Artificial intelligence approaches using natural language processing to advance EHR-based clinical research

Check for updates

Young Juhn, MD, MPH,[a,b] and Hongfang Liu, PhD[c] *Rochester, Minn*

The wide adoption of electronic health record systems in health care generates big real-world data that open new venues to conduct clinical research. As a large amount of valuable clinical information is locked in clinical narratives, natural language processing techniques as an artificial intelligence approach have been leveraged to extract information from clinical narratives in electronic health records. This capability of natural language processing potentially enables automated chart review for identifying patients with distinctive clinical characteristics in clinical care and reduces methodological heterogeneity in defining phenotype, obscuring biological heterogeneity in research concerning allergy, asthma, and immunology. This brief review discusses the current literature on the secondary use of electronic health record data for clinical research concerning allergy, asthma, and immunology and highlights the potential, challenges, and implications of natural language processing techniques. (J Allergy Clin Immunol 2020;145:463-9.)

Key words: *EHRs, asthma, allergy, immunology, informatics, data mining, machine learning, natural language processing, algorithms, artificial intelligence*

Over the past decade, electronic health record (EHR) systems have been increasingly implemented at US hospitals and clinics. Thus, the role of artificial intelligence (AI) (computer-handled human task) as a tool for enhancing clinical care and research is becoming important in a broad range of clinical practices. For example, a few recent AI works in the field of respiratory disease and other clinical areas highlight feasibility of interpretation of pulmonary function test results and their associated diagnosis, early identification of asymptomatic left ventricular dysfunction from 12-lead electrocardiogram, and early detection of atypia or carcinoma *in situ* from breast tissue biopsy.[1-3] These works may reflect only a small aspect of opportunities that EHR-based informatics research can offer in the future. Specifically, large amounts of detailed longitudinal patient information, including clinical history, laboratory tests, medications, interventions, and prognoses, are accumulated, updated, and available electronically. These large clinical databases are valuable data sources for clinical and translational research. Major initiatives have been established to exploit this crucial resource, including the Clinical and Translational Science Awards Program,[4] the Electronic Medical Records and Genomics Network, which links DNA biorepositories with EHRs,[5] the Clinical Data Research Networks Program supported by the Patient-Centered Outcomes Research Institute for comparative effectiveness research,[6] and the Observational Health Data Sciences and Informatics initiative (http://ohdsi.org/), which now contains a data network with more than 660 million subjects, all of which are normalized to the Observational Medical Outcomes Partnership Common Data Model. The great potential of these large initiatives highly depends on efficient and effective data mining approaches leading to knowledge discovery and enhancing clinical practice,[7,8] as a large amount of clinical information is locked in clinical narratives. For example, clinicians spend a significant amount of their time documenting clinical notes such as history of present illness, clinical history, narrative description on physical examination findings, pulmonary function test reports, and radiology or operative reports. Now, an online patient portal is an important platform for communication between clinicians and patients largely in free text and a valuable source for obtaining patient-reported outcomes. Experts from International Data Corporation have estimated that unstructured data accounts for more than 80% of currently available health care data.[9] Regarding this, natural language processing (NLP) techniques have been leveraged to extract information from clinical narratives for clinical research. Now, this capability of NLP systems enables AI-assisted cohort selection and/or matching from EHRs, which may make clinical

Abbreviations used
AI: Artificial intelligence
EHR: Electronic health record
ML: Machine learning
NLP: Natural language processing
PAC: Predetermined asthma criteria
PPV: Positive predictive value

From [a]the Precision Population Science Lab, Division of Community Pediatric and Adolescent Medicine, Department of Pediatric and Adolescent Medicine and [b]the Division of Allergy, Department of Medicine, Mayo Clinic and [c]the Division of Digital Health, Department of Health Sciences Research, Mayo Clinic, Rochester.

The work was supported by grants from the National Institute of Health (grant no. R01 HL126667) and the R21 grant (grant no. R21AI116839-01).

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Received for publication October 31, 2019; revised December 18, 2019; accepted for publication December 19, 2019.

Available online December 26, 2019.

Corresponding author: Young J. Juhn, MD, MPH, Precision Population Science Lab, Department of Pediatric and Adolescent Medicine/Internal Medicine/Health Sciences Research, Mayo Clinic, 200 First St SW, Rochester, MN 55905. E-mail: juhn.young@mayo.edu.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections

0091-6749/$36.00

© 2019 American Academy of Allergy, Asthma & Immunology

https://doi.org/10.1016/j.jaci.2019.12.897

trials or studies virtual or, at least, streamlined.[10-13] For example, the 2018 National NLP Clinical Challenges (n2c2) Shared-Task and Workshop on Cohort Selection for Clinical Trials was organized for competition. One group developed a rule-based NLP system for 13 selection criteria for a clinical trial using 202 training cohort and 86 testing cohort and achieved an F-measure of 0.90 (weighted mean of a test's precision [positive predictive value (PPV)] and recall [sensitivity]).[13] Given the labor-intensive nature of selecting a proper clinical trial cohort meeting the enrollment and exclusion criteria, this capability of an NLP system may potentially enable a virtual clinical trial and observational studies apart from enhancing our abilities for clinical trials. Also, because EHR-based research as a source of real-world evidence is complementary to traditional randomized controlled trial–based evidence as recently recognized by the Food and Drug Administration,[14] the utility of NLP as an efficient and effective approach leveraging EHRs for real-world evidence is likely to be increasingly recognized. In this review, first we will focus on discussing the current literature on NLP-based research in allergy, asthma, and immunology on the basis of our systematic literature review. We refer readers to the recent review on broad information extraction approaches (beyond allergy, asthma, and immunology) from EHRs for further understanding.[15,16] In the second part, we will introduce NLP to clinicians and highlight the role of NLP in data mining and knowledge discovery from EHRs using specific case studies. We will conclude this review with a discussion of the implications of NLP-based clinical research and care in allergy, asthma, and immunology.

SYSTEMATIC REVIEW FOR NLP-BASED RESEARCH IN ALLERGY, ASTHMA, AND IMMUNOLOGY

To assess the current state of EHR-based research using NLP in the field of asthma, allergy, and immunology, a comprehensive literature search of several databases from January 1, 2000, to August 13, 2019, English language, was conducted (see Table E1 in this article's Online Repository at www.jacionline.org for the detailed method for our systematic literature search and the summary of results of each study). Only 21 articles were included in this systematic review by excluding abstracts, review articles, research protocols, unavailable full articles, non–EHR- or non–NLP-based studies, and nonmatching topics. Overall, NLP-based research in the field of asthma, allergy, and immunology is limited in the literature. Herein, we summarize the current state of NLP-based research leveraging EHRs for allergic disorders and discuss the key findings on the limitations of NLP-based research from the literature review.

Asthma

NLP is most commonly applied for determining asthma status or outcomes as shown in Table E1. There are a total of 13 publications that used NLP logic (computer follows the preset rules in the program) or NLP machine learning (ML) (computer learns latent rules through human guidance called annotation [supervised] or no human guidance [unsupervised]) to determine asthma status or outcomes. Apart from 9 reports from our research group,[17-25] there are only 4 publications that used NLP algorithms in asthma status or outcomes.[26-28] As shown in Table E1, overall, the performance of NLP systems for asthma status and outcomes appears to vary depending on the nature of asthma concepts, outcomes, and NLP methods. Our previous work primarily applied NLP (logic or ML) to detect asthma status or prognosis on the basis of predetermined criteria instead of searching free-text terms as discussed in the following section.

Atopic dermatitis

Only 1 article used NLP to define atopic dermatitis on the basis of a predetermined criteria.[29] In this study, the authors developed an ML algorithm for identifying atopic dermatitis in adults from EHRs using Hanifin-Rajka and UK Working Party criteria. The study used 3 nonoverlapping data subsets stratified by class label: training (60%), validation (20%), testing (20%) of 562 adult patients. They reported precision (69%-84%), recall (51%-75%), and F (0.62-0.79) depending on structured code only, NLP only, and structured code-NLP and type of atopic dermatitis criteria.

Allergy entry in EHRs

Another area to which NLP has been applied is identification of allergy entry in EHRs. Even if standard terminologies for allergy entry could be searched and retrieved, such terminologies exist in an inconsistent or nonstandardized format (ie, lexical variation), posing challenges for identifying specific allergy status from EHRs For example, Systematized Nomenclature of Medicine—Clinical Terms (standardized clinical terminology system developed and maintained by SNOMED International, a nonprofit international organization, for the purpose of information entry into EHRs [*International Classification of Diseases* (*ICD*) is a classification system for information retrieval]) was used to match terms referring to allergy entry in EHRs but matching tends to be low (eg, 1 study reported 82% for SNOWMED-CT) because 6% of all allergies and 48% of food allergies were in a free-text format.[30] Because of this reason, NLP had to be used for addressing this challenge (second most common use cases for NLP after asthma), detection of specific allergy status from EHRs. There are 4 NLP-based studies (3 studies based on NLP logic for food allergy,[30] any allergy entries in emergency department setting,[31] and food/drug allergy[32] and 1 study based on supervised NLP-ML for any allergy misspelling[33]). Their overall performance of NLP algorithms was excellent (F-measures [mean of PPV and sensitivity] of 94%-99%[30,32,33] or slightly lower F-measures of 80%-92% depending on allergies[31]). These are promising results for application to clinical practice.

Other areas

There is 1 article for detecting smoking status related to asthma exacerbation in EHRs on the basis of NLP logic, with PPV of 79% and specificity of 90%.[34] One article used NLP and multiple ML classifiers to extract and identify anaphylaxis incidence from unstructured data (F-measure, 048-0.95).[35] Another article used NLP to extract and determine the incidence of hypersensitivity reaction to nonsteroidal anti-inflammatory drug from EHRs.[36]

There are several key findings for the current state of NLP-based research in the field of asthma, allergy, and immunology.

First, overall, NLP is underused in the field of asthma, allergy, and immunology. Although there is some NLP-based research in allergy and asthma, little research based on NLP approaches has been done in atopic dermatitis or allergic rhinitis. Apart from NLP

algorithms for asthma criteria and prognosis reported by our group discussed later,[20-23,37] NLP algorithms for complex concepts such as predetermined criteria for allergic disorders are significantly limited, recognizing that valuable information required for complex predetermined criteria for allergic disorders exists in a free-text format of EHRs. NLP is a useful tool for identifying a cohort with distinctive clinical characteristics in clinical conditions with significant heterogeneity such as asthma. For example, because NLP has a technical feature that researchers may modify the existing NLP algorithms for a condition (eg, asthma) to identify a subgroup of patients with distinctive clinical characteristics (eg, atopic asthma or asthma with impaired lung function), which may not be feasible with *ICD* codes for asthma.

Second, the current literature on EHR-based research using NLP has a tendency to not capture and use a temporal component for outcome events of interest. Because NLP leverages free texts of clinicians' notes in EHRs, it is more accurate than structured code for detecting the index date of allergic disorders. Because the status of allergic disorders in large-scale clinical studies is frequently determined by a survey question about a physician diagnosis of asthma (eg, "have you ever been told that you (or your child) have a diagnosis of asthma?") without knowing the index date, NLP is useful for detecting the index date (when one fulfilled predetermined criteria as shown in our research work discussed later) of asthma or other outcomes such as remission on a large scale. For clinical care and research, it is crucially important to determine the index or incidence date of outcome events (eg, asthma onset and asthma remission or relapse) because temporality is an essential aspect to discern the relationship between predictors and outcomes. In this respect, NLP systems have the capability for capturing temporal components of outcome events over time because NLP programs detect temporal components for each note for a specific concept.

Third, as one of the major challenges in informatics research, portability and external validation of AI algorithms are often difficult to assess because of patient confidentiality and technology transfer, only a few studies have addressed this challenge. Clinical practice and workflow vary across institutions, which results in different practice settings and reporting schemes for generating EHRs. Also, it has been demonstrated that clinical language is not homogeneous, but instead consists of heterogeneous characteristics (eg, syntactic variation or semantic variation).[25] All these factors pose challenges for applying an NLP algorithm developed from one institution to another and need to be considered in portability of NLP algorithms. This is an active research area for addressing these challenges.

Finally, not least, apart from its accuracy of EHR-based AI tools, another important aspect for assessing properness or utility of such tools is acceptability, user interface, usability, or workflow for the entire care process to ensure EHR-based AI tools are helpful rather than a task or burden to clinicians or researchers. Dr Homer Warner, a pioneer in medical informatics, said "Medical informatics = 10% technology + 10% medicine + 80% sociology," highlighting the crucial importance of designing and implementing informatics tools including NLP considering patient-centered, clinician-empowering, and team-based workflow. In this respect, there is no widely accepted standard metrics assessing or reporting these important domains. We further discuss this important issue and challenge in the later section.

THE CASE FOR NLP SYSTEMS AS AN EHR-BASED CLINICAL RESEARCH TOOL

Introduction for NLP

We discuss a brief overview of NLP to introduce the conceptual understanding of NLP for clinicians but do not intend to cover the topic of NLP for the purpose of performing NLP-based research. We refer readers to a few review articles that discuss NLP in depth.[38-40] NLP is a field in computer science, AI, and computational linguistics that enables interactions between computers and human languages and bridges the gap between clinical human language and computational systems. NLP, ML, and deep learning (a field of ML) all are part of the field of AI, and NLP uses ML and deep learning. NLP is computer-handled communication with humans through natural languages (human languages, not computer languages). NLP can be broadly defined as any tasks that computationally uses human languages such as written language (texts) or spoken language (speech) to detect the underlying concepts. In the context of clinical research, we limit it to EHRs in the clinical setting, which only encompass protected health information of patients in digital format created by patients, clinicians, and a broad range of other care team members. There are 2 approaches for NLP: one is being operated by rule-based (computer follows the preset rules in the program) versus ML approaches (computer learns latent rules through human guidance called annotation [supervised] or no human guidance [unsupervised]). Experts from International Data Corporation have estimated that unstructured data accounts for more than 80% of currently available health care data.[9] Although valuable clinical information (eg, history of present illness, clinical history, narrative description on physical examination findings, pulmonary function test report, and radiology or operation reports) is embedded in free texts of EHRs, it is labor intensive to extract such information and enter them into structured data. Thus, most, if not all, large-scale studies or population management strategies heavily rely on structured data and do not tap into unstructured data in EHRs, in a way preventing clinicians and researchers from complete and nuanced information in EHRs. There has been a concern about underutilization of unstructured data in EHRs, leading to persistent use of inaccurate structured data (eg, *ICD* codes) in both clinical research and care.[37,41] NLP can potentially address this challenge.

As shown in Fig 1 conceptually, NLP algorithms for cohort identification *extract* the information (or concepts) from EHRs, then *process* the extracted information, and finally *classify* patients into a subgroup according to rules or learners. These conceptual procedures for NLP are rather complex because NLP is not a single technique; rather, it consists of multiple techniques grouped together. To map words or phrases to concepts of interest, it requires careful text preprocessing and other NLP tasks to convert a document from its natural form into a bag of words.[38,39] A few examples of low-level NLP tasks (text preprocessing) include the following: (1) sentence boundary detection, which is typically defined by a period (title such as "Dr." might complicate this task), (2) tokenization (breaking a sentence into individual tokens) (eg, converting a sentence "patients have multiple history of wheezing episodes" into "patients," "have," "multiple," "history," "wheezing," "episodes"), (3) stemming (reducing word into a root form) (eg, converting "starting, started, and starts" to "start") and (4) lemmatization (process mapping a token) (eg, mapping "RBC, red blood cells" to "erythrocyte").

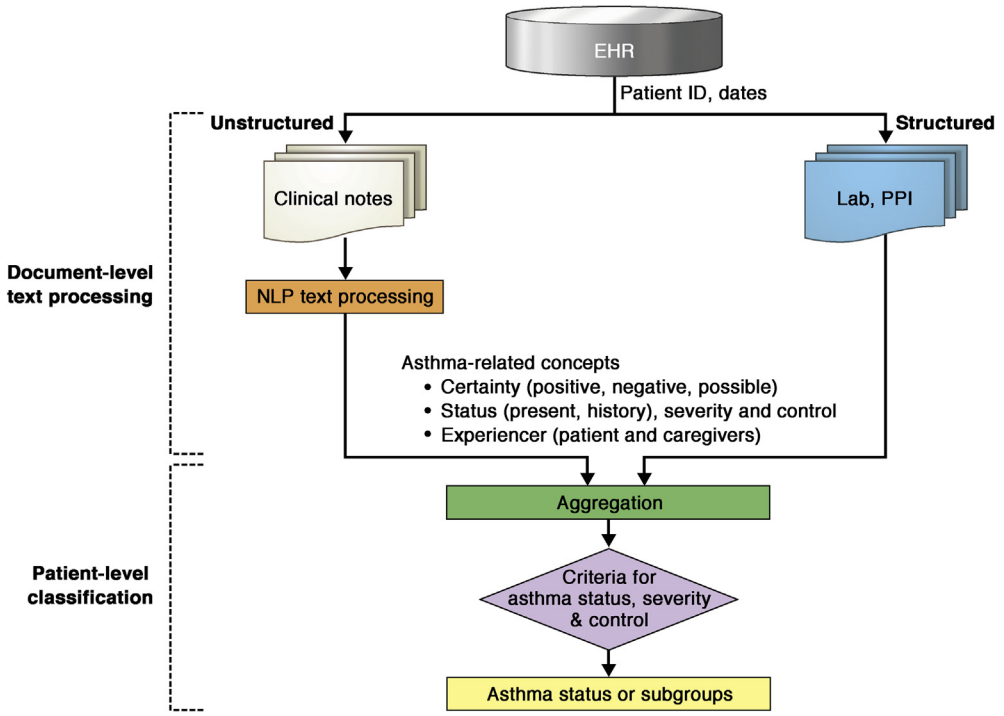

FIG 1. NLP Algorithms for document-level text processing and patient-level classification. NLP algorithms for cohort identification extract the information (or concepts) from EHRs, then process the extracted information, and finally classify patients into a subgroup according to rules or learners. Once information extraction and information processing from EHRs are completed, which occurs at a document level, aggregation and classification take place at a patient level on the basis of preestablished rules (eg, asthma criteria) (ie, a part of process developing NLP algorithm). *PPI*, Patient-provided information.

A few examples of higher-level NLP tasks include the following: (1) named entity recognition (identifying specific words or phrases such as disease, genes, or medication) and (2) setting up negation rules (eg, no history, denies history, absence of history). NLP is different from a simple key word search because a simple key word search will not be able to distinguish various forms of clinicians' notes in EHRs. For example, "patient has a known history of asthma" versus "there is no history of asthma," "patient reports a wheezing episode" versus "patient denies a wheezing episode," and "patient has a history of asthma" versus "sister has a history of asthma but no history of asthma in the patient." To differentiate these 2 sentences, NLP program applies a list of negation rules in addition to other NLP tasks described earlier. The National Library of Medicine provides several well-known knowledge infrastructure resources that facilitate these NLP tasks (eg, UMLS Metathesaurus for records synonyms and named entity recognition, and a text collection for word disambiguation).[39] Once information extraction and information processing from EHRs are completed, which occurs at a document level, aggregation and classification take place at a patient level on the basis of preestablished rules (eg, asthma criteria) (ie, a part of process developing NLP algorithm). A common criterion standard being used for NLP algorithm development is expert-based classification by annotating each specific feature (eg, wheezing episodes or physician diagnosis of asthma) by experts. This step typically requires some kind of annotation tool (eg, Anafora). Typically, training computer or learner takes multiple iterative processes until the computer correctly identifies (extract, process, and classify) the concept of interest using expert's annotation as a reference. NLP algorithms are developed or trained using a *developing or training cohort.* Once it reaches optimal performance in detecting the concept, the NLP algorithm is tested using an independent *testing cohort* (eg, a k-fold cross-validation method [eg, train/test split]). After the initial development and validation of NLP algorithms, it is necessary for investigators to further assess portability and external validation of NLP algorithms, which is one of the most challenging aspects before its implementation.

Use cases for NLP system in asthma research

As discussed earlier, as an example, we recently developed and validated NLP algorithms for 2 existing retrospective criteria for asthma: Predetermined Asthma Criteria (PAC) and Asthma Predictive Index described in Table I[20,22,23,37] and asthma prognosis.[21] Given the significant heterogeneity in determining asthma status for asthma care and research (eg, 60 different definitions in the literature) and the limited research leveraging free texts in EHRs for asthma research, our group developed NLP algorithms for PAC and Asthma Predictive Index by conducting a retrospective birth cohort study (training cohort = 430 and testing cohort = 500 from the 1997-2007 Mayo Clinic Birth Cohort [n = 8525]) that used clinician's manual chart review with annotation to apply both criteria as criterion standard and asthma status by NLP algorithms as a predictor. Despite the complex concepts of NLP-PAC to be captured from EHRs, NLP-PAC showed that sensitivity, specificity, PPV, and negative predictive value for asthma status by clinician's manual reviews

TABLE I. Two 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 3 conditions were present, and they were considered to have *probable* asthma if only the first 2 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 y or older)
   - Nasal polyps
   - Blood eosinophilia higher than 300/μL
   - Positive wheal and flare skin test results 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_1 or FVC less than 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[1]
   - Favorable clinical response to bronchodilator

Asthma Predictive Index

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 (≥4%)

FVC, Forced vital capacity.
Asthma Predictive Index (+): Frequent wheezing episodes (eg, 2 or more wheezing episodes per year) plus at least 1 of 2 major criteria or 2 of 3 minor criteria.

were 97%, 95%, 90%, and 98%, respectively. Therefore, our previous work on NLP algorithms for multiple asthma criteria demonstrates feasibility for developing NLP algorithms for patient-level classification of asthma status by applying complex clinical concepts on large scales. Furthermore, we demonstrated portability (generalizability) of NLP algorithms developed at our institution across different practice settings (different geographic area, patient population, documentation practice, and EHR system) by conducting a similar birth cohort study (training cohort = 298 and testing cohort = 297) in Sioux Falls, South Dakota.[20,25] Sensitivity, specificity, PPV, and negative predictive value for the NLP-PAC algorithm in predicting asthma status by manual chart review were 92%, 96%, 89%, and 97%, respectively, in a different practice setting.[20]

Importantly, NLP-PAC and NLP-API identify a group of asthmatic children with distinctive clinical and immunological characteristics (eg, persistent asthma, increased risk of asthma exacerbation, impaired lung function, increased risk of infection and T_H2-high immune profile among asthmatic children who met both NLP-PAC and NLP-API).[42]

For NLP algorithm for asthma prognosis (long-term remission, relapse, and persistent asthma; remission was defined by the absence of asthma events for 3 years consecutively) after the index date of asthma, sensitivity, PPV, and F-measure for asthma events were 94%, 98%, and 0.96%, respectively. An important feature of our NLP system for asthma criteria and prognosis is capability of detecting the index date of asthma or outcome events (ie, onset of asthma or remission or relapse), which is crucially important for an epidemiological investigation requiring temporality. These capabilities of NLP algorithms may enable scalable precision care or population management and reduce methodological heterogeneity in phenotyping in allergy, asthma, and immunology,[43] which often obscures true biological heterogeneity deterring translation of scientific findings into practice while reducing the burden of clinicians and researchers.

IMPLICATIONS OF NLP FOR EHR BASED ON CLINICAL RESEARCH AND CARE

Fig 2 illustrates modular components of EHRs (top row) that encompass specific clinical information in both structured (eg, laboratory data) and nonstructured (eg, clinical notes) formats (bottom row), and the multimodal capability of NLP can extract and process this information, and classify patients, enabling important clinical tasks (right column) such as quality report and clinical decision support. Thus, NLP might be an important method to make EHRs a helpful data source for addressing the needs and supporting important activities of clinicians and researchers while reducing their burden for chart review and data mining from EHRs (eg, 70% of clinicians using EHRs reported burnout).[8,9] For example, clinicians often need to meet regulatory compliance (eg, reporting asthma care quality measure to a state agency, Minnesota Community Measure[44]), which requires accurate determination of children with asthma as a denominator for the measure. NLP for asthma criteria could help clinicians efficiently complete this task while they still have control over asthma status suggested by NLP because NLP algorithms provide clinicians with evidence (ie, clinical notes supporting asthma diagnosis) for its logic. Another example is NLP might be able to retrieve and summarize pertinent information from EHRs for clinicians, enabling automated chart review leveraging free texts in EHRs, and thus, clinicians can make their management decision for asthma efficiently by not relying on manual chart review. Our recent study showed that automated chart review via the NLP system reduced clinicians' time for making their clinical decision by about 80%.[45] Application of AI or other information technology to clinical care and research is likely to be an important direction in the US health care system (eg, 21 Century Cures Act in 2016).[46] Because the performance of AI algorithms largely depends on high-quality input data, NLP as a data mining tool for EHRs will be a cornerstone for achieving high-value care and high-quality research, reducing methodological heterogeneity in phenotyping, a key step for advancing allergy and asthma research. The current interest and effort for establishing the interoperability standards for EHRs led by the Office of the National Coordinator for Health Information Technology under the 21 Century Cures Act has important implications on future EHR-driven clinical care and research.[47,48]

At the same time, there are also many challenges for NLP-based research to be addressed in the future: data quality issues (eg, missing and biases), privacy issue (hosting or transferring EHR data to cloud), algorithmic bias, lack of interoperability standards, and health information technology–related clinician burnout and workflow issues. For example, the 2019 NIH workshop on Machine Intelligence in Health Care highlighted trustworthiness, explainability, usability, transparency, and

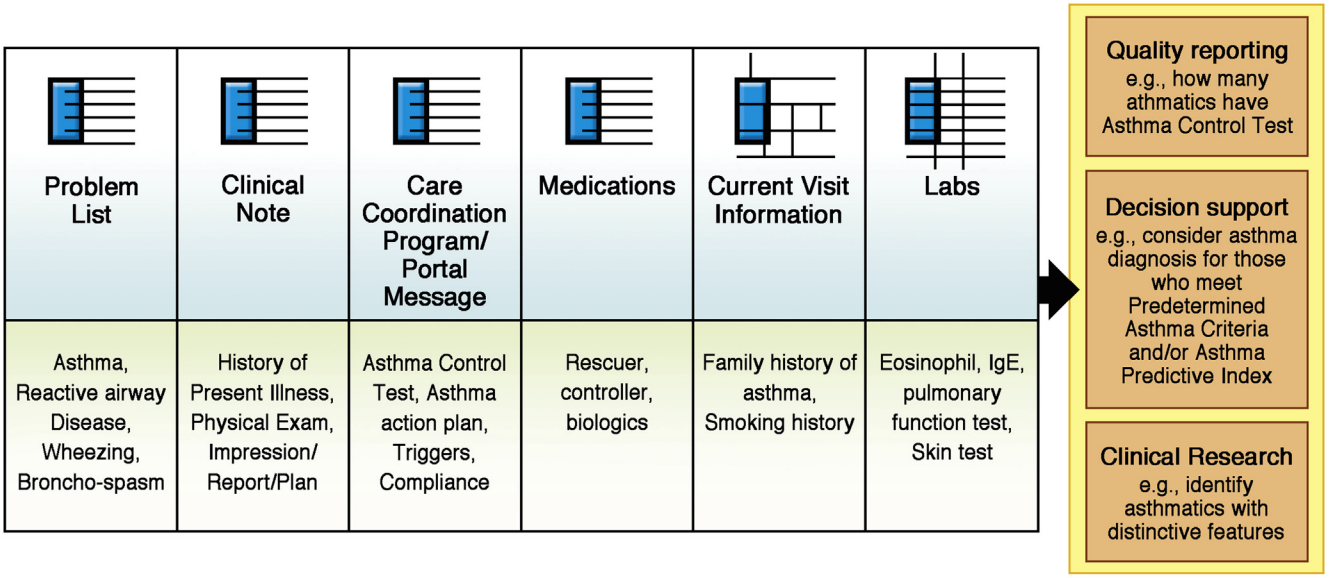

FIG 2. NLP Use cases for asthma. The figure illustrates modular components of EHRs (*top row*), which encompass specific clinical information in both structured (eg, laboratory data) and nonstructured (eg, clinical notes) formats (*bottom row*). It also shows the multimodal capability of NLP, which can extract and process this information, and classify patients, enabling important clinical tasks (right column) such as quality report and clinical decision support.

fairness of AI as potential challenges.[49] Although large-scale studies for asthma or allergy are primarily based on structured data (eg, *ICD* code) and are being routinely performed despite the known limitations, as discussed in our systematic review for the current literature, at present, research work leveraging free texts in EHRs via NLP is severely limited. Although NLP is not free from systematic biases from EHRs,[50,51] it may help us to identify algorithmic biases stemming from systematic biases of EHRs. Despite such limitations, EHR-based research as a source of real-world evidence is complementary to traditional randomized controlled trial–based evidence as recently recognized by the Food and Drug Administration.[15] Shortliffe and Sepúlveda[52] recently provided useful guidance for the must-have features of an informatics tool if it is to be accepted and integrated into routine workflow: (1) Black boxes are unacceptable (should provide why and how), (2) Time is a scarce resource, (3) Complexity and lack of usability thwart use, (4) Relevance and insight are essential, and (5) Delivery of knowledge and information must be respectful.

Given the current trends, impact, and potential of EHR-based research, innovative AI approaches for data mining and knowledge discovery from EHRs such as NLP should be pursued in education and training, clinical practice, and research in the field of allergy, asthma, and immunology. Recognizing unintended consequences, health information technology needs to be carefully and properly designed and implemented in a greater socioecological context of clinical care settings than a specific health information technology innovation itself.[53]

We are indebted to the review and comments by Drs James Li and Miguel Park in the Division of Allergy and Dr Chung Wi, Division of Community Pediatric and Adolescent Medicine and Precision Population Science Lab, at Mayo Clinic. Also, we thank Ms Kelly Okeson for administrative assistance for completing this manuscript.

REFERENCES

1. Topalovic M, Das N, Burgel PR, Daenen M, Derom E, Haenebalcke C, et al; Pulmonary Function Study Investigators; Pulmonary Function Study Investigators:. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. Eur Respir J 2019;53.
2. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nat Med 2019;25:70-4.
3. Mercan E, Mehta S, Bartlett J, Shapiro LG, Weaver DL, Elmore JG. Assessment of machine learning of breast pathology structures for automated differentiation of breast cancer and high-risk proliferative lesions. JAMA Netw Open 2019;2:e198777.
4. National Institutes of Health. Clinical and Translational Science Awards (CTSA): Foundation for Accelerated Discovery and Efficient Translation. 2011. Bethesda, Maryland. Available at: https://ncats.nih.gov/ctsa_2011/index.html. Accessed January 13, 2020.
5. Lemke AA, Wu JT, Waudby C, Pulley J, Somkin CP, Trinidad SB. Community engagement in biobanking: experiences from the eMERGE Network. Genomics Soc Policy 2010;6:35-52.
6. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21:578-82.
7. Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc 2019;26:106-14.
8. MedicalEconomics. EHRs should be a tool not a task. 2019. Available at: https://www.medicaleconomics.com/article/ehrs-should-be-tool-not-task. Accessed October 25, 2019.
9. Martin-Sanchez F, Verspoor K. Big data in medicine is driving big changes. Yearb Med Inform 2014;9:14-20.
10. Department of Biomedical Informatics, Harvard Medical School. National NLP Clinical Challenges (n2c2). 2019. Available at: https://n2c2.dbmi.hms.harvard.edu/. Accessed September 23, 2019.
11. Xiong Y, Shi X, Chen S, Jiang D, Tang B, Wang X, et al. Cohort selection for clinical trials using hierarchical neural network. J Am Med Inform Assoc 2019;26:1203-8.
12. Fang G, Annis IE, Elston-Lafata J, Cykert S. Applying machine learning to predict real-world individual treatment effects: insights from a virtual patient cohort. J Am Med Inform Assoc 2019;26:977-88.
13. Chen L, Gu Y, Ji X, Lou C, Sun Z, Li H, et al. Clinical trial cohort selection based on multi-level rule-based natural language processing system. J Am Med Inform Assoc 2019;26:1218-26.

14. Jarow JP, LaVange L, Woodcock J. Multidimensional evidence generation and FDA regulatory decision making: defining and using "real-world" data. JAMA 2017;318:703-4.
15. 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.
16. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc 2014;21:221-30.
17. 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.
18. 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.
19. Wu S, Liu S, Sohn S, Moon S, Wi C-i, Juhn Y, et al. Modeling asynchronous event sequences with RNNs. J Biomed Inform 2018;83:167-77.
20. 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.
21. Sohn S, Wi C-I, 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.
22. 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.
23. 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.
24. Sohn S, Wi CI, Juhn YJ, Liu H. Analysis of clinical variations in asthma care documented in electronic health records between staff and resident physicians. Stud Health Technol Inform 2017;245:1170-4.
25. Sohn S, Wang Y, Wi C-I, Krusemark EA, Ryu E, Ali MH, et al. Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions [published online ahead of print November 30, 2017]. J Am Med Inform Assoc. https://doi.org/10.1093/jamia/ocx138.
26. 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.
27. Himes BE, Kohane IS, Ramoni MF, Weiss ST. Characterization of patients who suffer asthma exacerbations using data extracted from electronic medical records. AMIA Annu Symp Proc 2008:308-12.
28. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019;25:433-8.
29. Gustafson E, Pacheco J, Wehbe F, Silverberg J, Thompson W. A machine learning algorithm for identifying atopic dermatitis in adults from electronic health records. IEEE Int Conf Healthc Inform 2017;2017:83-90.
30. Plasek JM, Goss FR, Lai KH, Lau JJ, Seger DL, Blumenthal KG, et al. Food entries in a large allergy data repository. J Am Med Inform Assoc 2016;23:e79-87.
31. Goss FR, Plasek JM, Lau JJ, Seger DL, Chang FY, Zhou L. An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes. AMIA Annu Symp Proc 2014; 2014:580-8.
32. Epstein RH, St Jacques P, Stockin M, Rothman B, Ehrenfeld JM, Denny JC. Automated identification of drug and food allergies entered using non-standard terminology. J Am Med Inform Assoc 2013;20:962-8.
33. Lai KH, Topaz M, Goss FR, Zhou L. Automated misspelling detection and correction in clinical free-text records. J Biomed Inform 2015;55:188-95.
34. Meystre SM, Deshmukh VG, Mitchell J. A clinical use case to evaluate the i2b2 Hive: predicting asthma exacerbations. AMIA Annu Symp Proc 2009;2009: 442-6.
35. Segura-Bedmar I, Colon-Ruiz C, Tejedor-Alonso MA, Moro-Moro M. Predicting of anaphylaxis in big data EMR by exploring machine learning approaches. J Biomed Inform 2018;87:50-9.
36. Blumenthal KG, Lai KH, Huang M, Wallace ZS, Wickner PG, Zhou L. Adverse and hypersensitivity reactions to prescription nonsteroidal anti-inflammatory agents in a large health care system. J Allergy Clin Immunol Pract 2017;5:737-43.e3.
37. Seol HY, Sohn S, Liu H, Wi C, Ryu E, Park MA, et al. Early identification of childhood asthma: the role of informatics in an era of electronic health records. Front Pediatr 2019;7:113.
38. Kimia AA, Savova G, Landschaft A, Harper MB. An introduction to natural language processing: how you can get more from those electronic notes you are generating. Pediatr Emerg Care 2015;31:536-41.
39. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011;18:544-51.
40. Friedman C, Rindflesch TC, Corn M. Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. J Biomed Inform 2013;46:765-73.
41. Pak HS. Unstructured data in health care. 2019. Available at: https://artificialintelligence.healthcaretechoutlook.com/cxoinsights/unstructured-data-in-healthcare-nid-506.html. Accessed October 22, 2019.
42. Seol HY, MC R, Wi C, Sohn S, Ryu E, Park MA, et al. Expert artificial intelligence-based natural language processing characterizes childhood asthma. BMJ Open Respiratory Research (In Press). 2020. http://dx.doi.org/10.1136/bmjresp-2019-000524.
43. 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 Respir J 2010;36:48.
44. Minnesota Community Measurement. 2017 Health Care Quality Report. Minneapolis (MN): Minnesota Community Measurement; 2017.
45. Juhn YWC, Sohn S, Ryu E, Park M, Fladager Muth J, Seol HY, et al. Asthma-Guidance and Prediction System (a-GPS) as a precision asthma care tool. Presented at: Annual American Academy of Allergy, Asthma and Immunology; Philadelphia, PA; March 16, 2020.
46. HIMSS. 21st Century Cures Act—a summary. 2017. Available at: https://www.himss.org/news/21st-century-cures-act-summary. Accessed October 25, 2019.
47. Tahir D. Provider groups: slow down info blocking rules. 2019. Available at: https://www.politico.com/newsletters/morning-ehealth/2019/09/18/provider-groups-slow-down-info-blocking-rules-478847. Accessed October 22, 2019.
48. Cohen J. ONC working on app privacy with Congress and White House. 2019. Available at: https://www.modernhealthcare.com/politics-policy/rucker-onc-working-app-privacy-congress-white-house. Accessed October 23, 2019.
49. Machine intelligence in healthcare. 2019. Available at: https://ncats.nih.gov/expertise/machine-intelligence#workshop. Accessed October 23, 2019.
50. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447.
51. Crown WH. Potential application of machine learning in health outcomes research and some statistical cautions. Value Health 2015;18:137-40.
52. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018;320:2199-200.
53. Bakken S. Building the evidence base on health information technology–related clinician burnout: a response to impact of health information technology on burnout remains unknown—for now. J Am Med Inform Assoc 2019; 26:1158.

The literature retrieval process for systematic review

A comprehensive search of several databases from January 1, 2000, to August 13th, 2019, English language, was conducted to understand the use of EHR for clinical research concerning allergy, asthma, and immunology in humans. The databases included Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, and Daily, Ovid EMBASE, Ovid Cochrane Central Register of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, Scopus, and Web of Science. The search strategy was designed and conducted by an experienced librarian with input from the authors. Controlled vocabulary supplemented with keywords was used to search for the use of AI approaches in asthma, immunology, and allergy using electronic health records. The specific key words used for literature search were based on the following three topics using a Boolean logic (AND):

1. Electronic health records and their derived searches (computer-based, computerized, computer-stored, medical record, medical records, patient record, patient records, health record, health records, automated chart review, EHR, EHRs, EMR, EMRs, physician order entries, or physician order entry); AND

2. Artificial intelligence (computer, machine, computational intelligence, computer reason, connectionist model, data mining, deep learning, expert system, fuzzy logic, informatics algorithm, informatics model, knowledge base, knowledge bases, knowledge engineering, knowledgebase, machine intelligence, machine learn, natural language processing, neural network, predictive modeling, and text mining); AND

3. Allergy, asthma and immunology (asthma, hypersensitivity, atopic dermatitis, atopic eczema, eczema, infantile eczema, allergic rhinitis, hay fever, allergen, allergens, allergology, immune, immunology, immunobiology, immunochemistry, neuroimmunology, photoimmunology)

This search returned 347 potential papers which met the search criteria. Both authors independently reviewed these papers (189 papers by YJJ and remaining papers by HL). From the review of abstracts, 44 papers met the initial screening criteria and were reviewed. Of the 44 papers, only 21 papers were included in this systematic review by excluding review papers (n=4), research protocol (n=1), unavailable full papers (n=2), non-EHRs based studies (n=6), non-matching topics (n=2) (allergy, asthma, and immunology), and non-NLP using papers (n=8).

eTable 1: The literature summary for EHR-based research using NLP in allergy, asthma and immunology between 2000 and 2019

Author and year of publication	Study Aim	Training cohort	Testing cohort	Study areas	Data extracted from EHRs	Type of data	Data mining methods	Validation methods (gold standard) for outcomes	Time comp	Internal validity for classification or prediction	Ext. validation
NLP-based research for asthma											
Afzal et al 2013(27)	Detect asthma by structured codes and free text	80% of 5032 children 5-18 years	20% of 5032 children 5-18 years	Asthma status	Asthma diagnosis, medications, diagnostic codes, and free texts	Both structured and unstructured	ML aggregating NLP, codes and standard terms	Expert-based asthma status by medical record review	No	PPV: 66%, sensitivity: 97%, specificity: 95%	No

Wu et al, 2014(18)	Detect predict index date of asthma	125 outpatients< 4 years	112 outpatients< 4 years	Asthma onset date	Asthma status by PAC and its index date	Unstructured	NLP-ML (cTAKES)	Expert-based asthma status and index date	Yes	PPV:80%, sensitivity: 92%, F1: 86%	No
Wu et al, 2013(19)	Detect asthma by NLP algorithm for asthma criteria	127 outpatients< 4 years	112 outpatients< 4 years	Asthma status	Asthma status in binary variable by Predetermined Asthma Criteria (PAC)	Unstructured	NLP logic and NLP ML (cTAKES)	Expert-based asthma status by PAC	Yes	sensitivity (recall): 81-85%, spec: 95-97%, PPV (precision): 84-88%, NPV: 94-95% depending on NLP	No
Himes et al 2009(56)	Predict COPD among asthma patients	9,349 adult outpatients	992 outpatients	ACOS	Age, race, sex, BMI, smoking status, WT, HT	Structured and unstructured	NLP logic and standard term (i2b2)	Expert-based COPD status	No	AUROC: 0.83	No
Himes et al 2008(28)	Characterize patients with asthma exacerbations	11356 adult asthmatics	1436 adult asthmatics	Asthma exacerbation	Age, BMI, race, sex and smoking history	Structured and unstructured	NLP logic and standard term (i2b2)	Administrative data (ICD code)	No	AUROC: 0.67	No
Wu et al 2018(20)	Classify childhood asthma prognosis	4013 children in birth cohort	Same cohort	Asthma prognosis	60 asthma-related named entities from free texts and asthma prognosis (no asthma, long-term remission, relapse, and persistent asthma)	Unstructured	NLP-logic	NLP-logic for asthma prognosis	Yes	PPV:55%, sensitivity: 60%, F1:57%	No

Wi et al 2018(21)	Test portability of NLP for Predetermined Asthma Criteria in a different practice setting	298 children in birth cohort	297 children in birth cohort	Asthma status	Asthma status by PAC and its index date	Structured and unstructured	NLP-logic and standard term	Expert-based asthma status by PAC	Yes	PPV:89%, sensitivity: 92%, specificity: 96%, NPV: 97%	Yes
Sohn et al 2018(22)	Classify childhood asthma prognosis	35 pediatric asthmatics	72 pediatric asthmatics	Asthma prognosis	Asthma status by PAC and its index date, asthma prognosis (remission, relapse and persistent asthma) of index dates of outcomes	Structured and unstructured	NLP-logic and standard term	Expert-based asthma prognosis	Yes	PPV: 98%, sensitivity 94%, F1: 96%	No
Kaur et al 2018(23)	Detect asthma status by Asthma Predictive Index (API)	87 children in birth cohort	427 children in birth cohort	Asthma status	Asthma status by API	Structured and unstructured	NLP-logic and standard term (ACE)	Expert-based asthma status	Yes	PPV: 88%, sensitivity: 86%, specificity 98%, NPV: 98%	No
Wi et al 2017(24)	Detect asthma status by Predetermined Asthma Criteria	430 children in birth cohort	500 children in birth cohort	Asthma status	Asthma status by PAC	Structured and unstructured	NLP-logic and standard term	Expert-based asthma status	Yes	PPV: 90%, sensitivity: 97%, specificity 95%, NPV: 98%	Yes
Sohn et al 2017(25)	Assess variations in asthma care documented by staff and pediatric residents	Not reported	9277 pediatric outpatients	Variations in asthma care	Asthma status by PAC and physician diagnosis and asthma outcomes	Structured and unstructured	NLP-logic and standard term	Expert-based asthma status	No	Cosine similarity for word (0.72), topic (0.99) and asthma-related concept (0.96)	No

Sohn et al, 2017(26)	Assess documentation variability and NLP system portability between two different health care settings	Not reported	298 children for each site	Documentation variations and portability of NLP system	Asthma status by PAC at each study sites by NLP	Structured and unstructured	NLP-logic and standard term	Expert-based asthma status	No	Cosine similarity for word (0.67), topic (0.94) and asthma-related concept (0.97), F1 for Mayo: 0.94, F1 for SCH: 0.91	No
Liang et al 2019(29)	Ascertain common childhood disease based on EHR following a knowledge model in China	3564 pediatric patients	2619 pediatric patients	Asthma diagnosis	Clinical notes (physical exam, etc.), laboratory data	Structured and unstructured	Deep learning NLP & Logistic regression	Expert-based system	No	Asthma was categorized separately as its own subcategory, and the accuracy ranged from 0.83 for cough variant asthma to 0.97 for unspecified asthma with acute exacerbation	No
NLP-based research for allergy entry in EHRs											
Epstein et al 2013(33)	Detect food and drug allergies from unstructured entries in EHRs	24599 perioperative notes	24857 perioperative notes	Food and drug allergy	Drug and food allergy entries	Unstructured	NLP logic (MTERMS)	Expert-based allergy status	No	PPV: 100%; sensitivity: 89 for free text entry and 96% for all allergy; F1: 94% for free text entry 98% for all allergy	No
Plasek et al 2016(31)	Detect food allergy by NLP and standardized terms	2730250 subjects	900 allergy entries	Food allergy	Food entries in allergy data repository	94% structured vs. 6% unstructured	NLP logic and standard term (MTERMS)	Expert-based food allergy status and SNOMED-CT and UNII matching	No	PPV: 100%, sensitivity: 97% F1: 99%	Yes

Lai et al 2015(34)	Detect misspelling and correction for allergies	2184 notes	442 EHR notes	Allergy entries	Misspelling for allergy and its correction	Unstructured	NLP-ML	Stanford Name Entity Recognition and other lexical resources	No	PPV: 96%, sensitivity: 93%, F1: 94%, accuracy: 88%	No
Goss et al 2014(32)	Detect allergy entries in EHR by NLP system	500 ED notes in EHRs	400 ED notes in EHRs	Allergy entries	Allergy entries in ED notes	Unstructured	NLP logic	Expert-based allergy status	No	PPV:71-100%, sensitivity: 69-93%, F1: 80-92% depending on allergies	No
NLP-based research for atopic dermatitis											
Gustafson 2017(30)	A Machine Learning Algorithm for Identifying Atopic Dermatitis in Adults from Electronic Health Records		Three non-overlapping data subsets stratified by class label: training (60%), validation (20%), testing (20%) of 562 adult patients.		Atopic dermatitis (AD) status	Structured and unstructured	NLP algorithms for AD concepts and ML classifiers	Expert-based system for criteria (Hanifin-Rajka and UK Working Party)	No	Precision (69-84%), recall (51-75%) and F (0.62-0.79) depending on code only, NLP only and code-NLP and type of AD criteria	No
NLP-based research for other areas											
Meystre et al 2009(35)	Detect smoking status	Not reported	900 sentences for smoking	Smoking status	Smoking status in free texts	Unstructured	NLP logic (i2b2)	Expert-based smoking status	No	PPV: 79%, specificity: 90%	No
Segura-Bedmar et al 2018(36)	Explore different machine learning techniques for classifying the incidences of anaphylaxis in Spain		75% of the data sets (217,890 records)	25% of the data sets	Anaphylaxis incidence	Structured and unstructured	NLP and standard terms and multiple ML classifiers	Expert-based system	No	Multiple ML classifiers assessed for performance. Precision (78-100%), recall (32-93%0 and F1(0.48-0.95)	No

Blumenthal et al 2017(37)	Determine the incidence and predictors of hypersensitivity reaction (HSR) to prescription of NSAIDs from allergy database of EHRs	62,719 patients were prescribed NSAIDs	Not applicable (no ML approach)	Incidence of HSR to prescription of NSAIDs from allergy database of EHRs	Structured and unstructured	NLP for standard terms and their mapping	Expert-based system	No	1,035 (1.7%) had an ADR, of which 189 (18.3%) were HSRs. Risk factors for HSR were assessed: Prior drug HSR history (OR: 1.8 [95% CI 1.3, 2.5]), female sex (OR 1.8 [95% CI 1.3, 2.4]), autoimmune disease (OR 1.7 [95% CI 1.1, 2.7]), and those prescribed the maximum standing NSAID dose (OR 1.5 [95% CI 1.1, 2.0])	No

ACE: Advanced Cohort Explorer; ACOS: Asthma-COPD Overlap Syndrome; ADR: Adverse Drug Reaction; API: Asthma Predictive Index; cTAKES: clinical Text Analysis and Knowledge Extraction System; eMERGE: Electronic Medical Records and Genomics; EHR: Electronic Health Records; GWAS: Genome-Wide Association Study; ML: Machine Learning; HSR: Hypersensitivity reaction; MTERMS: Medical Text Extraction, Reasoning, and Mapping System; NLP: Natural Language Processing; NSAIDs: Non-steroidal anti-inflammatory drugs; PAC: Predetermined Asthma Criteria; PPV: Positive Predictive Index; RR: Respiratory rate; SNOWMED-CT (Systematized Nomenclature of Medicine -- Clinical Terms); UNII: Unique Ingredient Identifier