



## A diagnostic codes-based algorithm improves accuracy for identification of childhood asthma in archival data sets

Hee Yun Seol, Chung-Il Wi, Euijung Ryu, Katherine S. King, Rohit D. Divekar & Young J. Juhn

To cite this article: Hee Yun Seol, Chung-Il Wi, Euijung Ryu, Katherine S. King, Rohit D. Divekar & Young J. Juhn (2020): A diagnostic codes-based algorithm improves accuracy for identification of childhood asthma in archival data sets, Journal of Asthma, DOI: [10.1080/02770903.2020.1759624](https://doi.org/10.1080/02770903.2020.1759624)

To link to this article: <https://doi.org/10.1080/02770903.2020.1759624>



Accepted author version posted online: 21 Apr 2020.  
Published online: 20 May 2020.



Submit your article to this journal [↗](#)



Article views: 24



View related articles [↗](#)



View Crossmark data [↗](#)



# A diagnostic codes-based algorithm improves accuracy for identification of childhood asthma in archival data sets

Hee Yun Seol, MD<sup>a,b</sup>, Chung-Il Wi, MD<sup>a</sup> , Euijung Ryu, PhD<sup>c</sup> , Katherine S. King, MS<sup>c</sup>, Rohit D. Divekar, MBBS, PhD<sup>d</sup>, and Young J. Juhn, MD, MPH<sup>e</sup> 

<sup>a</sup>Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, Minnesota, USA; <sup>b</sup>Department of Internal Medicine, Pusan National University Yangsan Hospital, Yangsan, Korea; <sup>c</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA; <sup>d</sup>Division of Allergic Disease, Mayo Clinic, Rochester, Minnesota, USA; <sup>e</sup>Department of Pediatric and Adolescent Medicine/Internal Medicine, Mayo Clinic, Rochester, Minnesota, USA

## ABSTRACT

**Objective:** While a single but truncated ICD code (493) had been widely used for identifying asthma in asthma care and research, it significantly under-identifies asthma. We aimed to develop and validate a diagnostic codes-based algorithm for identifying asthmatics using Predetermined Asthma Criteria (PAC) as the reference.

**Methods:** This is a retrospective cross-sectional study which utilized two different coding systems, the Hospital Adaptation of the International Classification of Diseases, Eighth Revision (H-ICDA) and the International Classification of Diseases, Ninth Revision (ICD-9). The algorithm was developed using two population-based asthma study cohorts, and validated in a validation cohort, a random sample of the 1976–2007 Olmsted County Birth Cohort. Performance of the diagnostic codes-based algorithm for ascertaining asthma status against manual chart review for PAC (gold standard) was assessed by determining both criterion and construct validity.

**Results:** Among eligible 267 subjects of the validation cohort, 50% were male, 70% white, and the median age at last follow-up was 17 (interquartile range, 8.7–24.4) years. Asthma prevalence by PAC through manual chart review was 34%. Sensitivity and specificity of the codes-based algorithm for identifying asthma were 82% and 98% respectively. Associations of asthma-related risk factors with asthma status ascertained by the code-based algorithm were similar to those by the manual review.

**Conclusions:** The diagnostic codes-based algorithm for identifying asthmatics improves accuracy of identification of asthma and can be a useful tool for large scale studies in a setting without automated chart review capabilities.

**Abbreviations:** PAC: Predetermined Asthma Criteria; H-ICDA: The Hospital Adaptation of the International Classification of Diseases, Eighth Revision; ICD-9: The International Classification of Diseases, Ninth Revision; REP: The Rochester Epidemiology Project; NLP-PAC: PAC based, natural language processing; ORs: Odds ratios; ICD-10: International Classification of Diseases, Tenth Revision

## ARTICLE HISTORY

Received 13 January 2020  
Revised 13 April 2020  
Accepted 19 April 2020

## KEYWORDS

Epidemiology; diagnostics; pediatrics

## Introduction

Asthma is currently one of the most common chronic illnesses in children and causes significant morbidity in children and adults, estimated to affect 6.2 million children and 18.4 million adults in the United States (1), and associated with significant resource utilization and burden on health care expenditures (2). Estimates of the prevalence of asthma and asthma-like symptoms vary widely across time and different countries around the world (3–7). However, comparison of these prevalence estimates has been difficult given the

current practice of various definition of asthma (eg 60 different definitions of childhood asthma among 122 published studies) (8,9).

Clinically, asthma is defined by a history of recurrent respiratory symptoms such as cough, wheeze, shortness of breath, and chest tightness that vary over time and in intensity, together with variable and reversible expiratory airflow obstruction (10,11), but it is challenging to incorporate this asthma criteria for ascertaining and identifying subjects with asthma from medical records in large-scale studies. The

emerging artificial intelligence approaches (eg natural language processing) (12,13) as a data mining tool for electronic health records (EHRs) (14) may become useful, but this novel tool may not be applicable to every EHRs system, and a significant amount of paper-based health records still remain in the archives. Therefore, large-scale epidemiologic studies on asthma primarily use various self-reported questionnaires to determine whether subjects have had asthma symptoms, have ever received a physician-diagnosis of asthma or ever used asthma medication. Recognizing the limitations of self-reported asthma status (eg recall bias and sampling error) (15,16), diagnostic codes, typically truncated into a single code (493), have been widely used for large-scale asthma studies, but they may significantly under- or over-identify individuals with asthma (5,17–21). Developing a practical but effective operational definition of asthma using a diagnostic code-based algorithm is useful for asthma care and research. To achieve this goal, we developed and validated a diagnostic codes-based algorithm to identify asthma using the Hospital Adaptation of the International Classification of Diseases, Eighth Revision (H-ICDA) (22) and International Classification of Diseases, Ninth Revision (ICD-9) (23) using the Predetermined Asthma Criteria (24–33) as a reference.

## Methods

### Study setting

This study was conducted in Olmsted County, Minnesota, USA, which has similar population characteristics to the US white population, with the exception that a large proportion of the population is employed in the healthcare field (34,35). Olmsted County provides an ideal location to perform population-based studies such as this under the auspices of the Rochester Epidemiology Project (REP) which has linked and archived the medical records of virtually all persons residing in Olmsted County, Minnesota, for more than half a century. It maintains an electronic index of medical diagnoses and surgical interventions, and has archived all addresses and demographic information over time (34). Over 95% of residents authorize their medical information for research use (36). Therefore, health care is virtually self-contained in our study setting.

### Study design

This is a cross-sectional study to develop and validate of a diagnostic code-based algorithm for identifying

asthma using the Predetermined Asthma Criteria (PAC) as a reference. Medical diagnosis data have been coded using two coding systems in our study setting; the Hospital Adaptation of the International Classification of Diseases, Eighth Revision (H-ICDA) coding system (22) and the International Classification of Diseases, Ninth Revision (ICD-9) coding system (23) during study periods (1976–2015). The Mayo Clinic and Olmsted Medical Center (OMC) which provide medical care in the community had used H-ICDA and ICD-9 (Mayo clinic had used H-ICDA until 2010 and other health care provider [Olmsted Medical Center] used it until 1986). To establish an optimal code algorithm for identifying patients with asthma, we used both coding systems for development, testing, and validation of an algorithm. A validation cohort was used for determining criterion validity and construct validity (see Statistical Analysis). Both Mayo Clinic and OMC IRBs approved this study.

### Study subjects

For the algorithm developing cohorts, we utilized previously used two population-based cohorts ( $n = 3,234$  for developing H-ICDA code-algorithm and  $n = 8,411$  for developing ICD-code algorithm) which defined asthma status by PAC through comprehensive manual chart review (24) and natural language processing algorithm for asthma status (NLP-PAC) (12,13). The developed H-ICDA and ICD-9 codes-based algorithms were further refined on an independent testing cohort ( $n = 2,627$ ) (25,31–33). Finally, we evaluated the validity of the diagnostic codes-based algorithm for identifying asthma against criteria-based manual chart review, using an independent validation cohort, a random sample of 300 children from the 1976–2007 Olmsted County Birth Cohort ( $n = 56,124$ ). The exclusion criteria for validation cohort were (1) absence of research authorization for using medical record, (2) insufficient medical record for determining asthma and (3) any conditions fulfilling the exclusion criteria in PAC (Table 1).

### Predetermined asthma criteria (PAC)

PAC, retrospective criteria for ascertaining asthma developed by renowned asthma researchers, Drs. John Yunginger and Charles Reed (Table 1), has been used extensively in research studying asthma epidemiology (25–33). PAC is conceptually similar to the 2015 Canadian Thoracic and Canadian Pediatric Society asthma criteria consisting of 1) recurrent wheezing

**Table 1.** Predetermined Asthma Criteria (PAC).

Patients were considered to have *definite* asthma if a physician had made a diagnosis of asthma and/or if each of the following three conditions were present, and they were considered to have *probable* asthma if only the first two conditions were present:

1. History of cough with wheezing, and/or dyspnea, OR history of cough and/or dyspnea plus wheezing on examination,
2. Substantial variability in symptoms from time to time or periods of weeks or more when symptoms were absent, and
3. Two or more of the following:
  - Sleep disturbance by nocturnal cough and wheeze
  - Nonsmoker (14 years or older)
  - Nasal polyps
  - Blood eosinophilia higher than 300/uL
  - Positive wheal and flare skin tests OR elevated serum IgE
  - History of hay fever or infantile eczema OR cough, dyspnea, and wheezing regularly on exposure to an antigen
  - Pulmonary function tests showing one FEV<sub>1</sub> or FVC less than 70% predicted and another with at least 20% improvement to an FEV<sub>1</sub> of higher than 70% predicted OR methacholine challenge test showing 20% or greater decrease in FEV<sub>1</sub>
  - Favorable clinical response to bronchodilator

Patients were excluded from the study if any of these conditions were present:

- Pulmonary function tests that showed FEV<sub>1</sub> to be consistently below 50% predicted or diminished diffusion capacity
- Tracheobronchial foreign body at or about the incidence date
- Hypogammaglobulinemia (IgG less than 2.0 mg/mL) or other immunodeficiency disorder
- Wheezing occurring only in response to anesthesia or medications

The following diseases excluded the patient from study if they occurred before the incidence date:

- Bullous emphysema or pulmonary fibrosis on chest radiograph
- PiZZ alpha<sub>1</sub>-antitrypsin
- Cystic fibrosis
- Other major chest disease such as juvenile kyphoscoliosis or bronchiectasis FVC forced vital capacity; FEV<sub>1</sub>, forced expiratory volume in 1 sec.

episodes or airflow obstruction, 2) reversibility to bronchodilator and 3) exclusion of alternative diagnoses (11). Index date was defined as the earliest date when the PAC was met.

### **Development and testing (refining) phase for the diagnostic codes-based algorithm**

As the key components of PAC are physician-diagnosed asthma and recurrent wheezing episodes with other respiratory symptoms within a three year period, conceptually we developed the diagnostic code-based algorithms to reflect these components.

#### **H-ICDA**

Through discussion with general pediatrician (YJJ), pulmonologist (HYS), and allergist (RDD), 255 codes of the H-ICDA related to 'Asthma', 'Bronchiolitis', 'Bronchospasm', 'Wheezing', 'Acute/chronic bronchitis', 'Chronic Obstructive Pulmonary Disease', and 'Pneumonia' were assessed in the development cohort for H-ICDA described above. Based on clinical importance/relevance of each diagnostic code and frequency of use in asthmatics, 50 codes were reviewed and classified to 5 Tiers system (Table 2). Reiterative sensitivity analyses were performed with the various combinations of the Tiers to optimize its performance based on sensitivity and specificity predicting asthma status by PAC. The majority of false positive cases for a diagnosis of asthma was from a single item of Tier 2 ('Bronchitis', 'Bronchospasm', and 'Wheezing').

Therefore we modified the algorithm to either 1) one or more of Tier 1 or 2) two or more of Tier 2 in a three-year period (Table 3).

#### **ICD-9**

For ICD-9, we followed a similar process as used for the H-ICDA (see Tables 2 and 4). Index date of the diagnostic code-based algorithm for identifying asthmatics was defined as the date of first Tier 1 or second Tier 2 episode in a three-year period, whichever came first.

Additional independent testing cohorts were used to test and further refine the developed algorithms (25,31–33). Any discrepancies between the codes-based algorithms and manual chart review were reviewed by two independent reviewers (H.Y.S and C. W.) for adjudicating and the algorithms were refined accordingly.

#### **Validation phase**

Two abstractors (H.Y.S. and E.F.J), blinded to the diagnostic codes, reviewed all of the medical records of a random sample of 300 subjects from the 1976–2007 Population-Based Birth Cohort up to September 1, 2015 to ascertain asthma status by PAC. Both abstractors were in 100% inter-observer agreement on asthma status using ten random samples. Then the newly developed and tested diagnostic codes-based algorithms were applied to the validation cohort.

**Table 2.** Tier system to figure out optimal code-based algorithms, H-ICDA.

| Group                     | Codes that may reflect Wheezing                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                           | Physician-Diagnosed Asthma (Tier 1)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Highly suspicious for Wheezing (Tier 2)                                                                                                                                                                 | Eosinophilic-Hypersensitivity-Allergic (Tier 3)                                                                                                                                                                                                                                                                                                                                                                                       | Obstructive (Tier 4)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Description (H-ICDA Code) | BRONCHITIS, ASTHMATIC, ACUTE (04890114)<br>BRONCHITIS, ASTHMATIC, SUBACUTE (04890116)<br>BRONCHITIS, ASTHMATIC (04900110)<br>BRONCHITIS, CHRONIC, ASTHMATIC (04912111)<br>BRONCHITIS, ALLERGIC (04930120)<br>ASTHMA, ALLERGIC (04930121)<br>HAY FEVER, WITH ASTHMA (04930122)<br>DISEASE, PULMONARY, ALLERGIC (04930123)<br>DISEASE, ALLERGIC LUNG (04930124)<br>DISEASE, REVERSIBLE OBSTRUCTIVE AIRWAY (ROAD) (04930125)<br>ASTHMA, BRONCHIAL, INTRINSIC (04931110)<br>ASTHMA, ENDOGENOUS (04931111)<br>ASTHMA, NOS—ASTH (04939110)<br>ASTHMA, WITH COUGH (04939111)<br>COUGH, ASTHMATIC (04939112)<br>ASTHMA, EXERCISE INDUCED (04939113)<br>ASTHMA, BRONCHIAL (04939120)<br>STATUS, ASTHMATICUS (04939130)<br>ASTHMA, STERIOD DEPENDENT/STERIOD THERAPY (04939140)<br>DISEASE, REACTIVE AIRWAY (LUNG) (04950116) | BRONCHIOLITIS, ACUTE (04891110)<br>BRONCHIOLITIS, CHRONIC (04919111)<br>BRONCHIOLITIS, NOS (04919110)<br>COUGH, BRONCHOSPASM (05197410)<br>BRONCHOSPASTIC (BRONCHIAL) (07793121)<br>WHEEZING (07786110) | ALVEOLITIS (ALLERGIC) (04950110)<br>PNEUMONITIS, HYPERSENSITIVITY (04950111)<br>BRONCHOPNEUMONIA, ALLERGIC (04950112)<br>DISEASE, HYPERSENSITIVITY, LUNG (04950113)<br>DISEASE, LUNG, HYPERSENSITIVITY (04950114)<br>DISEASE, HYPERACTIVE, AIRWAY (LUNG) (04950118)<br>IRRITABLE, BRONCHUS (04900220)<br>PNEUMONITIS, EOSINOPHILIC (05192110)<br>PNEUMONIA, EOSINOPHILIC (05192111)<br>INFILTRATE, PULMONARY, EOSINOPHILIC (05192120) | PNEUMONITIS, OBSTRUCTIVE (04860470)<br>BRONCHITIS, CHRONIC, OBSTRUCTIVE (04912110)<br>OBSTRUCTION, LUNG, WITH BRONCHITIS (04912112)<br>DISEASE, OBSTRUCTIVE, LUNG (04960110)<br>DISEASE, PULMONARY, CHRONIC OBSTRUCTIVE (04960111)<br>OBSTRUCTION, LUNG, CHRONIC (04960112)<br>OBSTRUCTION, AIRWAY, CHRONIC PULMONARY (04960113)<br>DISEASE, LUNG, CHRONIC OBSTRUCTIVE (04960114)<br>OBSTRUCTION, LUNG, AIRWAY, CHRONIC (04960115)<br>COLD, NOS (CHRONIC OBSTRUCTIVE LUNG DISEASE) (04960116)<br>COPD (CHRONIC OBSTRUCTIVE PULMONARY DISEASE) (04960117)<br>DISEASE, OBSTRUCTIVE AIRWAY (04960119)<br>COPD, STERIOD DEPENDENT/STERIOD THERAPY (04960120) |

**Table 3.** Medical Code-based algorithms for identifying people with asthma.

| Either 1) one or more Tier 1 OR 2) two or more of Tier 2 in a three-year period.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                         |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| H-ICDA Code                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                          | ICD-9 Code                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                         |
| Physician-Diagnosed Asthma (Tier 1)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Wheezing (Tier 2)                                                                                                                                                                        | Physician-Diagnosed Asthma (Tier 1)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Wheezing (Tier 2)                                                                                                                                                                                                                                       |
| BRONCHITIS, ASTHMATIC, ACUTE (04890114) BRONCHITIS, ASTHMATIC, SUBACUTE (04890116) BRONCHITIS, ASTHMATIC (04900110) BRONCHITIS, CHRONIC, ASTHMATIC (04912111) BRONCHITIS, ALLERGIC (04930110) ASTHMA, EXTRINSIC (04930120) ASTHMA, ALLERGIC (04930121) HAY FEVER, WITH ASTHMA (04930122) DISEASE, PULMONARY, ALLERGIC (04930123) DISEASE, ALLERGIC LUNG (04930124) DISEASE, REVERSIBLE OBSTRUCTIVE AIRWAY (ROAD) (04930125) ASTHMA, BRONCHIAL, INTRINSIC (04931110) ASTHMA, ENDOGENOUS (04931111) ASTHMA, NOS—ASTH (04939110) ASTHMA, WITH COUGH (04939111) COUGH, ASTHMATIC (04939112) ASTHMA, EXERCISE INDUCED (04939113) ASTHMA, BRONCHIAL (04939120) STATUS, ASTHMATICUS (04939130) ASTHMA, STEROID DEPENDENT/STEROID THERAPY (04939140) DISEASE, REACTIVE AIRWAY (LUNG) (04950116) | BRONCHIOLITIS, ACUTE (04891110) BRONCHIOLITIS, CHRONIC (04919111) BRONCHIOLITIS, NOS (04919110) COUGH, BRONCHOSPASM (05197410) BRONCHOSPASTIC (BRONCHIAL) (07793121) WHEEZING (07786110) | ASTHMA (493) EXTRINSIC ASTHMA (493.0) EXTRINSIC ASTHMA, UNSPECIFIED (493.00) EXTRINSIC ASTHMA WITH STATUS ASTHMATICUS (493.01) EXTRINSIC ASTHMA WITH ACUTE EXACERBATION (493.02) INTRINSIC ASTHMA (493.1) INTRINSIC ASTHMA, UNSPECIFIED (493.10) INTRINSIC ASTHMA WITH STATUS ASTHMATICUS (493.11) INTRINSIC ASTHMA WITH ACUTE EXACERBATION (493.12) CHRONIC OBSTRUCTIVE ASTHMA (WITH OBSTRUCTIVE PULMONARY DISEASE) (493.2) CHRONIC OBSTRUCTIVE ASTHMA, UNSPECIFIED (493.20) CHRONIC OBSTRUCTIVE ASTHMA, WITH STATUS ASTHMATICUS (493.21) CHRONIC OBSTRUCTIVE ASTHMA, WITH ACUTE EXACERBATION (493.22) OTHER FORMS OF ASTHMA (493.8) EXERCISE INDUCED BRONCHOSPASM (493.81) | ACUTE BRONCHIOLITIS (466.1) ACUTE BRONCHIOLITIS DUE TO RESPIRATORY SYNCYTIAL VIRUS (RSV) (466.11) ACUTE BRONCHIOLITIS DUE TO OTHER INFECTIOUS ORGANISMS (466.19) ACUTE BRONCHITIS AND BRONCHIOLITIS (466) ACUTE BRONCHOSPASM (519.11) WHEEZING (786.07) |

### Other variables

To assess construct validity, we collected well-established risk factors associated with asthma such as allergic rhinitis or atopic dermatitis by H-ICDA or ICD-9 and family history of asthma in the subject's medical records during study periods.

### Statistical analysis

Once the code-based algorithms were developed and refined, we assessed the performance of the algorithm for identifying asthma by testing both criterion and construct validity using an independent validation cohort. We first summarized the subject characteristics of the validation cohort. For criterion validity, the performance of the algorithm was assessed by using agreement rate, kappa index, sensitivity, specificity, positive predictive value, and negative predictive value against manual chart review for PAC as a reference. Additionally, we compared the association of the risk factors of asthma with asthma status ascertained by the developed code-based algorithm and manual chart review, respectively using logistic regression models (i.e. Construct validity). Odds ratios (ORs) and their corresponding 95% confidence intervals were

presented. All statistical analyses were performed using JMP statistical software package version 10 (SAS Institute, Inc., Cary, NC).

## Results

### Characteristics of study subjects

Out of 300 study subjects in the validation cohort, we excluded 33 subjects per our exclusion criteria. Among the 267 eligible study subjects, 134 (50%) were males, 185 (70%) white, and the median age at last follow-up date was 17.0 years (interquartile range, 8.7–24.4). Sixty five (24%), 47 (18%), and 45 (17%) had allergic rhinitis, atopic dermatitis, and family history of asthma (respectively) during the study period (Table 5). Ninety subjects (34%) met the PAC by manual chart review, while 78 (29%) met the newly developed diagnostic codes-based algorithm.

### Performance of the developed code-based algorithm (criterion validity)

Overall kappa index and agreement for asthma status between the diagnostic codes-based algorithm and PAC by manual chart review were 0.83 and 93%,

**Table 4.** Tier system to figure out optimal code-based algorithms, ICD-9.

| Group                    | Codes that may reflect Wheezing                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                          | Physician-Diagnosed Asthma (Tier 1)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Specific codes for bronchiolitis, bronchospasm, and wheezing (Tier 2)                                                                                                                                                                                          | Non-specific codes, might suggest bronchiolitis, bronchospasm, and wheezing (Tier 3)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Description (ICD-9 Code) | <p>ASTHMA (493) EXTRINSIC ASTHMA (493.0)</p> <p>EXTRINSIC ASTHMA, UNSPECIFIED (493.00)</p> <p>EXTRINSIC ASTHMA WITH STATUS ASTHMATICUS (493.01) EXTRINSIC ASTHMA WITH ACUTE EXACERBATION (493.02)</p> <p>INTRINSIC ASTHMA (493.1) INTRINSIC ASTHMA, UNSPECIFIED (493.10) INTRINSIC ASTHMA WITH STATUS ASTHMATICUS (493.11) INTRINSIC ASTHMA WITH ACUTE EXACERBATION (493.12) CHRONIC OBSTRUCTIVE ASTHMA (WITH OBSTRUCTIVE PULMONARY DISEASE) (493.2) CHRONIC OBSTRUCTIVE ASTHMA, UNSPECIFIED (493.20) CHRONIC OBSTRUCTIVE ASTHMA, WITH STATUS ASTHMATICUS (493.21) CHRONIC OBSTRUCTIVE ASTHMA, WITH ACUTE EXACERBATION (493.22) OTHER FORMS OF ASTHMA (493.8) EXERCISE INDUCED BRONCHOSPASM (493.81) ASTHMA, UNSPECIFIED (493.9) ASTHMA, UNSPECIFIED STATUS ASTHMATICUS (493.91) ASTHMA, UNSPECIFIED, WITH ACUTE EXACERBATION(493.92)</p> | <p>ACUTE BRONCHIOLITIS (466.1) ACUTE BRONCHIOLITIS DUE TO RESPIRATORY SYNCYTIAL VIRUS (RSV) (466.11) ACUTE BRONCHIOLITIS DUE TO OTHER INFECTIOUS ORGANISMS (466.19) ACUTE BRONCHITIS AND BRONCHIOLITIS (466) ACUTE BRONCHOSPASM (519.11) WHEEZING (786.07)</p> | <p>OTHER DISEASES OF TRACHEA AND BRONCHUS, NOT ELSEWHERE CLASSIFIED (519.1) OTHER DYSPNEA AND RESPIRATORY ABNORMALITY (786.09) OTHER SPECIFIED ALLERGIC ALVEOLITIS AND PNEUMONITIS (495.8) OTHER DISEASES OF THE RESPIRATORY SYSTEM, NOT OTHERWISE CLASSIFIED (519.8) EXTRINSIC ALLERGIC ALVEOLITIS (495) UNSPECIFIED ALLERGIC ALVEOLITIS AND PNEUMONITIS (495.9)</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                | <p>All additional codes which were suggested by REP for bronchiolitis, bronchospasm, and wheezing (Tier 4)</p> <p>OTHER CHRONIC BRONCHITIS (includes chronic bronchiolitis) (491.8) BRONCHITIS AND PNEUMONITIS DUE TO FUMES AND VAPORS (506.0) CHRONIC RESPIRATORY CONDITIONS DUE TO FUMES AND VAPORS (506.4) OTHER SPECIFIED ALVEOLAR AND PARIETOALVEOLAR PNEUMONOPATHIES (includes BOOP) (516.8) STRIDOR (786.1) acute/chronic bronchitis and COPD codes (Tier 5)</p> <p>ACUTE BRONCHITIS (466.0) CHRONIC BRONCHITIS (491) SIMPLE CHRONIC BRONCHITIS (491.0) MUCOPURULENT CHRONIC BRONCHITIS (491.1) OBSTRUCTIVE CHRONIC BRONCHITIS (491.2) OBSTRUCTIVE CHRONIC BRONCHITIS WITHOUT EXACERBATION (491.20) OTHER CHRONIC OBSTRUCTIVE CHRONIC BRONCHITIS WITH ACUTE EXACERBATION (491.21) OBSTRUCTIVE CHRONIC BRONCHITIS WITH ACUTE BRONCHITIS (491.22) BRONCHITIS (491.8) UNSPECIFIED CHRONIC BRONCHITIS (491.9) CHRONIC AIRWAY OBSTRUCTION, NOT ELSEWHERE CLASSIFIED (496)</p> |

**Table 5.** Characteristics of the validation cohort.

| Variables                                                  | (n = 267)        |
|------------------------------------------------------------|------------------|
| Age at last follow-up, years, median (interquartile range) | 17.0 (8.7, 24.4) |
| Male, n (%)                                                | 134 (50)         |
| White, n (%)                                               | 185 (70)         |
| Asthma, ascertained by abstractors based on PAC, n (%)     | 90 (34)          |
| Asthma, ascertained by codes algorithm, n (%)              | 78 (29)          |
| Allergic rhinitis, n (%)                                   | 65 (24)          |
| Atopic dermatitis, n (%)                                   | 47 (18)          |
| Family history of asthma, n (%)                            | 45 (17)          |

**Table 6.** Concordance of asthma status between the codes-based algorithm and PAC by manual chart review (n = 267) (Criterion Validity).

|                    | Unweighted Cohen's Kappa | Overall Agreement | Sensitivity | Specificity   |
|--------------------|--------------------------|-------------------|-------------|---------------|
| Overall (n = 267)  | 0.83                     | 0.93              | 82% (74/90) | 98% (173/177) |
| Male (n = 134)     | 0.83                     | 0.93              | 85% (41/48) | 97% (83/86)   |
| Female (n = 133)   | 0.82                     | 0.92              | 79% (33/42) | 99% (90/91)   |
| White (n = 185)    | 0.83                     | 0.92              | 85% (55/65) | 97% (116/120) |
| Non-White (n = 82) | 0.81                     | 0.93              | 76% (19/25) | 100% (57/57)  |

Sensitivity, specificity, positive predictive value, and negative predictive value were calculated as performance of the diagnostic codes-based algorithm with manual chart review as a reference.

**Table 7.** Association of asthma status of the code based algorithms and PAC by manual chart review with the known risk factors for asthma (n = 267) (Construct Validity).

|                                       | By the diagnostic codes-based algorithm |                    |                |         | By PAC by manual chart review |                    |                |         |
|---------------------------------------|-----------------------------------------|--------------------|----------------|---------|-------------------------------|--------------------|----------------|---------|
|                                       | No Asthma<br>(n = 189)                  | Asthma<br>(n = 78) | OR<br>(95% CI) | p-value | No Asthma<br>(n = 177)        | Asthma<br>(n = 90) | OR<br>(95% CI) | p-value |
| Age*, median<br>(interquartile range) | 14.3 (7.1, 22.7)                        | 21.8 (11.8, 27.0)  | 1.1 (1.0, 1.1) | <0.001  | 14.3 (6.8, 23.6)              | 19.9 (11.9, 25.8)  | 1.0 (1.0, 1.1) | 0.003   |
| Male, n (%)                           | 90 (48%)                                | 44 (56%)           | 1.4 (0.8, 2.4) | 0.191   | 86 (49%)                      | 48 (53%)           | 1.2 (0.7, 2.0) | 0.464   |
| White, n (%)                          | 126 (67%)                               | 59 (76%)           | 1.6 (0.9, 2.8) | 0.148   | 120 (68%)                     | 65 (72%)           | 1.2 (0.7, 2.2) | 0.459   |
| Allergic rhinitis, n (%)              | 33 (17%)                                | 32 (41%)           | 3.3 (1.8, 5.9) | <0.001  | 28 (16%)                      | 37 (41%)           | 3.7 (2.1, 6.7) | <0.001  |
| Atopic dermatitis, n (%)              | 29 (15%)                                | 18 (23%)           | 1.7 (0.9, 3.2) | 0.131   | 25 (14%)                      | 22 (24%)           | 2.0 (1.0, 3.7) | 0.036   |
| Family history of asthma, n (%)       | 22 (12%)                                | 23 (29%)           | 3.2 (1.6, 6.1) | <0.001  | 20 (11%)                      | 25 (28%)           | 3.0 (1.6, 5.8) | <0.001  |

CI = confidence interval; OR = odds ratio; \*Age at last follow-up, years.

respectively, compared to 0.77 and 83% for performance of Tier 1 codes (493 and other codes for asthma diagnosis). Sensitivity and specificity, for the code-based algorithm for asthma status were 82% and 98%, respectively whereas the performance of truncated Tier 1 codes for asthma showed 72% and 99%. These results were similar with regard to gender and race (Table 6).

### **Association of asthma status by the code-based algorithm and PAC with the known risk factors (construct validity)**

Association between selected well-established risk factors for asthma and asthma ascertained by the code-based algorithm was tested and compared with the results based on the PAC-based manual chart review. Subjects who had asthma ascertained by the code-based algorithm had significantly higher odds of having allergic rhinitis (OR: 3.3, 95% CI: 1.8 – 5.9) and a family history of asthma (OR = 3.2, 95% CI: 1.6 – 6.1) compared with those without asthma ( $P < 0.001$

in each). Asthma ascertained by the criteria-based manual chart review showed similar association results for allergic rhinitis and a family history of asthma (Table 7).

### **Time interval of index date between the code based algorithm and PAC based manual chart review**

The median difference of index date for asthma by the two algorithms (the code-based algorithms and PAC based manual chart review) was 0 days (interquartile range, 0–174). Two thirds and 88% of subjects had index dates by algorithm within 6 months and 3 years of index dates by PAC, respectively.

### **Discussion**

In this study, we showed a diagnostic codes-based algorithm significantly improved accuracy of asthma status. To our knowledge, this is the first study that developed and validated a diagnostic code-based

algorithm to identify people with diagnosed asthma or recurrent asthma-like symptoms, which can be used in large-scale studies.

Our study results show that the diagnostic code-based algorithm incorporating codes for asthma symptoms reasonably correlates with criteria based manual chart review, and this concordance is consistent across gender or ethnicity, suggesting criterion validity. The study findings suggest reasonable sensitivity (82%) and specificity (98%) for the algorithm in detecting asthma. Also, the association of risk factors for asthma (e.g. Allergic rhinitis, atopic dermatitis, or family history of asthma) with asthma status determined by the algorithm vs. criteria-based manual chart review was similar, supporting construct validity. For example, prevalence of allergic rhinitis, a risk factor for asthma, was same among asthmatics determined by the algorithm vs. PAC (41%).

Asthma prevalence in our validation cohort (29% by the algorithm and 34% by PAC) was consistent with the earlier report on prevalence of asthma-like symptoms among children based on a large-scale telephone survey in the US and Europe (37). This study assessed the prevalence of asthma or asthma-like symptoms such as wheezing, 32% of young children reported recurrent asthma-like symptoms. Among them, 28% of children reported weekly asthma symptoms and only 20% were diagnosed with asthma. Another potential reason for higher prevalence of asthma in our study setting compared to other studies (e.g. surveillance study for asthma prevalence) is that our study assessed life-time prevalence of asthma based on life-time medical records of the birth cohort, compared to other surveillance data (e.g. National Center for Health Statistics) asking parents of “current asthma” of their child. Diagnostic codes directly related to physician-diagnosed asthma (ie493, Tier 1 in our study) have been widely used for asthma ascertainment in the large-scale studies. All truncated codes under Tier 1 alone (493 and other codes for asthma diagnosis, see Table 3) showed poor sensitivity (72%) compared with those of the algorithm used in this study (82%, Tier 1 OR two or more Tier 2). Sensitivity for a single truncated diagnostic code for asthma (493 of ICD-9) is likely to be even lower, compared to that by the algorithm as Tier 1 in our study include broader codes than 493. It suggests that the proposed algorithm, especially two or more of Tier 2 over time, might identify patients with recurrent asthma symptoms who haven’t yet diagnosed asthma properly and are not identified by a widely used single truncated code for asthma (493). As

pointed out by Bisgaard et al (37). clinicians often use asthma symptoms as a diagnostic term without diagnosing children who have recurrent asthma symptoms for a long period of time which may limit access to preventive and therapeutic interventions for asthma.

Self-reported asthma, another commonly used method for asthma ascertainment in the large-scale study, may be subject to significant misclassification bias (15). Currently, laboratory tests such as eosinophil count, eNO, and lung function tests can be utilized to aid in the diagnosis of asthma, but not suitable for large-scale studies (38,39). Considering these, the developed algorithm might be a useful tool to identify individuals with asthma in large-scale studies. Alternatively, we and others recently reported an innovative artificial intelligence approach such as Natural Language Processing (NLP) algorithm for detecting asthma in large scale studies (12,13,40–44). However, our diagnostic code-based algorithm might be still complementary to NLP algorithms for asthma as many clinics and hospitals are not equipped with resources running NLP algorithms.

There were limitations in our study. First, it was performed at a single study setting. Generalizability of the algorithm is yet to be determined. Another limitation is prescription data was not available for the entire study period (1976–2015) in our clinical setting, so we did not include prescription data in designing the algorithm. International Classification of Diseases, Tenth Revision (ICD-10) has not been included in this study because it has only recently begun to be used in Mayo Clinic and other health care provider in our study setting. While our study showed feasibility and validation of using ICD codes to identify children with asthma more accurately beyond asthma diagnosis codes (egICD-9 of 493) using archival data sets, it may require a validation study to incorporate ICD-10 and ICD-11 for future studies. As there are algorithms for converting ICD codes(45), this should be feasible. The main strength of our study includes epidemiological advantages of our study setting. Since health care environment in our study setting is self-contained within the area and the REP has the ability to capture all inpatient and outpatient medical events for almost all residents of Olmsted County (36). Thus, diagnostic codes which were analyzed in this study capture clinic visits, ED visits, and hospitalizations across the full spectrum of disease, from symptoms through final diagnosis. Another strength is a multi-phase development, testing and validation process based on three relatively large independent cohorts, which improves scientific rigor for validation of the algorithm. Finally,

we used population-based birth cohorts for developing and validating the algorithm which included all medical events from birth of study subjects.

In conclusion, the diagnostic code-based algorithm improves accuracy of identification of asthma. In a setting where automated chart review is not feasible, this newly developed diagnostic code-based algorithm may be a useful tool for asthma ascertainment in asthma care and research.

## Acknowledgments

We thank Mrs. Kelly Okeson for her administrative assistance, Emily F. Johnson for her manual chart review, and Mrs. Julie C. Porcher for her review and helpful comments.

## Disclosure statement

The authors report no conflict of interest. This study was funded by the following: National Institute of Health (NIH)-funded R01 grant (R01 HL126667), R21 grant (R21AI116839-01), and T. Denny Sanford Pediatric Collaborative Research Fund. The resources of the Rochester Epidemiology Project (R01-AG34676) from the National Institute on Aging and CTSA Grant Number UL1 TR000135 from the National Center for Advancing Translational Sciences.

## ORCID

Chung-Il Wi  <http://orcid.org/0000-0001-8938-2997>  
 Euijung Ryu  <http://orcid.org/0000-0001-6281-8738>  
 Young J. Juhn  <http://orcid.org/0000-0003-2112-4240>

## Data availability statement

The datasets generated and/or analyzed during the current study are not publicly available as they include protected health information. Access to data could be discussed per the institutional policy after IRBs at Mayo Clinic and OMC approve it.

## References

1. National Health Interview Survey [Internet]. 2015.
2. Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma (GINA) Program. The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy*. 2004;59(5):469–478. doi:10.1111/j.1398-9995.2004.00526.x.
3. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. *Lancet*. 1998;351(9111):1225–1232.
4. Worldwide variations in the prevalence of asthma symptoms: the International Study of Asthma and Allergies in Childhood (ISAAC). *Eur Respir J*. 1998;12(2):315–335.
5. Lai CK, Beasley R, Crane J, Foliaki S, Shah J, Weiland S, the ISAAC Phase Three Study Group. Global variation in the prevalence and severity of asthma symptoms: phase three of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax*. 2009;64(6):476–483. doi:10.1136/thx.2008.106609.
6. Variations in the prevalence of respiratory symptoms, self-reported asthma attacks, and use of asthma medication in the European Community Respiratory Health Survey (ECRHS). *Eur Respir J*. 1996;9(4):687–695.
7. Sembajwe G, Cifuentes M, Tak SW, Kriebel D, Gore R, Punnett L. National income, self-reported wheezing and asthma diagnosis from the World Health Survey. *Eur Respir J*. 2010;35(2):279–286. doi:10.1183/09031936.00027509.
8. Sá-Sousa A, Jacinto T, Azevedo L, Morais-Almeida M, Robalo-Cordeiro C, Bugalho-Almeida A, Bousquet J, Fonseca J. Operational definitions of asthma in recent epidemiological studies are inconsistent. *Clin Transl Allergy*. 2014;4(1):24. doi:10.1186/2045-7022-4-24.
9. Van Wonderen KE, Van Der Mark LB, Mohrs J, Bindels PJE, Van Aalderen WMC, Ter Riet G. Different definitions in childhood asthma: how dependable is the dependent variable? *Eur Respiratory J*. 2010;36(1):48–56. doi:10.1183/09031936.00154409.
10. (GINA) GIfA. 2018. Global strategy for asthma management and prevention. Available from: <http://www.ginasthma.org/>.
11. Ducharme FM, Dell SD, Radhakrishnan D, Grad RM, Watson WT, Yang CL, Zelman M. Diagnosis and management of asthma in preschoolers: A Canadian Thoracic Society and Canadian Paediatric Society position paper. *Can Respir J*. 2015;22(3):135–143. doi:10.1155/2015/101572.
12. Wi C-I, Sohn S, Rolfes MC, Seabright A, Ryu E, Voge G, Bachman KA, Park MA, Kita H, Croghan IT, et al. Application of a natural language processing algorithm to asthma ascertainment. An automated chart review. *Am J Respir Crit Care Med*. 2017;196(4):430–437. doi:10.1164/rccm.201610-2006OC.
13. Wi C-I, Sohn S, Ali M, Krusemark E, Ryu E, Liu H, Juhn YJ. Natural language processing for asthma ascertainment in different practice settings. *J Allergy Clin Immunol Pract*. 2018;6(1):126–131. doi:10.1016/j.jaip.2017.04.041.
14. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *J Allergy Clin Immunol*. 2020;145(2):463–469. doi:10.1016/j.jaci.2019.12.897.
15. Miller JE. Predictors of asthma in young children: does reporting source affect our conclusions? *Am J Epidemiol*. 2001;154(3):245–250. doi:10.1093/aje/154.3.245.
16. Toren K, Palmqvist M, Lowhagen O, Balder B, Tunsater A. Self-reported asthma was biased in relation to disease severity while reported year of asthma onset was accurate. *J Clin Epidemiol*. 2006;59(1):90–93. doi:10.1016/j.jclinepi.2005.03.019.
17. Perry TT, Vargas PA, McCracken A, Jones SM. Underdiagnosed and uncontrolled asthma: findings in rural schoolchildren from the Delta region of Arkansas. *Ann Allergy Asthma Immunol*. 2008;101(4):375–381. doi:10.1016/S1081-1206(10)60313-4.

18. Montnemery P, Hansson L, Lanke J, Lindholm LH, Nyberg P, Lofdahl CG, et al. Accuracy of a first diagnosis of asthma in primary health care. *Fam Pract.* 2002;19(4):365–368. doi:10.1093/fampra/19.4.365.
19. de Marco R, Cerveri I, Bugiani M, Ferrari M, Verlato G. An undetected burden of asthma in Italy: the relationship between clinical and epidemiological diagnosis of asthma. *Eur. Respir. J.* 1998;11(3):599–605.
20. Looijmans-van den Akker I, van Luijn K, Verheij T. Overdiagnosis of asthma in children in primary care: a retrospective analysis. *Br J Gen Pract.* 2016;66(644):e152–e157. doi:10.3399/bjgp16X683965.
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.* 2010;20(1):79–83. doi:10.4104/pcrj.2010.00076.
22. Activities CoPaH. H-ICDA HAOIneAA. MI: National Center for Health Statistics; 1973.
23. World Health Organization. Manual of the International Classification of Diseases I, and Causes of Death, based on the recommendations of the ninth revision conference 1975, and adopted by the 29th World Health Assembly. Geneva: World Health Organization; 1977.
24. Yunginger JW, Reed CE, O'Connell EJ, Melton LJ, O'Fallon WM, Silverstein MD. A community-based study of the epidemiology of asthma. Incidence rates, 1964–1983. *Am Rev Respir Dis.* 1992;146(4):888–894. doi:10.1164/ajrccm/146.4.888.
25. Sheen YH, Rolfes MC, Wi C-I, Crowson CS, Pendegraft RS, King KS, Ryu E, Juhn YJ. Association of Asthma with Rheumatoid Arthritis: A Population-Based Case-Control Study. *J Allergy Clin Immunol Pract.* 2018;6(1):219–226. doi:10.1016/j.jaip.2017.06.022.
26. Umaretiya PJ, Swanson JB, Kwon HJ, Grose C, Lohse CM, Juhn YJ. Asthma and risk of breakthrough varicella infection in children. *Allergy Asthma Proc.* 2016; 37(3):207–215. doi:10.2500/aap.2016.37.3951.
27. Voge GA, Katusic SK, Qin R, Juhn YJ. Risk of Asthma in Late Preterm Infants: A Propensity Score Approach. *J Allergy Clin Immunol Pract.* 2015;3(6): 905–10. doi:10.1016/j.jaip.2015.03.018.
28. Bang DW, Yang HJ, Ryoo E, Al-Hasan MN, Lahr B, Baddour LM, Yawn BP, Juhn YJ. Asthma and risk of non-respiratory tract infection: a population-based case-control study. *BMJ Open.* 2013;3(10):e003857. doi:10.1136/bmjopen-2013-003857.
29. Juhn YJ, Kita H, Yawn BP, Boyce TG, Yoo KH, McGree ME, Weaver AL, Wollan P, Jacobson RM. Increased risk of serious pneumococcal disease in patients with asthma. *J Allergy Clin Immunol.* 2008; 122(4):719–723. doi:10.1016/j.jaci.2008.07.029.
30. Juhn YJ, Frey D, Li X, Jacobson R. Streptococcus pyogenes upper respiratory infection and atopic conditions other than asthma: a retrospective cohort study. *Prim Care Respir J.* 2012;21(2):153–158. doi:10.4104/pcrj.2011.00110.
31. Kim B-S, Mehra S, Yawn B, Grose C, Tarrell R, Lahr B, Juhn YJ. Increased risk of herpes zoster in children with asthma: a population-based case-control study. *J Pediatr.* 2013;163(3):816–821. doi:10.1016/j.jpeds.2013.03.010.
32. Kwon HJ, Bang DW, Kim EN, Wi C-I, Yawn BP, Wollan PC, Lahr BD, Ryu E, Juhn YJ. Asthma as a risk factor for zoster in adults: A population-based case-control study. *J Allergy Clin Immunol.* 2016; 137(5):1406–1412. doi:10.1016/j.jaci.2015.10.032.
33. Bang DW, Wi C-I, Kim EN, Hagan J, Roger V, Manemann S, Lahr B, Ryu E, Juhn YJ. Asthma status and risk of incident myocardial infarction: A population-based case-control study. *J Allergy Clin Immunol Pract.* 2016;4(5):917–923. doi:10.1016/j.jaip.2016.02.018.
34. Melton LJ. History of the Rochester Epidemiology Project. *Mayo Clin Proc.* 1996;71(3):266–274. doi:10.4065/71.3.266.
35. St. Sauver JL, Grossardt BR, Leibson CL, Yawn BP, Melton LJ, Rocca WA. Generalizability of epidemiological findings and public health decisions: an illustration from the Rochester Epidemiology Project. *Mayo Clin Proc.* 2012;87(2):151–160. doi:10.1016/j.mayocp.2011.11.009.
36. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ, Pankratz JJ, Brue SM, Rocca WA. Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. *Int J Epidemiol.* 2012; 41(6):1614–1624. doi:10.1093/ije/dys195.
37. Bisgaard H, Szefer S. Prevalence of asthma-like symptoms in young children. *Pediatr Pulmonol.* 2007; 42(8):723–728. doi:10.1002/ppul.20644.
38. Roufosse F, Weller PF. Practical approach to the patient with hypereosinophilia. *J Allergy Clin Immunol.* 2010; 126(1):39–44. doi:10.1016/j.jaci.2010.04.011.
39. Smith AD, Cowan JO, Filsell S, McLachlan C, Monti-Sheehan G, Jackson P, Taylor DR. Diagnosing asthma: comparisons between exhaled nitric oxide measurements and conventional tests. *Am J Respir Crit Care Med.* 2004;169(4):473–478. doi:10.1164/rccm.200310-1376OC.
40. Sohn S, Wang Y, Wi CI, Krusemark EA, Ryu E, Ali MH. Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. *J Am Med Inform Assoc.* 2018;25:353–359.
41. Kaur H, Sohn S, Wi C-I, Ryu E, Park MA, Bachman K, Kita H, Croghan I, Castro-Rodriguez JA, Voge GA, et al. Automated chart review utilizing natural language processing algorithm for asthma predictive index. *BMC Pulm Med.* 2018;18(1):34. doi:10.1186/s12890-018-0593-9.
42. Sohn S, Wi CI, Wu ST, Liu H, Ryu E, Krusemark E, Seabright A, Voge GA, Juhn YJ. Ascertainment of asthma prognosis using natural language processing from electronic medical records. *J Allergy Clin Immunol.* 2018;141(6):2292–2294.e3.
43. Sohn S, Wi C-I, Krusemark EA, Liu H, Ryu E, Wu S, Juhn YJ. Assessment of asthma progression determined by natural language processing to improve asthma care and research in the era of electronic medical records. *J Allergy Clin Immunol.* 2017;139:AB100.
44. Afzal Z, Engelkes M, Verhamme KMC, Janssens HM, Sturkenboom MCJM, Kors JA, Schuemie MJ. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. *Pharmacoepidemiol Drug Saf.* 2013;22(8):826–833. doi:10.1002/pds.3438.
45. Center for Medicare and Medicaid Services (CMS). General Equivalence Mappings: ICD-9-CM to and from ICD-10-CM and ICD-10-PCS; 2009.