



Letters

Performance of a computable phenotype for pediatric asthma using the problem list

Electronic health record (EHR) data have become a major tool for clinical research. Electronic health records allow researchers to use computable phenotypes, which use a computerized query of data elements and logical rules to identify a cohort in the EHR. Given the complexity of EHR data, it is important that computable phenotypes be well-validated. The limited computable phenotype literature in asthma has relied on encounter-based criteria, wherein individuals are only included if they had encounters with a diagnosis of asthma or used natural language processing that extracts information from narrative text in the EHR; however, this may not be accessible to some clinical researchers.^{1–3} Maintaining an accurate and updated problem list is now a requirement of many national technological and healthcare management standards.⁴ The problem list is a comprehensive and accessible list of the patient's medical diagnoses and conditions. Incorporating the problem list in the asthma computable phenotype is likely to capture patients with asthma who may not require frequent health care encounters for the management of their diseases. We hypothesize that children with asthma on the problem list but no encounters for an asthma diagnosis have milder disease. In this study, we sought to validate and compare a computable phenotype for asthma in children using encounter and problem list diagnoses for retrospective population-based study.

This study was evaluated by the Duke University Health System Institutional Review Board (No. Pro 00091342) and determined to be exempt under 45 Code of Federal Regulations 46.116.

Clinical EHR data from January 2014 to December 2019 were extracted from the Duke University Health System Epic System software, which serves as the primary health provider in the Durham County. Among children aged 5 to 18 years, we constructed computable phenotype cohorts that consisted of the following: (1) a prescription of 1 or more medications for asthma, and (2) diagnosis of asthma through either encounter or problem list. For the encounter phenotype, children required 1 inpatient or 2 outpatient encounters with an asthma-related *International Classification of Diseases, Ninth/Tenth Revision* code (available online). For the problem list phenotype, children required a problem list entry with an asthma-related *International Classification of Diseases, Ninth/Tenth Revision* code. We abstracted demographic and clinical information for comparison among the phenotypes, including age at inclusion in the cohort, sex, race and ethnicity, insurance status, body mass index, encounters for atopy or asthma exacerbations, and asthma medications (eTables 1 and 2).

Disclosures: The authors have no conflicts of interest to report.

Funding: This work was supported by funding from the National Institutes of Health (grant number R21 HL145415-02). Dr Tang also received support from the National Institutes of Health (grant number T32 AI007062).

<https://doi.org/10.1016/j.anai.2020.07.009>

1081-1206/© 2020 American College of Allergy, Asthma & Immunology. Published by Elsevier Inc. All rights reserved.

The computable phenotypes were combined for validation and 124 children with asthma were randomly selected, stratified by age group (5–7 vs >7 years old) and by phenotype (encounter vs problem list phenotype). Notably, 2 independent physicians reviewed each participant's chart and made a clinical judgment of asthma, no asthma, or need more information. If there was a disagreement, a third pediatric asthma expert (Dr Lang) reviewed the chart to adjudicate clinical asthma status. We calculated the positive predictive value (PPV) of the combined computable phenotype and assessed the percentage agreement among raters.

We identified 6395 children with asthma as follows: 50% were identified with both the encounter and problem list phenotype, 21% were identified with the encounter phenotype alone, and 33% were identified with the problem list phenotype alone (Table 1). Children who met the criteria for the problem list phenotype had, on average, fewer exacerbations requiring systemic corticosteroids and were less likely to have ever been prescribed controller therapy than those who met the criteria for the encounter phenotype. Other demographic characteristics, including ethnicity and insurance status, did not differ by phenotype.

Chart review confirmed the diagnosis of asthma in 121 of the 124 children in the validation cohort, resulting in a PPV of 0.98. Of the 3 children not confirmed to have asthma, 2 were not diagnosed because of a misclassification based on aerodigestive comorbidities or viral infection, and 1 was not diagnosed because of an EHR error (the review revealed no asthma encounters, problem list diagnosis, or medications). Stratification by age group or phenotype did not affect the PPV. The percentage agreement between the initial 2 reviewers was 60%. Chart review revealed that 1 reviewer accounted for most of the discrepancies. Performing a sensitivity analysis without this reviewer increased the reviewer agreement to 76%.

We validated a computable phenotype of asthma using encounter-based and problem list–based criteria in a pediatric asthma population with a PPV of 0.98. The problem list phenotype identified a substantial proportion of children with asthma who did not meet the criteria for the encounter phenotype and who had a comparatively milder disease (fewer asthma exacerbations or need for controller medication).

The operating characteristics of our combined computable phenotype are similar to other asthma computable phenotypes. Systematic reviews of validated asthma computable phenotypes identified a range of PPV from 0.46 to 1.00.^{1,2} Our results indicate that the problem list can be easily incorporated into computable phenotypes of asthma and maintain a comparably high PPV.

Limitations in using EHR data for computable phenotypes include misclassification bias and unmeasured confounding.

Table 1
Characteristics by Phenotype

Characteristics	Encounter phenotype	Problem list phenotype	Both encounter and problem list phenotype
Number of participants, (%)	1348 (21)	2108 (33)	2939 (50)
Age at initiation (y), median (IQR)	10.00 (7.00–13.00)	9.50 (6.00–14.00)	8.00 (6.00–11.00)
Male sex, n (%)	762 (56.5)	1163 (55.2)	1678 (57.1)
Race and ethnicity, n (%)			
Non-Hispanic white	142 (10.5)	259 (12.3)	389 (13.2)
Non-Hispanic black	877 (65.1)	1060 (50.3)	1776 (60.4)
Hispanic	216 (16.0)	545 (25.9)	528 (18.0)
Other/unknown	113 (8.4)	244 (11.6)	246 (8.4)
Insurance status, n (%)			
Private	433 (32.1)	892 (42.9)	944 (32.3)
Public	832 (61.7)	1113 (53.6)	1861 (63.6)
Self-pay/other	83 (6.2)	72 (3.5)	122 (4.2)
Atopic, n (%)	827 (61.4)	1024 (48.6)	2078 (70.7)
Obese, n (%)	333 (25.3)	528 (25.6)	805 (27.7)
Person-time (y), median (IQR)	2.94 (1.39–4.46)	2.02 (0.79–3.53)	3.39 (1.83–4.80)
Encounters/person-time, median (IQR)	4.11 (2.46–6.73)	4.29 (2.53–7.61)	4.93 (3.01–8.07)
Exacerbations/person-time, mean (SD)	1.27 (14.75)	0.27 (3.23)	1.14 (16.66)
Inpatient	0.15 (3.42)	0.00 (0.02)	0.20 (6.76)
ED	0.84 (14.25)	0.05 (1.62)	0.53 (15.08)
Urgent care	0.12 (0.65)	0.07 (0.66)	0.12 (0.99)
Outpatient	0.15 (1.10)	0.15 (2.72)	0.29 (1.94)
Medication, n (%)			
Rescue	1327 (98.4)	2060 (97.7)	2889 (98.3)
ICS or LTRA	938 (69.6)	1254 (59.5)	2429 (82.6)
Combination therapy	208 (15.4)	202 (9.6)	805 (27.4)

Abbreviations: ED, emergency department; ICS, inhaled corticosteroid; IQR, interquartile range; LTRA, leukotriene receptor antagonist.

Because the EHR is used primarily for clinical care and billing purposes and is not explicitly designed for research, it can suffer from missing or incorrect data. Studies have found that approximately 33% of adults and 45% of children with asthma are misdiagnosed.^{5,6} This challenge in diagnosis likely contributes to poor concordance among providers. As discussed in the results, 1 reviewer accounted for most of the discrepancies. Moreover, reviewers noted that they needed more information before making a clinical diagnosis of asthma in 16% (n = 20). The adjudicator ultimately settled discordance among the reviewers, and our validation only found 3 children misclassified as having asthma. Because our aim was to identify children with asthma with high PPV for clinical research, we did not include a control population, therefore we were unable to assess sensitivity. In addition, we recognize that the use of the problem list is not standardized and is utilized differently by providers.⁷ Other researchers may need to evaluate how providers use the problem list in their health care system.

This study validates the inclusion of the problem list in computable phenotypes for pediatric asthma with high PPV. Children identified using this approach in our cohort have milder asthma. Therefore, EHR studies designed to study a comprehensive

cohort representative of asthma in the population should consider incorporating the problem list.

Acknowledgments

The authors thank Anusha Vadlamudi, MD, Olga Hardin, MD, Matthew McCulloch, MD, and Nicole Koutlas, MD, for performing expert validation.

Supplementary Data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.anai.2020.07.009>.

Monica Tang, MD*

Benjamin A. Goldstein, PhD^{†,‡,§}

Jingye He, MA[†]

Jillian H. Hurst, PhD[‡]

Jason E. Lang, MD, MPH^{†,§,||}

*Division of Pulmonary, Critical Care Allergy, and Sleep

University of California San Francisco San Francisco, California

[†]Department of Biostatistics and Bioinformatics Duke University School of Medicine Durham, North Carolina

[‡]Children's Health and Discovery Initiative Department of Pediatrics Duke University School of Medicine Durham, North Carolina

[§]Department of Biostatistics and Bioinformatics Duke Clinical Research Institute Duke University School of Medicine Durham, North Carolina

^{||}Division of Allergy, Immunology, and Pulmonary Medicine Duke Children's Hospital and Health Center Duke University School of Medicine Durham, North Carolina monica.tang2@ucsf.edu

References

1. 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;49(6):1700204.
2. Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas JJ. Validation of asthma recording in electronic health records: a systematic review. *Clin Epidemiol*. 2017;9:643–656.
3. Wi CI, Sohn S, Ali M, et al. Natural language processing for asthma ascertainment in different practice settings. *J Allergy Clin Immunol Pract*. 2018;6(1):126–131.
4. Acker B, Bronnert J, Brown T, et al. Problem list guidance in the EHR. *J AHIMA*. 2011;82(9):52–58.
5. Aaron SD, Vandemheen KL, Boulet LP, et al. Overdiagnosis of asthma in obese and nonobese adults. *CMAJ*. 2008;179(11):1121–1131.
6. Yang CL, Simons E, Foty RG, Subbarao P, To T, Dell SD. Misdiagnosis of asthma in schoolchildren. *Pediatr Pulmonol*. 2017;52(3):293–302.
7. Rothschild AS, Lehmann HP, Hripscak G. Inter-rater agreement in physician-coded problem lists. *AMIA Annu Symp Proc*. 2005;2005:644–648.

Supplementary Data

eTable 1

Computable Phenotype for Exacerbation

Code	ICD-9	ICD-10	Medication
Asthma	493.	J45.	Systemic steroids
Respiratory abnormality, unspecified	786.00	R06.00	○ Prednisone
Shortness of breath (acute respiratory distress)	786.05	R06.02	○ Prednisolone
Tachypnea	786.06	R06.03	○ Dexamethasone
Wheezing	786.07	R06.82	○ Hydrocortisone
Other respiratory abnormalities (other forms of dyspnea, other abnormalities of breathing, unspecified abnormalities of breathing)	786.09	R06.2	○ Methylprednisolone
Cough	786.2	R06.09	
		R06.89	
		R06.9	
		R05	

Abbreviations: ICD-9, *International Classification of Diseases, Ninth Revision*; ICD-10, *International Classification of Diseases, Tenth Revision*.

NOTE: An exacerbation was defined as an encounter with an associated ICD-9/10 code and the use of any of the systemic steroids listed above.

eTable 2

Asthma Medications

Rescue	Albuterol dry powder inhaler (ProAir) Albuterol inhalant solution Albuterol metered dose inhaler (ProAir, Proventil, Ventolin) Albuterol oral tablet Albuterol/ipratropium bromide inhalant solution (DuoNeb) Albuterol/ipratropium bromide metered dose inhaler (Combivent) Ipratropium bromide inhalant solution (Atrovent) Ipratropium bromide metered dose inhaler (Atrovent) Levalbuterol inhalant solution (Xopenex) Levalbuterol metered dose inhaler (Xopenex)
ICS or LTRA	ICS Beclomethasone dipropionate metered dose inhaler (Qvar) Budesonide dry powder inhaler (Pulmicort) Budesonide inhalant solution (Pulmicort) Ciclesonide metered dose inhaler (Alvesco) Fluticasone furoate dry powder inhaler (Arnuity Ellipta) Fluticasone propionate dry powder inhaler (Flovent) Fluticasone propionate metered dose inhaler (Flovent) Mometasone furoate dry powder inhaler (Asmanex) LTRA Montelukast oral tablet (Singulair) Zileuton oral tablet (Zyflo)
Combination therapy	ICS + LTRA (above) ICS + LABA: salmeterol inhalant product (Serevent) ICS + LABA: tiotropium metered dose inhaler (Spiriva) ICS/LABA Budesonide/formoterol fumarate metered dose inhaler (Symbicort) Fluticasone furoate/vilanterol dry powder inhaler (Breco) Fluticasone propionate/salmeterol metered dose inhaler (Advair) Fluticasone propionate/salmeterol dry powder inhaler (Advair) Mometasone fumarate/formoterol furoate metered dose inhaler (Dulera) Biologic Omalizumab (Xolair) Mepolizumab (Nucala)

Abbreviations: ICS, inhaled corticosteroid; LABA, long-acting β -agonist; LAMA, long-acting muscarinic antagonist; LTRA, leukotriene receptor antagonist.