Early Identification of Childhood Asthma: The Role of Informatics in an Era of Electronic Health Records

Hee Yun Seol[1], Sunghwan Sohn[2], Hongfang Liu[2], Chung-Il Wi[1], Euijung Ryu[2], Miguel A. Park[3] and Young J. Juhn[4]*

[1] Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN, United States, [2] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States, [3] Division of Allergic Diseases, Mayo Clinic, Rochester, MN, United States, [4] Department of Pediatric and Adolescent Medicine and Internal Medicine, Mayo Clinic, Rochester, MN, United States

Emerging literature suggests that delayed identification of childhood asthma results in an increased risk of long-term and various morbidities compared to those with timely diagnosis and intervention, and yet this risk is still overlooked. Even when children and adolescents have a history of recurrent asthma-like symptoms and risk factors embedded in their medical records, this information is sometimes overlooked by clinicians at the point of care. Given the rapid adoption of electronic health record (EHR) systems, early identification of childhood asthma can be achieved utilizing (1) asthma ascertainment criteria leveraging relevant clinical information embedded in EHR and (2) innovative informatics approaches such as natural language processing (NLP) algorithms for asthma ascertainment criteria to enable such a strategy. In this review, we discuss literature relevant to this topic and introduce recently published informatics algorithms (criteria-based NLP) as a potential solution to address the current challenge of early identification of childhood asthma.

Keywords: early, identification, asthma, children, informatics, EHR

INTRODUCTION

Asthma is the most common chronic illness of childhood, affecting up to 17% of children and representing one of most burdensome chronic diseases in the US (1–6). At present, there are no signs of declining trends in the prevalence of asthma among children and adolescents; rather, they continue to increase in many parts of the world (5–7). Furthermore, a delayed diagnosis of asthma, especially in young children is common (8–10). The emerging literature suggests that childhood asthma increases long-term morbidities (11), which could be mitigated by timely therapeutic interventions. This current awareness leads to relevant and consequential questions, (1) what are the various long-term morbidities of childhood asthma that can be reduced through timely identification and therapeutic intervention, (2) what is the magnitude of delay in identifying childhood asthma and why does it occur, (3) what can be done to reduce or eliminate delay in identification of childhood asthma, and (4) what is the role of informatics in clinical care and research for early identification of childhood asthma? Even though currently almost all hospitals and most office-based physicians have adopted electronic health records (EHR) (12, 13), there is yet to evolve a strategic approach for applying informatics tools to identify early childhood asthma. The authors address each of the questions above, specifically the emerging role of informatics in improving early identification of childhood asthma in the EHR era.

Timely Identification and Intervention of Early Onset Asthma Might Reduce the Risk of Long-Term Morbidity

Early onset asthma is no longer just the domain of childhood but also affects long-term morbidity in adulthood. The Tasmanian study showed that children with early-onset asthma or wheezing episodes exhibited a lung function decline by the age of 7 years, compared to healthy children and this gap widened and persisted up to 45 years of age (11). Also, lower lung function at age 7 was associated with chronic obstructive pulmonary disease (COPD) and asthma-COPD overlap syndrome (ACOS). Long-term morbidity of childhood asthma has been corroborated by other studies (14, 15). The Tucson birth cohort study reported that transient wheezers during the first 6 years of life (which has been previously considered a non-pathological entity) demonstrated impaired lung function similar to those with persistent asthma and worse than late-onset wheezers at 16 years of age (16). Control of asthma through interventions (e.g., inhaled corticosteroid [ICS]) improved airway inflammation, respiratory symptoms, asthma exacerbations, rescue medication use, quality of life, and bronchial responsiveness, suggesting that early ICS intervention might have beneficial effects for those at risk of poor long-term asthma outcomes (17–19). In addition, the recent literature suggests that childhood asthma poses numerous health threats through asthma-associated infectious and inflammatory disease comorbidities (AIICs) (20–31). Both children and adults with asthma are at an increased risk of serious respiratory infections [e.g., pneumococcal pneumonia or invasive pneumococcal diseases (20–22), pertussis (23), and common upper respiratory infections (e.g., otitis media and strep infection) (24–26). Furthermore, AIICs are not limited to respiratory infections but also non-respiratory infections such as blood stream infection, appendicitis, herpes zoster (27–30) as well as inflammatory condition including Celiac disease (32). While these AIICs can cause a significant morbidity, the impact of AIICs in children with asthma is currently overlooked. Without a diagnosis of asthma, one has limited access to therapeutic interventions which may mitigate the risk of serious respiratory infections such as an AIIC (e.g., hospitalization-required severe pneumonia) (33). Thus, early identification of childhood asthma is a prerequisite step in the mitigation and prevention of poor long-term asthma outcomes (e.g., exacerbations, poorly controlled asthma and impaired lung function) and AIICs.

A Delay in Identification of Asthma in Children and Adolescents Is Common and Why it Happens

A delay in identification of childhood asthma, especially in young children is still common (8–10). We previously reported that almost two thirds of children under 18 years of age had a delay in asthma diagnosis, and the delay was as long as 3 years after one met Predetermined Asthma Criteria (see **Table 2-1**) (8). Bisgaard et al. revealed that 32% of children, ages 1–5 years in the US and Europe ($n = 9,490$), reported recurrent respiratory symptoms (cough, wheezing, and shortness of breath), and 28% of children reported weekly asthma symptoms of whom only

20% had an asthma diagnosis and only 9.5% were receiving ICS (34). Also, the Lung Health Survey showed that 7.5 % of high school students with recurrent and significant asthma symptoms were not diagnosed with asthma (35). Consequently, a delayed asthma diagnosis was associated with increased urgent care visits suggesting suboptimal care and limited access to proper asthma therapy (10, 36, 37).

Among complex causes for delaying diagnosis of pediatric asthma, both conceptual and operational predicaments can be potentially addressed or improved in the EHR era. Conceptually, there is a lack of consensus on asthma diagnosis. Case in point, the NIH-led workshop discussed the core and supplementary predictors and outcome variables for asthma research (38), but left asthma ascertainment criteria undefined. Currently, only a few asthma ascertainment criteria based on medical record review exist and provide a basis for developing informatics tools for these asthma criteria enabling early detection of childhood asthma as described in the next section. In 2015, the Canadian Thoracic Society and Canadian Pediatric Society criteria for asthma provided succinct and practical approaches to make a diagnosis of asthma for preschoolers (39), consisting of (1) existence of airflow obstruction (e.g., recurrent asthma-like symptoms or exacerbations), (2) reversibility of airflow obstruction (e.g., favorable bronchodilator response), and (3) no evidence for alternative diagnosis(**Table 1**). As an alternative asthma criteria, Predetermined Asthma Criteria (PAC) (**Table 2-1**) which is conceptually similar to the Canadian Thoracic Society criteria, consists of (1) airflow obstruction, (2) reversibility and variability of airflow obstruction (e.g., recurrent wheezing with respiratory symptoms), and (3) no evidence of alternative diagnosis. Its usefulness in detection of asthma has been proven in numerous asthma epidemiology (20, 23–31, 40–47). Also Asthma Predictive Index (API) (**Table 2-2**) which was developed to identify young children at risk of developing asthma can also be considered as feasibility of application of API to

---

TABLE 1 | Operational diagnostic criteria for asthma in children 1–5 years of age, a Canadian Thoracic Society and Canadian Pediatric Society.

1. Documentation of airflow obstruction

Preferred	Documented wheezing and other signs of airflow obstruction by physician or trained health care practitioner
Alternative	Convincing parental report of wheezing or other symptoms of airflow obstruction

2. Documentation of reversibility of airflow obstruction

Preferred	Documented improvement in signs of airflow obstruction to SABA ± oral corticosteroids by physician or trained health care practitioner
Alternative[†]	Convincing parental report of symptomatic response to a 3-month trial of a medium dose of ICS (with as-needed SABA)
Alternative[‡]	Convincing parental report of symptomatic response to SABA

3. No clinical evidence of an alternative diagnosis

[†]*In children with frequent symptoms and/or one or more exacerbation requiring rescue oral corticosteroids or a hospital admission;* [‡]*In children with mild intermittent symptoms and exacerbations, the diagnosis is only suggested because the accuracy of parental report of a short-term response to inhaled short-acting β2-agonists (SABA) may be unreliable due to misperception and spontaneous improvement of another condition. Because this is a weaker alternative diagnostic method, confirmation by direct observation when symptomatic is preferred. ICS Inhaled corticosteroids.*

TABLE 2 | Two asthma ascertainment criteria which were used for developing NLP algorithms.

2-1. 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 (airflow obstruction),
2. Substantial variability in symptoms from time to time or periods of weeks or more when symptoms were absent (reversibility and variability of airflow obstruction)
3. Two or more of the following:

- Favorable clinical response to bronchodilator
- Nonsmoker (14 years or older)
- Sleep disturbance by nocturnal cough and wheeze
- History of hay fever or infantile eczema OR cough, dyspnea, and wheezing regularly on exposure to an antigen
- Blood eosinophilia higher than 300 μL
- Positive wheal and flare skin tests OR elevated serum IgE
- Pulmonary function tests showing one FEV_1 or FVC <70% predicted and another with at least 20% improvement to an FEV_1 of higher than70% predicted OR methacholine challenge test showing 20% or greater decrease in FEV_1
- Nasal polyps

Patients were excluded from our previous study if any of these conditions were present (no evidence of alternative diagnosis):

- Pulmonary function tests that showed FEV_1 to be consistently below 50% predicted or diminished diffusion capacity
- Tracheobronchial foreign body at or about the incidence date
- Hypogammaglobulinemia (IgG <2.0 mg/mL) or other immunodeficiency disorder
- Wheezing occurring only in response to anesthesia or medications
- Bullous emphysema or pulmonary fibrosis on chest radiograph
- PiZZ alpha$_1$-antitrypsin
- Cystic fibrosis
- Bronchopulmonary dysplasia
- Mild pectus excavatum with respiratory symptoms
- Paradoxical vocal cord motion
- Other major chest disease such as juvenile kyphoscoliosis or bronchiectasis
  FVC forced vital capacity; FEV1, forced expiratory volume in 1 s.

2-2.ASTHMA PREDICTIVE INDEX (API)

Major Criteria	Minor Criteria
1. Physician diagnosis of asthma for parents	1. Physician diagnosis of allergic rhinitis for patient
2. Physician diagnosis of eczema for patient	2. Wheezing apart from colds
	3. Eosinophilia ($\geq$ 4%)

Asthma is determined by frequent wheezing episodes (two or more) plus at least one of two major criteria or two of three minor criteria.

retrospective studies has been recently established (48, 49). API includes frequent wheezing in addition to either one major risk factor (parental history of asthma or eczema) or two of three minor risk factors (eosinophilia, wheezing without colds, and allergic rhinitis). Apart from the conceptual challenges, operational challenges also contribute to the delayed diagnosis of

asthma. Even if asthma ascertainment criteria are in place, it is still challenging and extremely labor-intensive for a physician or other abstractor to comprehensively review, collect, and interpret all necessary information from structured (e.g., test results) and unstructured data (e.g., recurrent asthma symptoms or favorable bronchodilator responses written in clinicians' narrative free text) (50). Overcoming these challenges through informatics technologies provides a tremendous opportunity not only to enhance early identification of asthma but also to provide optimal asthma management through a clinical decision support system (CDSS) for clinicians and their care teams. Therefore, clinical informatics approaches, including recently developed natural language processing (NLP) algorithms for PAC and API, to be discussed in the next section, address both conceptual and operational barriers to early identification of asthma in children and adolescents for asthma care and research.

Informatics Approach for Early Identification of Asthma in an EHR Era: Natural Language Processing (NLP)

The amount of EHR data has grown exponentially which provides a tremendous opportunity to leverage EHR data for clinical research and practice. NLP, one of prominent informatics techniques, has been demonstrated to be a promising way to automate chart review enabling large-scale studies that require information embedded in clinical free text (50–53). Although NLP has been successfully applied in various clinical applications (54–56), it has been observed that NLP has been underutilized in EHR-based clinical research (57).

The criteria for asthma ascertainment are mostly based on clinical information such as a history of respiratory symptoms and relevant information which are largely embedded in EHR as free text. Due to the large volume of EHR (primarily free text information), manual chart review to ascertain patient asthma is very costly, time consuming, error-prone, and often impractical for point of care and population-based studies (50). The capability of an NLP algorithm to extract, process, and classify information in free text is a key feature in enabling early identification of asthma in the EHR era (58). However, asthma ascertainment utilizing informatics has not been fully explored. A research team applied a machine learning technique on EHR data (i.e., codes, drugs, and clinical text) in order to identify children with asthma (59). Their approach relies largely on physician-diagnosed asthma and does not take into account the patient's asthma symptoms that could precede the physician's asthma diagnosis. Thus, it is not suitable for early identification of asthma. Also, this approach is not able to provide physicians with evidence of the likelihood of asthma that would assist in their clinical decision making. In order to tackle these challenges and achieve a timely-diagnosis of asthma, we have developed NLP algorithms (both rule-based and machine learning algorithms) for two existing asthma ascertainment criteria, PAC and API as described in the following.

To develop and test NLP algorithms for asthma, the availability of asthma ascertainment criteria based on retrospective EHR data is prerequisite. To our knowledge, two
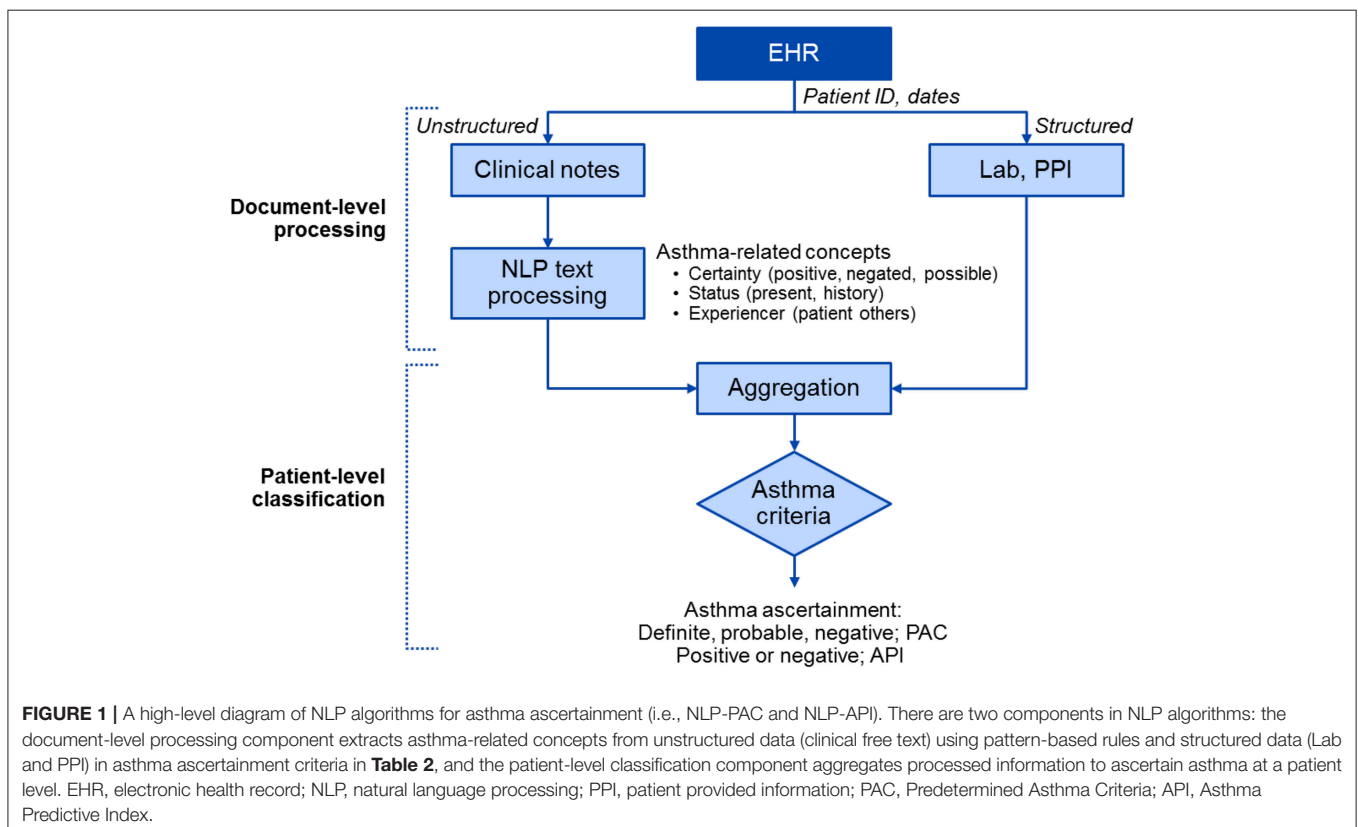
criteria (PAC and API) for asthma ascertainment using existing health records are suitable for developing their corresponding NLP algorithms. Both PAC and API are capable of determining the index date of asthma when the criteria are fulfilled, which is an important feature for clinical and epidemiological studies requiring temporal discernment for asthma as an exposure or outcome.

Conceptually, NLP algorithms extract, process and classify asthma-related events at the note level and then aggregate them to determine the patient asthma based on the given criteria (i.e., definite, probable, or negative of PAC; positive or negative of API) (**Figure 1**) (51, 60). Operationally, we applied an information extraction (IE)-based NLP pipeline for the development of NLP algorithms. For IE, we used various resource-driven tools which utilize domain-specific knowledge engineering such as MedTaggerIE for information in clinical documents (61) and MedTime for temporal information (62). Our original work on developing and testing an initial prototype NLP algorithm for PAC encompassed both rule-based and machine learning systems (50, 63).

We assessed performance of the prototype NLP algorithms for PAC in comparison with ICD code-based asthma ascertainment (50). The results of *the initial prototype* of our NLP algorithm for PAC demonstrated that NLP algorithms significantly outperformed ICD-code based asthma ascertainment in both validity and timely identification of asthma. Importantly, based on the asthma index date, defined as the earliest date of the constellation of symptoms meeting the PAC, 85% of

children had experienced a delayed physician diagnosis by ICD code-based asthma ascertainment, compared to 27% by NLP algorithms. Subsequent to the testing of this initial prototype NLP we have been able to improve the performance. Our recently published study shows the performance of the NLP for PAC algorithm at 90% of PPV and 98% of NPV (51) which suggests a significant improvement over the prototype NLP algorithm. As this enables automated comprehensive chart review for the ascertainment of asthma almost real time and on a large scale, the commentary accompanying the paper considered this informatics work as "a giant step" for leveraging EHR for asthma care and research (64). This is especially true in the early identification of childhood asthma which allows timely intervention allowing mitigation of long-term and serious consequences of childhood asthma as discussed above.

Along these lines, we recently developed and tested an NLP algorithm for the API (60) using the manual chart review to ascertain asthma based on API as a reference. The performance of the NLP algorithm for API was 88% of PPV and 98% of NPV. Moreover, both NLP for PAC and API algorithms provide evidence (i.e., part of clinical text containing asthma-related events) for asthma ascertained by the algorithms, providing verifiable information for decision making of the clinicians. The promising results portend the potential implementation of these NLP algorithms for asthma care in a real-world setting as it greatly reduces the burden on clinicians in reviewing large volumes of EHR to discern asthma status. Currently, these two

FIGURE 1 | A high-level diagram of NLP algorithms for asthma ascertainment (i.e., NLP-PAC and NLP-API). There are two components in NLP algorithms: the document-level processing component extracts asthma-related concepts from unstructured data (clinical free text) using pattern-based rules and structured data (Lab and PPI) in asthma ascertainment criteria in **Table 2**, and the patient-level classification component aggregates processed information to ascertain asthma at a patient level. EHR, electronic health record; NLP, natural language processing; PPI, patient provided information; PAC, Predetermined Asthma Criteria; API, Asthma Predictive Index.

NLP algorithms are being further tested before implementation in our asthma care practice.

To ensure whether these NLP algorithms can be generalizable to other study settings with a different population, different clinical practice, and even different EHR system, we tested the performance of our current NLP algorithm for PAC in a different study setting with a different EHR system and demonstrated portability (PPV: 89% and NPV97% at a hospital in a different state) (65, 66). These results provide a new opportunity for large scale (multi-site) and automated identification of childhood asthma which may potentially address challenges associated with a delayed identification of asthma resulting in delayed therapeutic interventions and possible long-term morbidity.

Implications of NLP Algorithms on Asthma Care and Research for Early Identification of Childhood Asthma

In addressing the existing challenges for early identification of childhood asthma as described above, NLP algorithms have potential impact for asthma care. *First,* incorporating our NLP algorithms for the two asthma ascertainment criteria as part of CDSS would support clinicians or their care teams according to their needs and workflow (e.g., alerting clinicians at a point of care or population surveillance). Also, those NLP algorithms provide retrieved events or evidence with temporal components supporting an asthma diagnosis, so clinicians can then review and discuss such evidence supporting the diagnosis of asthma with parents or caregivers. *Another important implication* of NLP algorithms for PAC and API is the ability for computational phenotyping subgroups of asthmatic children with distinctive clinical characteristics (i.e., stratification) based on the two independent asthma ascertainment criteria enabling precision medicine. For example, we recently reported that children who met both PAC and API criteria by NLP algorithms had clinical features characterizing children at high-risk for conditions such as atopy, poor asthma control, increased asthma exacerbation, impaired lung function, and high risk of asthma-associated infectious and inflammatory comorbid conditions (AIICs), compared to those who met only one criteria or healthy controls (32, 67). This computational phenotyping approach on a large scale is a novel way of identifying a high-risk population for either point-of-care or population surveillance. *Third,* along these lines, extending the application of NLP algorithms from early identification of asthma to the optimal management of asthma through *stratification and prognostication* of asthma is an important step to improve outcomes, care quality and ultimately care costs. Our group has developed and tested the multiple necessary NLP algorithms enabling optimal management of childhood asthma such as an NLP algorithm for asthma prognosis (e.g., likelihood of remission or relapse), triggers of asthma exacerbations (e.g., allergic sensitizations), and clinicians' adherence to guidelines (68–71). Currently, our group is testing the utility of an innovative asthma management strategy leveraging these NLP algorithms implemented in clinical work flow. *Lastly,* computational phenotyping through NLP algorithms is likely to significantly enhance clinical research capabilities for asthma. The algorithms have the potential to enable virtual clinical studies (e.g., clinical trial or observational studies) as they can automatically identify subjects who meet the enrollment criteria and determine outcomes collected prospectively or retrospectively. The conceptual feasibility of emulating clinical trials using EHR has already been demonstrated (72). This might truly revolutionize the current way of conducting clinical studies, especially randomized clinical trials and is an active research area under clinical research informatics.

SUMMARY

In this review, we summarized currently reported informatics approaches for the early identification of asthma and their potential to assist in timely identification so as to reduce the risk of long-term morbidity of asthma through improved management strategies. Given the current magnitude and impact of delayed identification of childhood asthma, we believe that the utilization of informatics approaches and techniques on EHR provides a great potential to improve clinical decision making in the early identification of asthma, allowing proper and timely diagnosis and management of asthma.

To enhance current asthma care and research through the use of rapidly emerging technologies to achieve early identification of asthma, multiple changes are necessary. *First,* awareness of the significance of early identification of childhood asthma is needed from its currently overlooked and misunderstood status, especially long-term impact of transient wheezing episodes and AIICs. *Second*, research effort for the development of new informatics tools, further tested and refined existing tools, and implementation of the clinical informatics tools such as NLP algorithms to the real world setting should be widely shared and supported. *Lastly,* not only the researchers but also clinicians who deliver direct individual patient care need to realize and adapt to the rapid advancement of informatics approaches leveraging EHR to enhance the current state of asthma care and research. Given the potential value and impact of informatics tools in the pursuit of delivering the best possible care to the patient, clinicians and researchers are strongly encouraged to seek a partnership with informatics teams to maximize the benefit of the patients they care for. The informatics approaches and tools for asthma discussed above can be implemented to enhance current asthma care and research, while they may help us make novel biomedical discoveries for asthma. We hope this review paper ignites such interest and endeavor.

AUTHOR CONTRIBUTIONS

HS and SS drafted the manuscript, HL, C-IW, ER, MP, and YJ provided critical input to the manuscript, and all the authors approved the final version.

REFERENCES

1. Valet RS, Gebretsadik T, Carroll KN, Wu P, Dupont WD, Mitchel EF, et al. High asthma prevalence and increased morbidity among rural children in a medicaid cohort. *Ann Allergy Asthma Immunol*. (2011) 106:467–73. doi: 10.1016/j.anai.2011.02.013

2. The Center for Disease Control and Prevention. Vital signs: asthma prevalence, disease characteristics, and self-management education — United States, 2001–2009. *Morb Mortal Wkly Rep*. (2011) 60:547–52.

3. Lethbridge-Çejku M, Vickerie J. Summary health statistics for U.S. adults: national health interview survey, 2003. *Vital Health Stat*. (2005) 10:1–161.

4. Stanton MW. *The High Concentration of U.S. Health Care Expenditures 2006*. Available online at: https://meps.ahrq.gov/data_files/publications/ra19/ra19.pdf

5. Center for Disease Control and Prevention. Forecasted state-specific estimates of self-reported asthma prevalence–United States, 1998. *Morb Mortal Wkly Rep*. (1998) 47:1022–5.

6. Anandan C, Nurmatov U, van Schayck OCP, Sheikh A. Is the prevalence of asthma declining? systematic review of epidemiological studies. *Allergy*. (2010) 65:152–67. doi: 10.1111/j.1398-9995.2009.02244.x

7. Asher MI, Montefort S, Bjorksten B, Lai CK, Strachan DP, Weiland SK, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC phases one and three repeat multicountry cross-sectional surveys. *Lancet*. (2006) 368:733–43. doi: 10.1016/S0140-6736(06)69283-0

8. Molis WE, Bagniewski S, Weaver AL, Jacobson RM, Juhn YJ. Timeliness of diagnosis of asthma in children and its predictors. *Allergy*. (2008) 63:1529–35. doi: 10.1111/j.1398-9995.2008.01749.x

9. Jones A, Sykes A. The effect of symptom presentation on delay in asthma diagnosis in children in a general practice. *Respir Med*. (1990) 84:139–42. doi: 10.1016/S0954-6111(08)80017-0

10. Yeatts K, Davis KJ, Sotir M, Herget C, Shy C. Who gets diagnosed with asthma? frequent wheeze among adolescents with and without a diagnosis of asthma. *Pediatrics*. (2003) 111(5 Pt 1):1046–54. doi: 10.1542/peds.111.5.1046

11. Bui DS, Burgess JA, Lowe AJ, Perret JL, Lodge CJ, Bui M, et al. Childhood lung function predicts adult chronic obstructive pulmonary disease and asthma-chronic obstructive pulmonary disease overlap syndrome. *Am J Respir Crit Care Med*. (2017) 196:39–46. doi: 10.1164/rccm.201606-1272OC

12. Henry JW, Pylypchuk Y, Searcy T, Patel V. *Adoption of Electronic Health Record Systems Among U.S. Non-Federal Acute Care Hospitals: 2008-2015*. The Office of the National Coordinator for Health Information Technology. (2016).

13. Office of the National Coordinator for Health Information Technology. *Office-based Physician Electronic Health Record Adoption,' Health IT Quick-Stat #50*. (2016) Available online at: dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php (accessed December, 2016).

14. Phelan PD, Robertson CF, Olinsky A. The melbourne asthma study: 1964-1999. *J. allergy clin. immunol*. (2002) 109:189–94. doi: 10.1067/mai.2002.120951

15. Sears MR, Greene JM, Willan AR, Wiecek EM, Taylor DR, Flannery EM, et al. A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. *N Engl J Med*. (2003) 349:1414–22. doi: 10.1056/NEJMoa022363

16. Grad R, Morgan WJ. Long-term outcomes of early-onset wheeze and asthma. *J Allergy Clin Immunol*. (2012) 130:299–307. doi: 10.1016/j.jaci.2012.05.022

17. Castro-Rodriguez JA, Rodrigo GJ. Efficacy of inhaled corticosteroids in infants and preschoolers with recurrent wheezing and asthma: a systematic review with meta-analysis. *Pediatrics*. (2009) 123:e519–25. doi: 10.1542/peds.2008-2867

18. Haahtela T, Tamminen K, Kava T, Malmberg LP, Rytila P, Nikander K, et al. Thirteen-year follow-up of early intervention with an inhaled corticosteroid in patients with asthma. *J Allergy Clin Immunol*. (2009) 124:1180–5. doi: 10.1016/j.jaci.2009.09.036

19. Suruki RY, Daugherty JB, Boudiaf N, Albers FC. The frequency of asthma exacerbations and healthcare utilization in patients with asthma from the UK and USA. *BMC Pulm Med*. (2017) 17:74. doi: 10.1186/s12890-017-0409-3

20. Juhn YJ, Kita H, Yawn BP, Boyce TG, Yoo KH, McGree ME, et al. Increased risk of serious pneumococcal disease in patients with asthma. *J Allergy Clin Immunol*. (2008) 122:719–23. doi: 10.1016/j.jaci.2008.07.029

21. Talbot TR, Hartert TV, Mitchel E, Halasa NB, Arbogast PG, Poehling KA, et al. Asthma as a risk factor for invasive pneumococcal disease. *N Engl J Med*. (2005) 352:2082–90. doi: 10.1056/NEJMoa044113

22. Juhn YJ. Risks for infection in patients with asthma (or other atopic conditions): is asthma more than a chronic airway disease? *J Allergy Clin Immunol*. (2014) 134:247–57; quiz 58–9. doi: 10.1016/j.jaci.2014.04.024

23. Capili CR, Hettinger A, Rigelman-Hedberg N, Fink L, Boyce T, Lahr B, et al. Increased risk of pertussis in patients with asthma. *J Allergy Clin Immunol*. (2012) 129:957–63. doi: 10.1016/j.jaci.2011.11.020

24. Bjur KA, Lynch RL, Fenta YA, Yoo KH, Jacobson RM, Li X, et al. Assessment of the association between atopic conditions and tympanostomy tube placement in children. *Allergy Asthma Proc*. (2012) 33:289–96. doi: 10.2500/aap.2012.33.3529

25. Frey D, Jacobson R, Poland G, Li X, Juhn Y. Assessment of the association between pediatric asthma and streptococcus pyogenes upper respiratory infection. *Allergy Asthma Proc*. (2009) 30:540–5. doi: 10.2500/aap.2009.30.3268

26. 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) 21:153–8. doi: 10.4104/pcrj.2011.00110

27. Kim BS, Mehra S, Yawn B, Grose C, Tarrell R, Lahr B, et al. Increased risk of herpes zoster in children with asthma: a population-based case-control study. *J Pediatr*. (2013) 163:816–21. doi: 10.1016/j.jpeds.2013.03.010

28. Kwon HJ, Bang DW, Kim EN, Wi CI, Yawn BP, Wollan PC, et al. Asthma as a risk factor for zoster in adults: a population-based case-control study. *J Allergy Clin Immunol*. (2016) 137:1406–12. doi: 10.1016/j.jaci.2015.10.032

29. Bang DW, Yang HJ, Ryoo E, Al-Hasan MN, Lahr B, Baddour LM, et al. Asthma and risk of non-respiratory tract infection: a population-based case-control study. *BMJ Open*. (2013) 3:e003857. doi: 10.1136/bmjopen-2013-003857

30. Hasassri ME, Jackson ER, Ghawi H, Ryoo E, Wi CI, Bartlett MG, et al. Asthma and risk of appendicitis in children: a population-based case-control study. *Acad Pediatr*. (2017) 17:205–11. doi: 10.1016/j.acap.2016.09.006

31. 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) 37:207–15. doi: 10.2500/aap.2016.37.3951

32. Patel B, Wi CI, Hasassri ME, Divekar R, Absah I, Almallouhi E, et al. Heterogeneity of asthma and the risk of celiac disease in children. *Allergy Asthma Proc*. (2018) 39:51–8. doi: 10.2500/aap.2018.39.4100

33. O'Byrne PM, Pedersen S, Carlsson LG, Radner F, Thoren A, Peterson S, et al. Risks of pneumonia in patients with asthma taking inhaled corticosteroids. *Am J Respir Crit Care Med*. (2011) 183:589–95. doi: 10.1164/rccm.201005-0694OC

34. Bisgaard H, Szefler S. Prevalence of asthma-like symptoms in young children. *Pediatric Pulmonol*. (2007) 42:723–8. doi: 10.1002/ppul.20644

35. Ownby DR, Tingen MS, Havstad S, Waller JL, Johnson CC, Joseph CL. Comparison of asthma prevalence among African American teenage youth attending public high schools in rural Georgia and Urban Detroit. *J Allergy Clin Immunol*. (2015) 136:595–600. e3. doi: 10.1016/j.jaci.2015.02.007

36. Wright AL, Stern DA, Kauffmann F, Martinez FD. Factors influencing gender differences in the diagnosis and treatment of asthma in childhood: the Tucson children's respiratory study. *Pediatr Pulmonol*. (2006) 41:318–25. doi: 10.1002/ppul.20373

37. Lynch BA, Van Norman CA, Jacobson RM, Weaver AL, Juhn YJ. Impact of delay in asthma diagnosis on health care service use. *Allergy Asthma Proc.* (2010) 31:e48–52. doi: 10.2500/aap.2010.31.3358

38. Busse WW, Morgan WJ, Taggart V, Togias A. Asthma outcomes workshop: overview. *J Allergy Clin Immunol.* (2012) 129(3 Suppl.):S1–8. doi: 10.1016/j.jaci.2011.12.985

39. Ducharme F, Dell S, Radhakrishnan D, Grad R, Watson W, Yang C, et al. Diagnosis and management of asthma in preschoolers: a Canadian thoracic society and Canadian paediatric society position paper. *Paediatr Child Health.* (2015) 20:353–71. doi: 10.1093/pch/20.7.353

40. Silverstein MD, Reed CE, O'Connell EJ, Melton LJ, O'Fallon WM, Yunginger JW. Long-term survival of a cohort of community residents with asthma. *N Engl j med.* (1994) 331:1537–41. doi: 10.1056/NEJM199412083312301

41. Hunt LW, Jr., Silverstein MD, Reed CE, O'Connell EJ, O'Fallon WM, Yunginger JW. Accuracy of the death certificate in a population-based study of asthmatic patients. *JAMA.* (1993) 269:1947–52. doi: 10.1001/jama.1993.03500150059027

42. Yunginger J, Reed, CE, O'Connell, EJ, Melton, J, O'Fallon, WM, Silverstein, MD. A community-based study of the epidemiology of asthma: incidence rates, 1964-1983. *Am Rev Respir Dis.* (1992) 146:888–94. doi: 10.1164/ajrccm/146.4.888

43. Yawn BP, Yunginger JW, Wollan PC, Reed CE, Silverstein MD, Harris AG. Allergic rhinitis in Rochester, Minnesota residents with asthma: frequency and impact on health care charges. *J Allergy Clin Immunol.* (1999) 103(1 Pt 1):54–9. doi: 10.1016/S0091-6749(99)70525-7

44. Bauer B, Reed CE, Yunginger JW, Wollan P, Silverstein MD. Incidence and outcomes of asthma in the elderly: a population-based study in Rochester, MN. *Chest.* (1997) 111:303–10. doi: 10.1378/chest.111.2.303

45. Juhn YJ, Qin R, Urm S, Katusic S, Vargas-Chanes D. The influence of neighborhood environment on the incidence of childhood asthma: a propensity score approach. *J Allergy Clin Immunol.* (2010) 125:838–43. e2. doi: 10.1016/j.jaci.2009.12.998

46. Sheen YH, Rolfes MC, Wi CI, Crowson CS, Pendegraft RS, King KS, et al. Association of asthma with rheumatoid arthritis: a population-based case-control study. *J Allergy Clin Immunol Pract.* (2018) 6:219–26. doi: 10.1016/j.jaip.2017.06.022

47. 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) 3:905–10. doi: 10.1016/j.jaip.2015.03.018

48. Wi CI, Krusemark EA, Voge G, Sohn S, Liu H, Ryu E, et al. Usefulness of asthma predictive index in ascertaining asthma status of children using medical records: an explorative study. *Allergy.* (2018) 73:1276–83. doi: 10.1111/all.13403

49. Wi CI, Park MA, Juhn YJ. Development and initial testing of asthma predictive index for a retrospective study: an exploratory study. *J Asthma.* (2015) 52:183–90. doi: 10.3109/02770903.2014.952438

50. Wu ST, Sohn S, Ravikumar KE, Wagholikar K, Jonnalagadda SR, Liu H, et al. Automated chart review for asthma cohort identification using natural language processing: an exploratory study. *Ann Allergy Asthma Immunol.* (2013) 111:364–9. doi: 10.1016/j.anai.2013.07.022

51. Wi CI, Sohn S, Rolfes MC, Seabright A, Ryu E, Voge G, et al. Application of a natural language processing algorithm to asthma ascertainment. an automated chart review. *Am J Respir Crit Care Med.* (2017) 196:430–7. doi: 10.1164/rccm.201610-2006OC

52. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA.* (2011) 306:848–55. doi: 10.1001/jama.2011.1204

53. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc.* (2005) 12:448–57. doi: 10.1197/jamia.M1794

54. Sohn S, Clark C, Halgrim SR, Murphy SP, Chute CG, Liu H. MedXN: an open source medication extraction and normalization tool for clinical text. *J Am Med Inform Assoc.* (2014) 21:858–65. doi: 10.1136/amiajnl-2013-002190

55. Sohn S, Kocher JP, Chute CG, Savova GK. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. *J Am Med Inform Assoc.* (2011) 18(Suppl. 1):i144–9. doi: 10.1136/amiajnl-2011-000351

56. Sohn S, Savova GK, Liu H. Mayo clinic smoking status classification system: extensions and improvements. In: *AMIA Annal Symposium.* San Francisco, California (2009).

57. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. *J Biomed Inform.* (2018) 77:34–49. doi: 10.1016/j.jbi.2017.11.011

58. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc.* (2011) 18:544–51. doi: 10.1136/amiajnl-2011-000464

59. Afzal Z, Engelkes M, Verhamme KM, Janssens HM, Sturkenboom MC, Kors JA, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. *Pharmacoepidemiol Drug Saf.* (2013) 22:826–33. doi: 10.1002/pds.3438

60. Kaur H, Sohn S, Wi CI, Ryu E, Park MA, Bachman K, et al. Automated chart review utilizing natural language processing algorithm for asthma predictive index. *BMC Pulm Med.* (2018) 18:34. doi: 10.1186/s12890-018-0593-9

61. Liu H, Bielinski SJ, Sohn S, Murphy S, Wagholikar KB, Jonnalagadda SR, et al. An information extraction framework for cohort identification using electronic health records. *AMIA Jt Summits Transl Sci Proc.* (2013) 2013:149–53.

62. Sohn S, Wagholikar KB, Li D, Jonnalagadda SR, Tao C, Komandur Elayavilli R, et al. Comprehensive temporal information detection from clinical text: medical events, time, and TLINK identification. *J Am Med Inform Assoc.* (2013) 20:836–42. doi: 10.1136/amiajnl-2013-001622

63. Wu ST, Juhn YJ, Sohn S, Liu H. Patient-level temporal aggregation for text-based asthma status ascertainment. *J Am Med Inform Assoc.* (2014) 21:876–84. doi: 10.1136/amiajnl-2013-002463

64. Beuther DA, Krishnan JA. Finding asthma: building a foundation for care and discovery. *Am J Respir Crit Care Med.* (2017) 196:401–2. doi: 10.1164/rccm.201704-0840ED

65. Wi CI, Sohn S, Ali M, Krusemark E, Ryu E, Liu H, et al. Natural language processing for asthma ascertainment in different practice settings. *J Allergy Clin Immunol Pract.* (2018) 6:126–31. doi: 10.1016/j.jaip.2017.04.041

66. Sohn S, Wang Y, Wi CI, Krusemark EA, Ryu E, Ali MH, et al. Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. *J Am Med Inform Assoc.* (2017) 25:353–9. doi: 10.1093/jamia/ocx138

67. Wi CI, Sohn S, Liu H, Park MA, Kita H, Croghan IT, et al. Assessment of heterogeneity of childhood asthma using medical informatics approaches. *J Allergy Clin Immunol.* (2017) 139:AB202. doi: 10.1016/j.jaci.2016.12.657

68. Sohn S, Wi CI, Wu ST, Liu H, Ryu E, Krusemark E, et al. Ascertainment of asthma prognosis using natural language processing from electronic medical records. *J Allergy Clin Immunol.* (2018) 141:2292–4. e3. doi: 10.1016/j.jaci.2017.12.1003

69. Juhn YMS, Wi C, Fu S, Weston J, Porcher J, Johnson EF, et al. Automated chart review for identifying pre- and peri-natal risk factors associated with childhood asthma. In: *American Thoracic Society 2018 International Conference.* San Diego, CA (2018).

70. Wi CILS, Moon S, Seol HY, Sohn S, Ryu E, Liu H, et al. Automated chart review for identifying factors associated with childhood asthma by utilizing electronic medical records. In: *American Academy of Allergy Asthma & Immunology Annual Meeting.* Orlando, FL (2018).

71. Wi CJY, Moon S, Porcher J, Seol HY, Choi BS, Ryu E, et al. Automated chart review for assessing adherence to asthma guidelines for asthma triggers. In: *European Respiratory Society International Congress.* Paris, France (2018).

72. Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol.* (2018) 96:12–22. doi: 10.1016/j.jclinepi.2017.11.021