	1.19 (0.69-1.89)	0.49	29.1 (15.7-54.1)	1.15 (0.57-2.04)	0.66	34.6 (15.5-77.0)	1.26 (0.56-2.85)	0.57

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,847, [c] n=1,760

Table E4 Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by primary health insurance (n=4607)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Public (Medicare and Medicaid) (n=3,950)									
Reference period	25.2 (24.4-26.0)	Reference		25.2 (24.2-26.2)	Reference		25.1 (23.9-26.5)	Reference	
Time after exacerbation: 1-7 days	126.5 (107-149.4)	4.89 (4.09-5.78)	<0.001	144.7 (118.8-176.2)	5.60 (4.54-6.83)	<0.001	95.8 (70.0-131.1)	3.69 (2.63-5.02)	<0.001
8-14 days	50.4 (38.7-65.6)	1.95 (1.47-2.52)	<0.001	49.7 (35.5-69.6)	1.92 (1.34-2.66)	<0.001	51.6 (33.6-79.1)	1.99 (1.25-2.98)	0.002
15-28 days	38.5 (31.1-47.7)	1.49 (1.19-1.84)	<0.001	36.6 (27.7-48.2)	1.41 (1.05-1.85)	0.02	41.8 (29.8-58.5)	1.62 (1.13-2.24)	0.006
Private (n=457)									
Reference period	24.6 (22.4-27.1)	Reference		24.6 (21.9-27.7)	Reference		24.6 (21.0-28.9)	Reference	
Time after exacerbation: 1-7 days	128.4 (78.6-209.5)	5.17 (3.00-8.27)	<0.001	140.3 (79.7-247)	5.64 (2.97-9.67)	<0.001	102.2 (38.4-272.4)	4.16 (1.27-9.87)	0.005
8-14 days	32.1 (12-85.5)	1.29 (0.40-3.03)	0.61	35.1 (11.3-108.8)	1.41 (0.35-3.70)	0.56	25.6 (3.6-181.5)	1.04 (0.06-4.64)	0.97
15-28 days	40.1 (21.6-74.5)	1.62 (0.80-2.87)	0.14	35.1 (15.8-78.1)	1.41 (0.55-2.90)	0.41	51.1 (19.2-136.2)	2.08 (0.64-4.94)	0.15

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,847, [c] n=1,760

Table E5 Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by household income quartile (n=4,607)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Quartile 1 (lowest) (n=1,681)									
Reference period	25.3 (24.1-26.6)	Reference		25.2 (23.6-26.8)	Reference		25.4 (23.5-27.5)	Reference	
Time after exacerbation: 1-7 days	113.6 (87.6-147.3)	4.36 (3.30-5.64)	<0.001	134.9 (99.7-182.6)	5.21 (3.75-7.03)	<0.001	78.8 (47.5-130.7)	2.97 (1.69-4.82)	<0.001
8-14 days	37.9 (24.2-59.4)	1.45 (0.89-2.22)	0.11	41.8 (24.3-71.9)	1.61 (0.88-2.67)	0.09	31.5 (14.2-70.1)	1.19 (0.47-2.44)	0.67
15-28 days	38.9 (28.4-53.2)	1.50 (1.07-2.03)	0.01	40.2 (27.1-59.4)	1.55 (1.01-2.26)	0.03	36.8 (21.8-62.1)	1.40 (0.78-2.30)	0.22
Quartile 2 (n=1,046)									
Reference period	25.0 (23.5-26.6)	Reference		25.1 (23.1-27.1)	Reference		24.9 (22.5-27.6)	Reference	
Time after exacerbation: 1-7 days	150.1 (110.5-203.9)	5.88 (4.22-7.94)	<0.001	159.2 (109.9-230.6)	6.21 (4.14-8.93)	<0.001	133.7 (77.6-230.3)	5.27 (2.87-8.8)	<0.001
8-14 days	51.3 (30.4-86.6)	2.01 (1.13-3.27)	0.01	51.2 (26.6-98.4)	2.00 (0.95-3.63)	0.04	51.4 (21.4-123.5)	2.03 (0.72-4.4)	0.12
15-28 days	34.8 (22.2-54.6)	1.36 (0.83-2.09)	0.18	31.3 (17.3-56.5)	1.22 (0.63-2.11)	0.51	41.1 (20.6-82.3)	1.62 (0.73-3.05)	0.18
Quartile 3 (n=998)									
Reference period	24.6 (23.1-26.3)	Reference		24.9 (22.9-27.1)	Reference		24.3 (26.9-27.6)	Reference	
Time after exacerbation: 1-7 days	143.5 (104.4-197.2)	5.72 (4.06-7.83)	<0,001	146.4 (98.1-218.4)	5.80 (3.73-8.56)	<0.001	138.8 (234.3-230.3)	5.60 (3.11-9.22)	<0.001
8-14 days	64.2 (39.9-103.3)	2.56 (1.52-4.01)	<0.001	73.2 (41.6-128.9)	2.90 (1.54-4.92)	<0.001	49.6 (119.1-123.5)	2.00 (0.71-4.34)	0.13
15-28 days	49.1 (33.4-72.1)	1.96 (1.29-2.84)	<0.001	36.6 (20.8-64.4)	1.45 (0.77-2.46)	0.21	69.4 (117.2-82.3)	2.80 (1.56-4.61)	<0.001
Quartile 4 (highest) (n=697)									
Reference period	25.2 (23.3-27.2)	Reference		25.1 (22.8-27.7)	Reference		25.3 (22.3-28.6)	Reference	

Time after exacerbation: 1-7 days	118.2 (77.1-181.3)	4.53 (2.83-6.84)	<0.001	151 (93.9-242.9)	5.72 (3.36-9.08)	<0.001	61.5 (23.1-163.9)	2.41 (0.74-5.68)	0.08
8-14 days	56.3 (30.3-104.7)	2.16 (1.07-3.82)	0.02	35.5 (13.3-94.7)	1.35 (0.42-3.17)	0.56	92.3 (41.5-205.4)	3.61 (1.42-7.45)	0.02
15-28 days	31.0 (17.2-55.9)	1.19 (0.61-2.05)	0.58	35.5 (17.8-71.1)	1.35 (0.61-2.55)	0.41	23.1 (7.4-71.5)	0.90 (0.22-2.37)	0.86

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,847, [c] n=1,760

Table E6 Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke stratified by state (n=4,607)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Florida (n=2,454)									
Reference period	25.1 (24.1-26.1)	Reference		25.2 (24.0-26.6)	Reference		24.8 (23.2-26.6)	Reference	
Time after exacerbation: 1-7 days	127.2 (102.5-158)	4.94 (3.93-6.13)	<0.001	127.5 (97.2-167.4)	4.91 (3.67-6.42)	<0.001	126.7 (88.6-181.2)	5.00 (3.39-7.09)	<0.001
8-14 days	41.9 (28.7-61.1)	1.63 (1.08-2.33)	0.01	46.6 (29.7-73.1)	1.79 (1.10-2.74)	0.01	33.8 (16.9-67.5)	1.33 (0.61-2.50)	0.42
15-28 days	34.9 (26.1-46.8)	1.36 (0.99-1.8)	0.04	30.7 (20.7-45.4)	1.18 (0.77-1.72)	0.41	42.2 (27.2-65.4)	1.67 1.03-2.53)	0.03
Nebraska (n=58)									
Reference period	24.1 (18.4-31.5)	Reference		22.9 (16.6-31.6)	Reference		27.3 (16.7-44.6)	Reference	
Time after exacerbation: 1-7 days	77.8 (11.0-552.5)	3.18 (0.18-14.53)	0.25	110.9 (15.6-787.6)	4.85 (0.27-22.4)	0.12	0	NA[d]	NA[d]
8-14 days	155.7 (38.9-622.4)	6.36 (1.04-20.55)	0.01	221.9 (55.5-887.2)	9.70 (1.57-31.8)	0.002	0	NA[d]	NA[d]
15-28 days	77.8 (19.5-311.2)	3.18 (0.52-10.28)	0.11	110.9 (27.7-443.6)	4.85 (0.79-15.9)	0.03	0	NA[d]	NA[d]
New York (n=2,095)									
Reference period	25.0 (23.9-26.2)	Reference		24.9 (23.5-26.4)	Reference		25.2 (23.5-27.1)	Reference	
Time after exacerbation: 1-7 days	132.1 (106.4-164.1)	5.19 (4.11-6.45)	<0.001	168.9 (132.7-215)	6.70 (5.15-8.55)	<0.001	69.6 (42.6-113.6)	2.69 (1.57-4.28)	<0.001
8-14 days	56.4 (40.5-78.5)	2.21 (1.55-3.05)	<0.001	51.2 (33-79.4)	2.03 (1.26-3.08)	0.002	65.2 (39.3-108.2)	2.53 (1.44-4.07)	<0.001
15-28 days	40.3 (30.5-53.2)	1.59 (1.18-2.08)	0.001	38.4 (26.8-54.9)	1.52 (1.03-2.15)	0.02	43.5 (28.1-67.4)	1.70 (1.05-2.58)	0.02

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,847, [c] n=1,760

[d] Nebraska had only 16 cases of ischemic stroke in the reference period.

Table E7. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke in patients with asthma exacerbation, excluding patients with concurrent chronic obstructive pulmonary disease (n=4,030)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Reference period	25.0 (24.2-25.8)	Reference		25.0 (24.0-26.0)	Reference		25.1 (23.8-26.4)	Reference	
Time after exacerbation: 1-7 days	133.2 (113.5-156.4)	5.21 (4.4-6.12)	<0.001	153.4 (127-185.2)	6.01 (4.91-7.27)	<0.001	99.6 (73.6-134.8)	3.89 (2.81-5.23)	<0.001
8-14 days	52.4 (40.6-67.6)	2.05 (1.57-2.63)	<0.001	55.4 (40.5-75.8)	2.17 (1.55-2.94)	<0.001	47.4 (30.6-73.5)	1.85 (1.15-2.8)	0.006
15-28 days	38.2 (30.9-47.2)	1.50 (1.20-1.84)	<0.001	36.2 (27.5-47.7)	1.42 (1.06-1.86)	0.01	41.5 (29.8-57.8)	1.63 (1.14-2.24)	0.004

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,460, [c] n=1,570

Table E8. Incidence rate and adjusted incidence rate ratio (IRR) for acute myocardial infarction and ischemic stroke, excluding the pre-exposure period from the reference period (n=4,542)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Reference period	25.0 (24.3-25.8)	Reference		25.0 (24.1-26.0)	Reference		25.0 (23.8-26.2)	Reference	
Time after exacerbation: 1-7 days	130.8 (112.2-152.4)	5.11 (4.35-5.97)	<0.001	149.4 (124.8-179)	5.84 (4.82-7)	<0.001	99.1 (74.2-132.3)	3.88 (2.85-5.15)	<0.001
8-14 days	51.1 (40.0-65.2)	2.00 (1.54-2.53)	<0.001	51.9 (38.2-70.5)	2.03 (1.46-2.73)	<0.001	49.6 (32.9-74.6)	1.94 (1.25-2.86)	0.002
15-28 days	38.7 (31.7-47.2)	1.51 (1.23-1.84)	<0.001	36.1 (27.8-46.8)	1.41 (1.07-1.82)	0.01	43.1 (31.6-58.8)	1.69 (1.21-2.29)	0.001

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,801, [c] n=1,741

Table E9. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke, excluding patients who died within 30 days of outcome (n=4,570)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Reference period	25.0 (24.2-25.7)	Reference		25.0 (24.0-26.0)	Reference		25.0 (23.8-26.2)	Reference	
Time after exacerbation: 1-7 days	127.6 (109.4-148.9)	5.02 (4.27-5.86)	<0.001	147.1 (122.8-176.2)	5.78 (4.77-6.94)	<0.001	94.2 (70.1-126.5)	3.71 (2.7-4.95)	<0.001
8-14 days	50.4 (39.5-64.4)	1.98 (1.53-2.52)	<0.001	51.1 (37.6-69.4)	2.01 (1.45-2.7)	<0.001	49.2 (32.7-74.1)	1.94 (1.25-2.86)	0.002
15-28 days	37.4 (30.6-45.8)	1.47 (1.19-1.80)	<0.001	34.3 (26.3-44.7)	1.35 (1.02-1.75)	0.02	42.8 (31.4-58.4)	1.69 (1.21-2.28)	0.001

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,827, [c] n=1,743

Table E10. Incidence rate and adjusted incidence rate ratio (aIRR) for acute myocardial infarction and ischemic stroke, excluding patients with prolonged hospitalization for asthma exacerbation (n=4,398)

	Composite outcome			Acute myocardial infarction[b]			Ischemic stroke[c]		
Time interval	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value	Incidence rate[a] (95% CI)	aIRR (95% CI)	P-value
Reference period	25 .0 (24.3-25.8)	Reference		25.0 (24.1-26.0)	Reference	<0.001	25.0 (23.8-26.3)	Reference	
Time after exacerbation: 1-7 days	127.8 (109-149.7)	4.99 (4.22-5.85)	<0.001	145.7 (120.9-175.7)	5.70 (4.67-6.88)	<0.001	97.1 (72-131)	3.78 (2.75-5.07)	<0.001
8-14 days	47.6 (36.7-61.7)	1.86 (1.41-2.39)	<0.001	47.7 (34.4-66.1)	1.87 (1.32-2.55)	<0.001	47.4 (30.9-72.8)	1.85 (1.16-2.77)	0.009
15-28 days	38.0 (30.9-46.7)	1.49 (1.20-1.82)	<0.001	35.8 (27.4-46.7)	1.40 (1.05-1.82)	0.02	41.8 (30.3-57.7)	1.63 (1.16-2.23)	<0.001

Abbreviations: aIRR, adjusted incidence rate ratio; CI, confidence intervals

[a] Incident rate per 100 person-years, [b] n=2,709, [c] n=1,689

REFERENCES

1. Most Recent National Asthma Data | CDC [Internet]. 2019 [cited 2019 Apr 20]. Available from:https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm

2. Heart Disease Facts & Statistics | cdc.gov [Internet]. 2018 [cited 2019 Apr 20]. Available from: https://www.cdc.gov/heartdisease/facts.htm

3. HCUPnet [Internet]. HCUPnet. [cited 2018 Nov 27]. Available from: https://hcupnet.ahrq.gov

4. Schanen J, Iribarren C, Shahar E, Punjabi N, Rich S, Sorlie P, et al. Asthma and incident cardiovascular disease: the Atherosclerosis Risk in Communities Study. Thorax. 2005 Aug;60(8):633–8.

5. Carter P, Lagan J, Fortune C, Bhatt DL, Vestbo J, Niven R, et al. Association of cardiovascular disease with respiratory disease. J Am Coll Cardiol. 2019 Feb 27;

6. Onufrak SJ, Abramson JL, Austin HD, Holguin F, McClellan WM, Vaccarino LV. Relation of adult-onset asthma to coronary heart disease and stroke. Am J Cardiol. 2008 May 1;101(9):1247–52.

7. Chung W-S, Lin C-L, Chen Y-F, Ho F-M, Hsu W-H, Kao C-H. Increased stroke risk among adult asthmatic patients. Eur J Clin Invest. 2014 Nov;44(11):1025–33.

8. Iribarren C, Tolstykh IV, Eisner MD. Are patients with asthma at increased risk of coronary heart disease? Int J Epidemiol. 2004 Aug;33(4):743–8.

9. Van Eeden S, Leipsic J, Paul Man SF, Sin DD. The relationship between lung inflammation and cardiovascular disease. Am J Respir Crit Care Med. 2012 Jul 1;186(1):11–6.

10. Dalager-Pedersen M, Søgaard M, Schønheyder HC, Nielsen H, Thomsen RW. Risk for myocardial infarction and stroke after community-acquired bacteremia: a 20-year population-based cohort study. Circulation. 2014 Apr 1;129(13):1387–96.

11. Musher DM, Abers MS, Corrales-Medina VF. Acute infection and myocardial infarction. N Engl J Med. 2019 10;380(2):171–6.

12. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, et al. Acute myocardial infarction after laboratory-confirmed influenza infection. N Engl J Med. 2018 Jan 25;378(4):345–53.

13. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. Eur Respir J. 2018;51(3).

14. Goto T, Shimada YJ, Faridi MK, Camargo CA, Hasegawa K. Incidence of acute cardiovascular event after acute exacerbation of COPD. J Gen Intern Med. 2018 Sep;33(9):1461–8.

15. Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. BMJ. 2016 Sep 12;354:i4515.

16. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. Stat Med. 2006 May 30;25(10):1768–97.

17. HCUP-US SID Overview [Internet]. [cited 2019 Apr 17]. Available from: https://www.hcup-us.ahrq.gov/sidoverview.jsp

18. Hasegawa K, Gibo K, Tsugawa Y, Shimada YJ, Camargo CA. Age-related differences in the rate, timing, and diagnosis of 30-day readmissions in hospitalized adults with asthma exacerbation. Chest. 2016 Apr;149(4):1021–9.

19. Hasegawa K, Tsugawa Y, Chang Y, Camargo CA. Risk of an asthma exacerbation after bariatric surgery in adults. J Allergy Clin Immunol. 2015 Aug;136(2):288-294.e8.

20. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. PloS One. 2014;9(3):e92286.

21. Juhn Y, Kung A, Voigt R, Johnson S. Characterisation of children's asthma status by ICD-9 code and criteria-based medical record review. Prim Care Respir J J Gen Pract Airw Group. 2011 Mar;20(1):79–83.

22. Tattersall MC, Guo M, Korcarz CE, Gepner AD, Kaufman JD, Liu KJ, et al. Asthma predicts cardiovascular disease events: the multi-ethnic study of atherosclerosis. Arterioscler Thromb Vasc Biol. 2015 Jun;35(6):1520–5.

23. Strand LB, Tsai MK, Wen CP, Chang S-S, Brumpton BM. Is having asthma associated with an increased risk of dying from cardiovascular disease? A prospective cohort study of 446 346 Taiwanese adults. BMJ Open. 2018 31;8(5):e019992.

24. Xu M, Xu J, Yang X. Asthma and risk of cardiovascular disease or all-cause mortality: a meta-analysis. Ann Saudi Med. 2017;37(2):99–105.

25. Castillo JR, Peters SP, Busse WW. Asthma exacerbations: pathogenesis, prevention, and treatment. J Allergy Clin Immunol Pract. 2017 Aug;5(4):918–27.

26. Singh AM, Busse WW. Asthma exacerbations. 2: aetiology. Thorax. 2006 Sep;61(9):809–16.

27. Yokoyama A, Kohno N, Fujino S, Hamada H, Inoue Y, Fujioka S, et al. Circulating interleukin-6 levels in patients with bronchial asthma. Am J Respir Crit Care Med. 1995 May;151(5):1354–8.

28. Medoff BD, Sauty A, Tager AM, Maclean JA, Smith RN, Mathew A, et al. IFN-gamma-inducible protein 10 (CXCL10) contributes to airway hyperreactivity and airway inflammation in a mouse model of asthma. J Immunol Baltim Md 1950. 2002 May 15;168(10):5278–86.

29. Vijayakumar J, Subramanian S, Singh P, Corsini E, Fontanez S, Lawler M, et al. Arterial inflammation in bronchial asthma. J Nucl Cardiol Off Publ Am Soc Nucl Cardiol. 2013 Jun;20(3):385–95.

30. Qian F-H, Zhang Q, Zhou L-F, Liu H, Huang M, Zhang X-L, et al. High-sensitivity C-reactive protein: a predicative marker in severe asthma. Respirol Carlton Vic. 2008 Sep;13(5):664–9.

31. Fujita M, Ueki S, Ito W, Chiba T, Takeda M, Saito N, et al. C-reactive protein levels in the serum of asthmatic patients. Ann Allergy Asthma Immunol Off Publ Am Coll Allergy Asthma Immunol. 2007 Jul;99(1):48–53.

32. Badimon L, Peña E, Arderiu G, Padró T, Slevin M, Vilahur G, et al. C-reactive protein in atherothrombosis and angiogenesis. Front Immunol. 2018;9:430.

33. Camargo CA, Barr RG, Chen R, Speizer FE. Prospective study of inhaled corticosteroid use, cardiovascular mortality, and all-cause mortality in asthmatic women. Chest. 2008 Sep;134(3):546–51.

34. Hung M-J, Mao C-T, Hung M-Y, Chen T-H. Impact of asthma on the development of coronary vasospastic angina: A population-based cohort study. Medicine (Baltimore). 2015 Oct;94(42):e1880.

35. Salpeter SR, Ormiston TM, Salpeter EE. Cardiovascular effects of beta-agonists in patients with asthma and COPD: a meta-analysis. Chest. 2004 Jun;125(6):2309–21.

36. Olenchock BA, Fonarow GG, Pan W, Hernandez A, Cannon CP, Get with the guidelines steering committee. Current use of beta blockers in patients with reactive airway disease who are hospitalized with acute coronary syndromes. Am J Cardiol. 2009 Feb 1;103(3):295–300.

37. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory therapy with Canakinumab for atherosclerotic disease. N Engl J Med. 2017 21;377(12):1119–31.

38. U.S. Census Bureau QuickFacts: New York; UNITED STATES [Internet]. [cited 2019 Apr 23]. Available from: https://www.census.gov/quickfacts/fact/table/wa,ne,fl,ny,US/PST045218

39. Hasegawa K, Tsugawa Y, Brown DFM, Camargo CA. A population-based study of adults who frequently visit the emergency department for acute asthma. California and Florida, 2009-2010. Ann Am Thorac Soc. 2014 Feb;11(2):158–66.

40. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. Am Heart J. 2004 Jul;148(1):99–104.

41. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. Med Care. 2005 May;43(5):480–5.

42. Tirschwell DL, Longstreth WT. Validating administrative data in stroke research. Stroke. 2002 Oct;33(10):2465–70.

SUMMARY OF PAPER 1 AND PAPER 2 CONCLUSION

In paper 1, based on the data from a multicenter prospective cohort of 1,016 infants with bronchiolitis, we developed four machine learning models to predict severity of illness. By using prehospitalization data as predictors, these models consistently yielded superior performance—a higher AUC, net reclassification, sensitivity, and specificity—in predicting positive pressure ventilation and intensive treatment outcomes over the reference model. Specifically, these advanced machine learning models correctly predicted a larger number of infants with higher severity—with a fewer number of false-positives—who would not be appropriately predicted by the conventional models. Moreover, the machine learning models also achieved a greater net benefit across wide ranges of clinical thresholds. Although an external validation is warranted, the current study lends support to the application of machine learning models to the prediction of acute severity in infants with bronchiolitis. Machine learning models have a potential to enhance clinicians' decision-making ability and hence to improve clinical care and optimize resource utilization in this high morbidity population.

In paper 2, in this self-controlled case series study using large population-based data of adults with asthma, we found that asthma exacerbation was associated with an increased incidence of acute myocardial infarction and ischemic stroke—particularly in the first 1-week period. The observed patterns were consistent across the different patient subgroups and statistical assumptions. For clinicians, our findings provide the opportunities for applying cardiovascular prevention measures to patients with severe asthma exacerbation during hospitalization and transition to outpatient care. Our observations should also facilitate further investigations into the mechanisms linking

asthma exacerbation to acute cardiovascular events, which could, in turn, lead to the development of effective preventive strategies in this large patient population.

OVERALL DISCUSSION AND PERSPECTIVES

In paper 1, we applied machine learning models to predict high acuity care among infants hospitalized for bronchiolitis. By using the prehospitalization data as predictors, these models consistently yielded superior performance in predicting positive pressure ventilation and intensive treatment outcomes over the reference model. In paper 2, we applied causal inference approach using self-controlled case series design and found that asthma exacerbation was associated with an increased incidence of acute myocardial infarction and ischemic stroke.

Further, an integration of machine learning with causal inference approach using multi-omics data is crucial to investigate the complex biological mechanisms of bronchiolitis and asthma. As future directions of research, we aim to delineate the complex pathobiology of bronchiolitis and childhood asthma. While bronchiolitis has been considered a single disease with similar mechanisms, recent studies have shown heterogeneity in clinical presentations, chronic morbidities (e.g., subsequent risk of asthma), and upper airway microbiome, cytokine, and metabolome profiles among infants with bronchiolitis. However, these studies have relied solely on single-level data (e.g., only clinical or microbiome data). To our knowledge, no study has used clinical, virus, and multi-level omics data to investigate the endotypes of bronchiolitis and their longitudinal relation with chronic airway morbidities. The inadequate understanding of bronchiolitis endotypes during infancy—an important period of lung development—has hindered efforts to develop endotype-specific bronchiolitis treatment and asthma prevention strategies in this high-risk population. By using unsupervised machine learning models, we aim to identify biologically- and clinically-meaningful endotypes of bronchiolitis through integrating

clinical, virus, nasopharyngeal microbiome, cytokine and metabolome data. Besides, it is also crucial to apply advanced causal inference and causal discovery approach to infer causal relations between host response, respiratory viruses, and the microbiome. The integration of these two approaches would enable us to develop targeted bronchiolitis treatment and asthma prevention strategies in this high-risk population.