Published in final edited form as:

Ann Allergy Asthma Immunol. 2020 November ; 125(5): 611–613.e1. doi:10.1016/j.anai.2020.07.009.

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

Monica Tang, MD¹, Benjamin A. Goldstein, PhD^{2,3,4}, Jingye He, MA², Jillian H. Hurst, PhD³, Jason E. Lang, MD, MPH^{3,4,5}

¹University of California San Francisco, San Francisco, CA,

²Department of Biostatistics and Bioinformatics, Duke University, Durham, NC,

³Children's Health and Discovery Initiative, Department of Pediatrics, Duke University, Durham, NC,

⁴Duke Clinical Research Institute, Duke University, Durham, NC,

⁵Division of Allergy, Immunology, and Pulmonary Medicine, Duke University School of Medicine, Duke Children's Hospital and Health Center, Durham, North Carolina.

Keywords

asthma; asthma diagnosis; electronic health records; algorithms

Electronic health record (EHR) data have become a major tool for clinical research. EHRs allow researchers to use computable phenotypes, which utilize 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 have had encounter(s) with a diagnosis of asthma, or uses natural language processing which extracts information from narrative text in the EHR but 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 standards⁴. 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 healthcare encounters for management of their disease. We hypothesize that children with asthma on the problem list but no encounters with an asthma diagnosis have milder disease. In this study, we sought to validate

Corresponding Author Information: Monica Tang, 400 Parnassus Ave, Box 0336, San Francisco, CA 94143, Telephone number: 415-514-8044, Fax number: 415-353-3529, monica.tang2@ucsf.edu.

Conflicts of interest: None

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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 (DUHS) Institutional Review Board (No. Pro 00091342) and determined to be exempt under 45 CFR 46.116.

Clinical EHR data from January 2014 to December 2019 were extracted from DUHS EPIC system, which serves as the primary health provider in Durham County. Among children 5–18 years old, we constructed computable phenotype cohorts consisting of 1) a prescription of one or more medications for asthma and 2) diagnosis of asthma through either encounter or problem list. For the encounter phenotype, children required one inpatient or two outpatient encounters with an asthma-related ICD9/10 code (available online). For the problem list phenotype, children required a problem list entry with an asthma-related ICD9/10 code. We abstracted demographic and clinical information for comparison between the phenotypes, including age at inclusion in the cohort, gender, race/ethnicity, insurance status, BMI, encounters for atopy or asthma exacerbations, and asthma medications (e-Supplement).

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

We identified 6395 children with asthma: 50% were identified with both the encounter and problem list phenotype; 21% were identified with the encounter phenotype alone; 33% were identified with the problem list phenotype alone (Table 1). Children who met 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 compared to those who met 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 three children not confirmed to have asthma, two were due to a misclassification based on aerodigestive comorbidities or viral infection and one was due to EHR error (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 two reviewers was 60%. Chart review revealed that one reviewer accounted for a majority of the discrepancies. Performing a sensitivity analysis without this reviewer increased 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 significant proportion of children with asthma who did not meet

criteria for the encounter phenotype and who had 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 0.46–1.00^{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. As 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, one reviewer accounted for a majority 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 between reviewers and our validation only found three children misclassified as having asthma. As our aim was to identify children with asthma with high PPV for clinical research, we did not include a control population, so we were unable to assess sensitivity. Additionally, we recognize that use of the problem list is not standardized and is utilized differently by providers⁷. Other researchers may need to evaluate how providers use the problem list in their healthcare system. This study validates inclusion of the problem list in computable phenotypes for pediatric asthma with high PPV. Children identified using this approach in our cohort were milder. Therefore, EHR studies designed to study a comprehensive cohort representative of asthma in the population should consider incorporating the problem list.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge Anusha Vadlamudi, MD, Olga Hardin, MD, Matthew McCulloch, MD, and Nicole Koutlas, MD for their work in performing the expert validation.

Funding Source: This work was supported by funding from NIH R21 HL145415–02. Monica Tang also received support from NIH T32 AI007062.

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. *The European respiratory journal*. 2017;49(6).
2. Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas IJ. Validation of asthma recording in electronic health records: a systematic review. *Clin Epidemiol*. 2017;9:643–656. [PubMed: 29238227]
3. CI W, S S, M A, et al. Natural Language Processing for Asthma Ascertainment in Different Practice Settings. *The journal of allergy and clinical immunology In practice*. 2018;6(1).
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. [PubMed: 19015563]
6. Yang CL, Simons E, Foty RG, Subbarao P, To T, Dell SD. Misdiagnosis of asthma in schoolchildren. *Pediatric pulmonology*. 2017;52(3):293–302. [PubMed: 27505297]
7. Rothschild AS, Lehmann HP, Hripcsak G. Inter-rater agreement in physician-coded problem lists. *AMIA Annu Symp Proc*. 2005:644–648. [PubMed: 16779119]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:**Characteristics by Phenotype**

	Encounter Phenotype	Problem List Phenotype	Both Encounter and Problem List Phenotype
Number of participants, (%)	1348 (21)	2108 (33)	2939 (50)
Age at Initiation (years), median [IQR]	10.00 [7.00, 13.00]	9.50 [6.00, 14.00]	8.00 [6.00, 11.00]
Male, n (%)	762 (56.5)	1163 (55.2)	1678 (57.1)
Race/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 (years), 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: LTRA, leukotriene receptor antagonist; ICS, inhaled corticosteroid; IQR, interquartile range; sd, standard deviation.