

# Evaluation of Asthma Status Classification Consistency in Electronic Health Records by Asthma Specialist Review versus Rules-Based Algorithms

Kimberly D. Lactaoen<sup>1,2</sup>; Lizbeth F. Gomez<sup>1,2</sup>; Patrick K. Gleeson<sup>3</sup>; Julian Rosé<sup>3</sup>; Latoya Jeanpierre<sup>4</sup>; Samantha Leong<sup>4</sup>; Andrea J. Apter<sup>3</sup>; Rebecca A. Hubbard<sup>5</sup>; Gary E. Weissman<sup>1,2,3</sup>; Blanca E. Himes<sup>6,\*</sup>

<sup>1</sup>Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA; <sup>2</sup>Palliative and Advanced Illness (PAIR) Center; <sup>3</sup>Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pennsylvania, Philadelphia; <sup>4</sup>Division of Allergy and Immunology, Children’s Hospital of Philadelphia; <sup>5</sup>Department of Biostatistics, School of Public Health, Brown University, Providence, RI; <sup>6</sup>National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), Bethesda, MD\*

## Introduction

Electronic Health Record (EHR) data provides extensive information that can be used for research studies, although care must be taken to ensure its biased nature is considered in study design<sup>1</sup>. Previous studies have identified asthma patients according to EHR data using rules-based algorithms and machine learning techniques, with manual chart review often used as the reference standard<sup>2-4</sup>. We sought to determine whether asthma specialists would consistently classify people as having asthma based on EHR-derived data, and whether select rules-based algorithms were consistent with these classifications.

## Methods

- 600 PennMedicine patients with at least one encounter between Jan 1, 2017, and Aug 31, 2023, containing an an International Classification of Diseases, Tenth Revision (ICD-10) code for asthma (i.e., J45\*) were selected as follows: 200 had no record of short-acting beta agonist (SABA) or inhaled corticosteroid (ICS); 200 had record of SABA and no ICS; 200 had record of ICS and no SABA.
- Four asthma specialists (1 pulmonologist, 3 allergists/immunologists) helped create an asthma classification guide with definitions for *Definite/Highly Probable*, *Probable*, *Probably Not/No*, and *Unknown*. A dichotomous classification definition was also used by assigning *asthma* to those with *Definite/Highly Probable* and *Probable* levels.
- Each patient’s EHRs were independently reviewed and labelled for asthma by two specialists. Disagreements were resolved via discussion and consensus when possible, or by including the review of a third specialist when necessary.
- Asthma classifications via manual chart review were compared to common rules-based asthma classification schema.

\*Dr. Himes contributed to this article as an employee of the University of Pennsylvania. The views expressed are her own and do not necessarily represent the views of the National Institutes of Health or the United States Government.

## Results

- Table 1** summarizes patient characteristics. Those with an ICD-10 code for asthma and SABA were more likely to be female, while patients with an ICD-10 for asthma and ICS were more likely to be aged 55y or greater and to be White.
- After initial chart review, 465 of 600 records had consistent classifications (weighted κ-coefficient=0.74). After attempted consensus, 593 of 600 records had consistent classifications (weighted κ-coefficient=0.98). Remaining disagreements were resolved by a third reviewer.
- Using the selection criteria of asthma ICD-10, SABA, and no ICS led to the greatest likelihood of asthma classification among specialists, with 91% having *Definite/Highly Probable* or *Probable* classifications (**Table 2**).
- Three of 12 rules-based algorithms with variations of the presence of ICD-10 codes and medications within the EHR had >90% *Definite/Highly Probable* or *Probable* classifications by specialist chart review (**Table 2**). Of note, the number of people with such classifications was much smaller than the starting sample size of 600.

## Discussion

- Initial chart review showed moderate agreement in classification amongst asthma specialists, underscoring the difficulty in properly classifying asthma based on EHR data alone.
- Comparison of classification results from rules-based algorithms versus asthma specialists shows that care must be taken when using such rules for EHR-based studies of people with asthma.

## Future Directions

Create a machine learning algorithm to predict asthma status according to gold standard four-level classifications based on EHR-derived data, including that extracted via NLP.

| Table 1. Patient characteristics stratified by selection criteria. |                                |                                     |                                    |                                 |
|--|--------------------------------|-------------------------------------|------------------------------------|---------------------------------|
| Demographic characteristic   | Selection Criteria             |                                     |                                    |                                 |
|  | ICD-10<br>N = 200 <sup>‡</sup> | ICD-10/SABA<br>N = 200 <sup>‡</sup> | ICD-10/ICS<br>N = 200 <sup>‡</sup> | Overall<br>N = 600 <sup>‡</sup> |
| Age  |                                |                                     |                                    |                                 |
| 18-34  | 77 (39%)                       | 63 (32%)                            | 36 (18%)                           | 176 (29%)                       |
| 35-54  | 60 (30%)                       | 73 (37%)                            | 54 (27%)                           | 187 (31%)                       |
| 55-74  | 55 (28%)                       | 52 (26%)                            | 81 (41%)                           | 188 (31%)                       |
| 75+  | 8 (4.0%)                       | 12 (6.0%)                           | 29 (15%)                           | 49 (8.2%)                       |
| Sex  |                                |                                     |                                    |                                 |
| Female   | 131 (66%)                      | 146 (73%)                           | 112 (56%)                          | 389 (65%)                       |
| Male   | 69 (35%)                       | 54 (27%)                            | 88 (44%)                           | 211 (35%)                       |
| Unknown/Other  | 0 (0%)                         | 0 (0%)                              | 0 (0%)                             | 0 (0%)                          |
| Race   |                                |                                     |                                    |                                 |
| AIAN   | 0 (0%)                         | 1 (0.5%)                            | 1 (0.5%)                           | 2 (0.3%)                        |
| Asian  | 6 (3.0%)                       | 7 (3.5%)                            | 6 (3.0%)                           | 19 (3.2%)                       |
| Black  | 48 (24%)                       | 59 (30%)                            | 24 (12%)                           | 131 (22%)                       |
| NHPI   | 0 (0%)                         | 1 (0.5%)                            | 0 (0%)                             | 1 (0.2%)                        |
| White  | 116 (58%)                      | 104 (52%)                           | 146 (73%)                          | 366 (61%)                       |
| Unknown/Other  | 30 (15%)                       | 28 (14%)                            | 23 (12%)                           | 81 (14%)                        |
| <sup>‡</sup> n (%)   |                                |                                     |                                    |                                 |

| Table 2. Consistency of asthma specialist versus rules-based algorithm classifications of asthma. Shown are the counts for each asthma specialist classification against selected rules-based algorithms, along with the corresponding percentage. |                                  |          |                 |          |              |
|--|----------------------------------|----------|-----------------|----------|--------------|
| Rules-based Algorithm<br>(Number with asthma according to rule)  | Asthma Specialist Classification |          |                 |          | Dichotomous* |
|  | Definite/Highly Probable         | Probable | Probably Not/No | Unknown  | Has Asthma   |
| J45* ICD-10, no SABA or ICS (N=200)  | 80 (40%)                         | 71 (36%) | 20 (10%)        | 29 (15%) | 151 (76%)    |
| J45* ICD-10 and SABA, no ICS (N=200)   | 136 (68%)                        | 45 (23%) | 8 (4%)          | 11 (6%)  | 181 (91%)    |
| J45* ICD-10 and ICS, no SABA (N=200)   | 89 (45%)                         | 55 (28%) | 13 (7%)         | 43 (22%) | 144 (72%)    |
| ≥ 2 J45* ICD-10 and SABA (N=136)   | 96 (71%)                         | 30 (22%) | 3 (2%)          | 7 (5%)   | 126 (93%)    |
| Active J45* ICD-10 in Problem List and SABA (N=136)  | 101 (74%)                        | 28 (21%) | 1 (1%)          | 6 (4%)   | 129 (95%)    |
| Primary J45* ICD-10 and SABA (N=127)   | 92 (72%)                         | 27 (21%) | 3 (2%)          | 5 (4%)   | 119 (94%)    |
| *The dichotomous classification is based on the summation of <i>Definite/Highly Probable</i> and <i>Probable</i> labels via manual chart review  |                                  |          |                 |          |              |

## References

<sup>1</sup>Hubbard RA, Lou C, Himes BE: The Effective Sample Size of EHR-Derived Cohorts Under Biased Sampling. Modern Statistical Methods for Health Research. Zhao Y, Chen D (eds.). Springer International Publishing, 2021

<sup>2</sup>Greenblatt RE, Zhao EJ, Henrickson SE, Apter AJ, Hubbard RA, Himes BE: Factors Associated with Exacerbations among Adults with Asthma According to Electronic Health Record Data. Asthma Research and Practice 5: 1, 2019.

<sup>3</sup>Nissen F, Quint J, Wilkinson S, Müllerova H, Smeeth L, Douglas IJ. Validation of asthma recording in electronic health records: a systematic review. Clin Epidemiol. 2017 Dec;Volume 9:643–56.

<sup>4</sup>Al Sallakh MA, Vasileiou E, Rodgers SE, Lyons RA, Sheikh A, Davies GA. Defining asthma and assessing asthma outcomes using electronic health record data: a systematic scoping review. Eur Respir J. 2017 Jun;49(6):1700204.

## Funding

This work was supported by NHLBI R01HL162354.

## Contact

kimberly.lactaoen@pennmedicine.upenn.edu

