The use of natural language processing to identify Tdap-related local reactions at five health care systems in the Vaccine Safety Datalink

Chengyi Zheng[a,*], Wei Yu[a], Fagen Xie[a], Wansu Chen[a], Cheryl Mercado[a], Lina S. Sy[a], Lei Qian[a], Sungching Glenn[a], Gina Lee[a], Hung Fu Tseng[a], Jonathan Duffy[b], Lisa A. Jackson[c,1], Matthew F. Daley[d], Brad Crane[e], Huong Q. McLean[f], Steven J. Jacobsen[a]

[a]Kaiser Permanente Southern California, Pasadena, CA, USA

[b]Centers for Disease Control and Prevention, Atlanta, GA, USA

[c]Kaiser Permanente Washington, Seattle, WA, USA

[d]Kaiser Permanente Colorado, Denver, CO, USA

[e]Kaiser Permanente Northwest, Portland, OR, USA

[f]Marshfield Clinic Research Institute, Marshfield, WI, USA

Abstract

Objective: Local reactions are the most common vaccine-related adverse event. There is no specific diagnosis code for local reaction due to vaccination. Previous vaccine safety studies used non-specific diagnosis codes to identify potential local reaction cases and confirmed the cases through manual chart review. In this study, a natural language processing (NLP) algorithm was developed to identify local reaction associated with tetanus-diphtheria-acellular pertussis (Tdap) vaccine in the Vaccine Safety Datalink.

Methods: Presumptive cases of local reactions were identified among members 11 years of age using ICD-9-CM codes in all care settings in the 1–6 days following a Tdap vaccination between 2012 and 2014. The clinical notes were searched for signs and symptoms consistent with local

[*]Corresponding author at: Department of Research and Evaluation, Kaiser Permanente Southern California, 100 S Los Robles Ave, 2nd Floor, Pasadena, CA, 91101, USA. Chengyi.X.Zheng@kp.org (C. Zheng).

Author's contributions

All authors have made substantial contributions to all of the following: (1) the conception and design of the study, or acquisition, analysis, or interpretation of data, (2) drafting the article and revising it critically, and (3) final approval of the version to be submitted. [1]previously Group Health Research Institute.

Conflicts of interest

The authors have no conflicts of interest relevant to this article to disclose.

Author declaration

This work was presented at the 2017 American Medical Informatics Association (AMIA) annual symposium. Otherwise, this work has not been published previously, that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to ethical standards. The authors do not have permission to share data.

reaction. Information on the timing and the location of a sign or symptom was also extracted to help determine whether or not the sign or symptom was vaccine related. Reactions triggered by causes other than Tdap vaccination were excluded. The NLP algorithm was developed at the lead study site and validated on a stratified random sample of 500 patients from five institutions.

Results: The NLP algorithm achieved an overall weighted sensitivity of 87.9%, specificity of 92.8%, positive predictive value of 82.7%, and negative predictive value of 95.1%. In addition, using data at one site, the NLP algorithm identified 3326 potential Tdap-related local reactions that were not identified through diagnosis codes.

Conclusion: The NLP algorithm achieved high accuracy, and demonstrated the potential of NLP to reduce the efforts of manual chart review in vaccine safety studies.

1. Introduction

The Vaccine Safety Datalink (VSD) project plays a critical role in monitoring adverse events after vaccinations [1]. Currently, safety outcomes in the VSD are primarily assessed using structured health care data such as diagnosis codes. However, diagnosis codes can be unreliable in terms of specificity and sensitivity [2,3]. Because of these concerns about the coding accuracy of structured data [4], manual review of medical records is often required to confirm the diagnosis, or provide additional detail that can aid in the outcome attribution. The process is labor-intensive, time-consuming, costly, and introduces additional variability.

Local reactions (also known as injection site reactions) are the most common vaccine adverse events (VAE) [5] and are included as primary outcome measures in vaccine clinical trials. Although local reactions are generally mild, patients may seek medical care for more severe manifestations of local reactions [6] which adds burden to healthcare systems and may contribute to vaccine hesitancy for some patients. Many factors could affect the risk of local reactions including type of vaccine, age of the patient, and location of the injection [7,8].

Study of vaccine-related local reactions can provide information on how to administer vaccine to minimize the risk of local reactions. For example, studies have evaluated whether there is reduced risk for local reaction based on the limb the vaccine was given, or reduced risk of local reaction specific to the first or second dose in a vaccine series. It can also help clinicians to provide better instruction to patients regarding possible local reactions following vaccination.

Based on the Brighton Collaboration case definition [9], local reaction is any morphological or physiological change at or near the injection site such as pain or redness. Since local reaction includes a group of signs and symptoms that are commonly associated with non-vaccine causes, and there is a lack of International Classification of Diseases (ICD) diagnosis codes for local reactions due to vaccination, it is difficult to identify vaccine-

related local reaction cases. Previous studies used a group of non-specific codes to identify potential cases of local reactions [7,8,10] and confirmed the cases by chart review [7]. However, due to the large number of possible cases identified when using non-specific codes, manual review of all suspected cases required significant resources.

Natural language processing (NLP) is a field of computer science that is focused on the processing of human language. NLP systems have been used to capture clinically important data that are unavailable or inaccurate in structured data fields from electronic health records (EHR) [11–13]. This study examined the feasibility of applying NLP to clinical notes to identify a medical condition (medically-attended local reactions) relevant to vaccine safety research.

2. Methods

2.1. Study population

The study was conducted among members at five U.S. health care systems participating in the VSD Project [14]. Each participating VSD site routinely creates structured datasets containing demographic and medical information (e.g., vaccinations, diagnoses) on its members [1]. For this study, participating sites also created text datasets of clinical notes from their EHR for the study population. When available, clinical notes associated with physical encounters (e.g., outpatient visits) and virtual encounters (email and telephone) were included. Kaiser Permanente Southern California (KPSC) served as the lead study site. Two other KP sites (Kaiser Permanente Colorado, Kaiser Permanente Northwest) and two non-KP sites (Group Health, Marshfield Clinic) participated. The Institutional Review Board of each participating organization approved this study.

Past studies of vaccine-related local reactions relied on diagnosis codes and used a risk window of 1–6 days after vaccination [8,15,16]. Events on the day of vaccination (Day 0) were excluded since Day 0 diagnosis codes often represent conditions with onset prior to vaccination [8,15,16]. In this study, the overall study population included patients of 11 years of age or older with tetanus–diphtheria–acellular pertussis (Tdap) vaccination between 2012 and 2014 (inclusive). We created three mutually exclusive sub-populations based on the codes and the risk window. The first two sub-populations were identified by ICD-9-CM codes (Table 1). These ICD-9-CM codes were previously used in other VSD studies [8,15,16]. To exclude conditions that were present before vaccination, presumptive cases with any of the codes in the 30 days prior to vaccination were excluded. The third sub-population included all vaccinees who were not included in the first two groups, who may or may not have had a medical encounter in the six days after vaccination, but who did not have any of the Table 1 ICD-9-CM codes assigned in the six days after vaccination. The purpose of this group was to examine the "false negative" rate of the traditional code-based approach.

2.1.1. Search by codes on Days 1–6—This group was created based on an approach similar to that of past studies [7,8]. Presumptive cases were identified using ICD-9-CM codes documented in all care settings in the 1–6 days following Tdap vaccination. For the selected patients, chart notes on Days 1–6 after vaccination were included.

The training dataset included 250 patients from KPSC. The size of training dataset was determined based on experience from past studies and it was created using the same criteria used in developing the reference standard (Appendix A).

The validation dataset included 250 different patients from KPSC and an additional 250 patients from the other four sites in proportion to their population size. Based on the sample size calculation with an assumption of 75% sensitivity and specificity, 500 cases would have a margin of error of 5%. Patients were randomly selected from the presumptive cases. The results of chart review (described below) were used to create the training and validation datasets.

2.1.2. Search by codes on Day 0—Using KPSC data only, we identified patients with a local reaction code only on the day of vaccination. All the chart notes from the day of vaccination were included.

2.1.3. Broad search without codes—This population included KPSC Tdap vaccinees who were not in the first two sub-populations. All the chart notes from the 1–6 days after vaccination were included.

2.2. Reference standard

Information from the medical record was abstracted by experienced chart abstractors according to an abstraction manual, and results were documented in the Research Electronic Data Capture (REDCap) system [17]. The REDCap system served as the database to record data abstracted from the medical record. Participating sites adhered to their standard practices for ensuring chart review quality. Quality review of completed abstraction forms against source medical records by a secondary abstractor was performed on 10% of the cases. A KPSC physician adjudicated cases that could not be confirmed as local reaction by the abstractors.

A presumptive case was considered a confirmed Tdap-related local reaction if 1) information in the medical record indicated the presence of signs or symptoms (S/S) consistent with a local reaction, including pain, tenderness, redness, warmth, swelling, itch, rash, induration, ulceration, lymphadenopathy, inflammation, and cellulitis; 2) those S/S were confirmed to be in the limb injected with Tdap; 3) the reaction was not known to have had onset before vaccination; and 4) the reaction did not clearly have another cause (e.g., injury, infection).

Chart abstraction of the training dataset and NLP algorithm development occurred in parallel. The NLP algorithm was built and refined based on the incremental release of training data. This process allowed identification of any issues or inconsistencies between chart abstraction and algorithm development. It also enabled us to measure the NLP algorithm's performance incrementally. Chart abstraction of the validation dataset was performed after training data abstraction was finished. The results of the chart review for the validation dataset served as the reference standard against which NLP was compared.

2.3. NLP algorithm

A rule-based NLP algorithm was developed and iteratively improved using the training dataset. The algorithm was implemented on a NLP system that was internally developed by KPSC which was based on NLTK [18], pyConText/NegEx [19], and Stanford NLP [20]. The final NLP program was executed locally at each participating site. Results without protected health information were sent back to KPSC for analysis.

First, the clinical notes were pre-processed through section detection, sentence separation, and tokenization (i.e., segmenting text into linguistic units such as words and punctuation). Second, keywords were compiled based on published case definitions and ontologies [21], and enriched by the training data to capture additional linguistic variations such as abbreviations and misspellings (Appendix B). Third, using these compiled terms, pattern matching was used to identify vaccination, S/S of local reaction, site(s) of vaccination and reaction, and cause. Negated terms and pre-existing conditions were identified and excluded. The site of reaction was compared to the vaccination site coded in the structured data, and excluded if the sites did not match. S/S with causes other than Tdap or clearly stated to be unrelated to vaccination were excluded. To identify a possible relationship between the outcome (local reaction) and the cause (e.g., Tdap), spatial information (e.g., specific body location) and temporal information (e.g., onset time) were also captured. The evidence (Table 2) identified was combined and assigned an output level between 1 and 8, with smaller values indicating stronger probability of being a true positive case (Table 3). A sample clinical note with NLP-identified concepts and relationship is provided in Appendix C.

For the Day 0 search, only output Levels 1–5 were treated as positive to increase specificity. The NLP algorithm was further modified to determine the temporal relationship between vaccination and S/S occurring on the same day. Since the vaccination data only contained the date without the time of vaccination [22], the timestamp of the notes was used to determine the sequence of events. Tdap was routinely administered at the end of a clinical encounter which typically lasted 15–30 minutes. Therefore, we grouped notes within a 30-minute window and treated them as a single encounter. Vaccine-related local reactions often were documented in a follow-up encounter. The S/S identified by NLP in the first encounter on Day 0 were classified as pre-vaccination symptoms since the vaccination likely had not yet been given based on chart review. The S/S identified in later encounters were classified as post-vaccination symptoms.

For the broad search, without the restriction of diagnosis codes, the S/S identified in Level 8 were often not caused by vaccine. Therefore, only output Levels 1–7 were considered true positive cases.

2.4. Analysis

For the Day 0 search, all the NLP positive cases were selected for chart review. For the broad search, we randomly selected 200 NLP positive cases for chart review.

The final NLP algorithm was tested on the manually chart reviewed reference standard validation dataset. We then used the numbers of true positive (TP), false positive (FP), true

negative (TN), and false negative (FN) cases to estimate the NLP algorithm's sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) [23,24]. The positive (or negative) likelihood ratios (LRs) indicate the degree of increase (or decrease) of the probability of having the outcome, if the test was positive (or negative) [25]. Confidence intervals were calculated using MedCalc [24]. To provide a more accurate estimate of NLP performance for the population from which the sample was drawn, the accuracy measurements were weighted to incorporate sampling fractions.

3. Results

Our study population included 1,904,595 patients 11 years of age who received Tdap. The KPSC population included 1,240,594 patients. Among the KPSC population, 5271 patients had a local reaction code on Days 1–6, of which 89.4%, 9.2% and 1.4% were diagnosed in a clinic, emergency room, and inpatient setting, respectively. Among the 5271 patients, 37.4% received multiple vaccines, and 4.6% were missing vaccination site information. The median number of notes per patient during Days 1–6 was 4 (interquartile range: 2–7). Some patients had over 100 notes due to their inpatient stay.

3.1. NLP performance for search by codes on Days 1–6

Of the 500 patients in the validation dataset, 134 (26.8%) were chart-confirmed as Tdap-related local reactions (Table 4). The NLP algorithm achieved an overall weighted sensitivity of 87.9%, specificity of 92.8%, PPV of 82.7%, and NPV of 95.1%. Compared to KPSC, the two other KP sites had lower PPV and the two non-KP sites had lower sensitivity (Appendix D). The frequencies of identified evidence types and NLP output levels are listed in Tables 2 and 3. The accuracies of the NLP output levels decreased as the levels increased.

After comparing the results of NLP and chart review, we identified ten cases of chart review errors caused by human error [26,27]. We corrected the chart review errors (n = 10), removed cases where notes were either not found (n = 9) or were missing key information (n = 4), and reported the results adjusted for the chart review errors in Appendix E. The adjusted estimates reflect the performance of the NLP algorithm after removing factors not related to NLP. The remaining NLP-related errors (n = 32) were caused by variable data quality in the notes and incorrect identification of cause or S/S (Table 5). The types of chart review errors are summarized in Appendix F. The frequencies of S/S identified by NLP are listed in Appendix G. The note types that were included or excluded are listed in Appendix H. NLP performance varied depending on the diagnosis code of interest (Appendix I). Of the 5271 patients, 10% had symptoms identified on the mismatched site.

3.2. NLP performance for search by codes on Day 0

Of the 45,744 cases with a Day 0 local reaction diagnosis code, the NLP algorithm identified 14 Day 0 cases. Five out of these 14 cases were confirmed as true Day 0 local reactions (PPV 35.7%).

3.3. NLP performance for broad search without codes

Of the 1,189,579 KPSC Tdap vaccinees that were not identified by the code-based approaches, 363,749 (30.6%) people had at least one note on Days 1–6. Among them, the NLP algorithm identified 3326 potential Tdap-related local reactions, and 59% were from virtual encounter notes. The PPV was 79.5% (159/200). Error analysis of the 41 false positive cases is presented in Appendix J. The estimated number of Tdap-related local reaction cases that did not have a local reaction code was 2644 (3326*79.5%); 1084 from physical encounters and 1560 from virtual encounters. In comparison, the estimated number of positive cases using codes was 1581 (5271*30.0%).

3.4. Estimated incidence rates

For the medically-attended Tdap-related local reactions, the estimated incidence rate based on the NLP search by codes on Days 1–6 was 11.9 per 10,000 vaccines (n = 2267, Table 6) for the overall population. For the KPSC population only, the estimated incidence rate based on the three NLP searches was 41.2 per 10,000 vaccinees (n = 1770 + 14 + 3326).

4. Discussion

In this study, the NLP system was deployed and executed at multiple institutions and, without transferring clinical notes, achieved reasonable accuracy in identifying a specific vaccine-related adverse event. This study demonstrates the feasibility of using NLP to reduce the potential burden of conducting manual chart review in future vaccine safety studies.

A text mining system for the Vaccine Adverse Event Reporting System (VAERS) was developed and achieved sensitivity of 83.1% and specificity of 73.9% for classification of anaphylaxis reports [28]. VAERS collects self-reported VAE from the public and is useful for safety signal detection and hypothesis generation. However, such spontaneous reporting systems only account for a small fraction of all adverse events [29], especially for less severe conditions such as local reactions [30]. Compared to VAERS reports, clinical notes cover comprehensive information not limited to VAE. Therefore, it is more difficult to identify VAE from clinical notes compared to VAERS.

Only one prior study used NLP to identify VAE from clinical notes [31]. The authors achieved PPVs of 64% and 74% for VAE and gastrointestinal-related VAE, respectively. However, they were not able to examine NPV. The NLP algorithm identified only 29.5% as many cases of gastrointestinal-related VAE as compared to the code-based approach. Therefore, manual chart review was still needed for both NLP and code identified cases. In addition, the study was performed at a single institution.

4.1. Overcoming the limitation of diagnosis codes

"False negatives" of diagnosis codes are not commonly investigated in vaccine safety studies. This is one type of outcome misclassification; if this occurred in a differential manner between vaccinated and comparison groups, this could lead to significantly biased results. The NLP algorithm identified more positive cases (n = 2644) from the broad search

of patients without ICD-9-CM codes than the code-based approach (n = 1581). Thus, NLP was able to identify what might be missed when outcomes of interest are identified through codes alone.

"False positives" of diagnosis codes were also identified in this study. The confirmed positive rate of codes was 28.2% by chart review (Table 4) and 31.0% (2267 of 7315 from Table 6) by NLP. To confirm the thousands of potential cases identified by diagnosis codes, manual chart review is infeasible. This study further demonstrates the limitation of diagnosis codes and the value of NLP for identifying vaccine-related local reactions.

Telephone encounters were often not coded by ICD-9-CM but they were an important data source in past vaccine safety studies [31,32]. However, the information documented in these notes was often not as complete as other types of clinical notes such as progress notes. This led to some false positive cases in the broad search due to the lack of a "possible" category in the reference standard. Among the false positive cases, 37% were confirmed as vaccine-related local reactions but lacked evidence to attribute the reaction to Tdap (15 of 41 cases from Appendix I).

NLP had the lowest PPV for lymphadenopathy and lowest NPV for complications of medical care. Diagnosis codes also had different PPV (0–73%) based on chart review. Future studies may utilize these findings to exclude false positive cases or strategically allocate chart review resources.

4.2.   Incidence rate estimation

The solicited local reaction rates in the clinical trials were about 70% for pain, 22% for erythema, and 20% for swelling for the two approved Tdap vaccines [33,34]. Most local reaction cases had minor symptoms where no medical attention was needed. In this study, the estimated incidence rate for local reactions (41.2 per 10,000 Tdap vaccinees ≈ 0.4%) provides a measure of medical attention associated with the vaccination, albeit an underestimate of the true incidence of local reactions. By moderating patients' perceptions of the risk of vaccination, it may help to reduce vaccine hesitancy [35]. It is worth noting that patients with different dosing intervals of toxoid vaccines may have different rates of medically-attended local reactions, although we did not examine that in this study. In comparison, other studies of Tdap-related local reactions have found that the code-based rate for pregnant women was 15.4/10,000 [10], and the rate for patients aged   65 years was 51.2/10,000 [8]. The chart-based rate for patients 9–25 years of age was 2.6/10,000 [16].

4.3.   Identifying rare events and temporality

We demonstrated the potential for NLP to tease out the temporal relationship for Day 0 events. For most Day 0 visits with local reaction codes, the adverse event had occurred before the administration of vaccine [22]. Because of this, in most vaccine safety studies, Day 0 cases are only included for rare and acute adverse events that are treated in emergency department or inpatient settings [22,36]. In this study, the estimated Day 0 event rate was 0.0004% for all Tdap vaccinees. For such a rare event, it is difficult to create a validation dataset to measure sensitivity. Therefore, our reported result must be interpreted with

caution. Despite this, based on the validated performance of NLP, the number of false negative cases should be small.

4.4. Performance variation among study sites

Variation among the study sites affected NLP performance. For example, some chart notes from one site were missing sentence-ending punctuation, such as a period between consecutive sentences. It is likely that the reduced performance at non-KPSC sites was the result of the lack of training data from these sites. A training dataset that represents all sites would help to reduce performance variation [37].

We allocated half of the validation cases among four participating sites based on their population size (Table 6). The small number of cases from Non-KP site 2 resulted in wider confidence intervals for the NLP performance measurements. For future NLP work, oversampling from smaller sites may help to narrow the confidence intervals for accuracy measurements.

As shown in Table 6, there was site variation in the prevalence of diagnosis codes and the availability of chart notes. Non-KPSC sites, especially the two non-KP sites, had lower NLP-based incidence rates than KPSC. Some of the contributing factors may be related to the differences in health care systems, utilization and clinical documentation patterns, and clinical visits to external facilities not documented in the EHR. Some of the false negative cases at the two non-KP sites were due to scanned images of external visits that were not available to NLP.

4.5. Beyond free text chart notes and NLP

In this study, the NLP process was mainly limited to chart notes, while chart abstractors reviewed all the information in the EHR. Potentially all structured and semi-structured data could be utilized in the algorithm. In this study, vaccine injection site and date were obtained from the structured data in the EHR. This information helped to rule out a vaccine-related cause if the reaction site did not match that of the injection site, or if the S/S began prior to vaccination. In addition, machine learning could be used together with NLP to improve accuracy by including structured (e.g. codes, care settings) and unstructured data (e.g. type and number of symptoms) [38].

4.6. Potential applications

Although this study demonstrated the higher accuracy of NLP on identifying one VAE compared to the diagnosis codes, NLP is not 100% accurate. Therefore, case confirmation of NLP results may still be needed for certain studies. It is worth noting that manual review (even with independent double review) may not be 100% accurate. Many published studies solely relied on the diagnosis codes which are often not accurate. Therefore, the usage of NLP in vaccine safety studies must consider the pros and cons of NLP and the specific requirements of the study. For studies that require manual chart review, NLP can be used to replace or facilitate manual chart review by narrowing down cases if the diagnosis codes lack specificity. Currently, we often proceed with studies assuming the coding of diagnoses

is complete. NLP could be applied to a population without ICD codes to increase the sensitivity of detecting conditions.

NLP is fast and can be automatically executed, therefore it may be well-suited for vaccine safety surveillance [39,40], and possibly identifying safety signals from clinical data, in a manner complementary to VAERS [41,42]. NLP can handle large amounts of data, permitting long-term follow-up of millions of vaccinees [41] and identification of rare outcomes [43,44]. NLP can be applied to identify attributions of adverse events [41,44], assess outcomes or confounders [43], determine temporality [44–46], and identify reasons for vaccine hesitancy [35,43].

4.7. Limitations

There are limitations to this study worth consideration. First, due to limitation of resources, we were not able to perform double chart abstraction for all the cases. Second, the small sample sizes for non-KPSC sites resulted in wide confidence intervals. Third, NLP has lower PPV when applied for Day 0 and without diagnosis codes.

Based on the error analyses of NLP and chart review (Appendix F and J), one common challenge was to correctly attribute the causes of identified S/S. Local reaction-related S/S are also associated with many possible causes. Causes are not always clearly known or documented in the clinical notes. Based on the analysis of the validation data, vaccine reaction was only stated in 31% of the positive cases (Table 2). The frequent co-administrations of other vaccines at the same time added another level of difficulty for cause attribution.

5. Conclusion

With more EHRs becoming available in different health care systems, new methods may be needed for conducting vaccine safety studies that have traditionally relied upon structured data and manual chart review. In this study, we demonstrated the uses of an NLP algorithm to identify VAE from the clinical notes at five VSD sites. These results suggest that NLP has the potential to reduce the efforts of manual chart review in vaccine safety studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

[1]. Baggs J, Gee J, Lewis E, Fowler G, Benson P, Lieu T, Naleway A, Klein NP, Baxter R, Belongia E, Glanz J, Hambidge SJ, Jacobsen SJ, Jackson L, Nordin J, Weintraub E, The Vaccine Safety Datalink: a model for monitoring immunization safety, Pediatrics 127 (Suppl. 1) (2011) S45–53. [PubMed: 21502240]

[2]. Verstraeten T, DeStefano F, Chen RT, Miller E, Vaccine safety surveillance using large linked databases: opportunities, hazards and proposed guidelines, Expert Rev. Vaccines 2 (2003) 21–29. [PubMed: 12901594]

[3]. Mullooly JP, Donahue JG, DeStefano F, Baggs J, Eriksen E, Predictive value of ICD-9-CM codes used in vaccine safety research, Methods Inf. Med 47 (2008) 328–335. [PubMed: 18690366]

[4]. Andrade SE, Scott PE, Davis RL, Li DK, Getahun D, Cheetham TC, Raebel MA, Toh S, Dublin S, Pawloski PA, Hammad TA, Beaton SJ, Smith DH, Dashevsky I, Haffenreffer K, Cooper WO, Validity of health plan and birth certificate data for pregnancy research, Pharmacoepidemiol. Drug Saf 22 (2013) 7–15. [PubMed: 22753079]

[5]. Kroger A, Duchin J, Vázquez M, General Best Practice Guidelines for Immunization: Best Practices Guidance of the Advisory Committee on Immunization Practices (ACIP), Centers for Disease Control and Prevention, 2017.

[6]. Cook IF, Best vaccination practice and medically attended injection site events following deltoid intramuscular injection, Hum. Vaccin. Immunother 11 (2015) 1184–1191. [PubMed: 25868476]

[7]. Jackson LA, Yu O, Nelson JC, Dominguez C, Peterson D, Baxter R, Hambidge SJ, Naleway AL, Belongia EA, Nordin JD, Baggs J, Vaccine Safety Datalink T., Injection site and risk of medically attended local reactions to acellular pertussis vaccine, Pediatrics 127 (2011) e581–587. [PubMed: 21300679]

[8]. Tseng HF, Sy LS, Qian L, Marcy SM, Jackson LA, Glanz J, Nordin J, Baxter R, Naleway A, Donahue J, Weintraub E, Jacobsen SJ, Vaccine Safety Datalink T., Safety of a tetanus-diphtheria-acellular pertussis vaccine when used off-label in an elderly population, Clin. Infect. Dis 56 (2013) 315–321. [PubMed: 23196953]

[9]. Gidudu J, Kohl KS, Halperin S, Hammer SJ, Heath PT, Hennig R, Hoet B, Rothstein E, Schuind A, Varricchio F, Walop W, Brighton S, Collaboration Local Reactions Working Group for a Local Reaction at or near Injection, A local reaction at or near injection site: case definition and guidelines for collection, analysis, and presentation of immunization safety data, Vaccine 26 (2008) 6800–6813. [PubMed: 18950670]

[10]. Sukumaran L, McCarthy NL, Kharbanda EO, McNeil MM, Naleway AL, Klein NP, Jackson ML, Hambidge SJ, Lugg MM, Li R, Weintraub ES, Bednarczyk RA, King JP, DeStefano F, Orenstein WA, Omer SB, Association of Tdap vaccination with acute events and adverse birth outcomes among pregnant women with prior tetanus-containing immunizations, Jama 314 (2015) 1581–1587. [PubMed: 26501534]

[11]. Wang X, Hripcsak G, Markatou M, Friedman C, Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study, J. Am. Med. Inform. Assoc 16 (2009) 328–337. [PubMed: 19261932]

[12]. Jensen PB, Jensen LJ, Brunak S, Mining electronic health records: towards better research applications and clinical care, Nature reviews, Genetics 13 (2012) 395–405.

[13]. Doan S, Conway M, Phuong TM, Ohno-Machado L, Natural language processing in biomedicine: a unified system architecture overview, Methods Mol. Biol 1168 (2014) 275–294. [PubMed: 24870142]

[14]. McNeil MM, Gee J, Weintraub ES, Belongia EA, Lee GM, Glanz JM, Nordin JD, Klein NP, Baxter R, Naleway AL, Jackson LA, Omer SB, Jacobsen SJ, DeStefano F, The Vaccine Safety Datalink: successes and challenges monitoring vaccine safety, Vaccine 32 (2014) 5390–5398. [PubMed: 25108215]

[15]. Jackson LA, Yu O, Belongia EA, Hambidge SJ, Nelson J, Baxter R, Naleway A, Gay C, Nordin J, Baggs J, Iskander J, Frequency of medically attended adverse events following tetanus and diphtheria toxoid vaccine in adolescents and young adults: a Vaccine Safety Datalink study, BMC Infect. Dis 9 (2009) 165. [PubMed: 19804643]

[16]. Jackson LA, Yu O, Nelson J, Belongia EA, Hambidge SJ, Baxter R, Naleway A, Nordin J, Baggs J, Iskander J, Risk of medically attended local reactions following diphtheria toxoid containing vaccines in adolescents and young adults: a Vaccine Safety Datalink study, Vaccine 27 (2009) 4912–4916. [PubMed: 19567245]

[17]. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG, Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support, J. Biomed. Inform 42 (2009) 377–381. [PubMed: 18929686]

[18]. Bird S, Loper Edward, Klein Ewan, Natural Language Processing With Python, O'Reilly Media Inc., Sebastopol, CA, 2009.

[19]. Chapman BE, Lee S, Kang HP, Chapman WW, Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm, J. Biomed. Inform 44 (2011) 728–737. [PubMed: 21459155]

[20]. Manning CD, Surdeanu Mihai, Bauer John, Finkel Jenny, Bethard Steven J., McClosky David, The stanford CoreNLP natural language processing toolkit, 52nd Annual Meeting of the Association for Computational Linguistics (2014) 55–60.

[21]. Bodenreider O, The unified medical language system (UMLS): integrating biomedical terminology, Nucleic Acids Res 32 (2004) D267–270. [PubMed: 14681409]

[22]. Jacobsen SJ, Sy LS, Ackerson BK, Chao CR, Slezak JM, Cheetham TC, Takhar HS, Velicer CM, Hansen J, Klein NP, An unmasking phenomenon in an observational post-licensure safety study of adolescent girls and young women, Vaccine 30 (2012) 4585–4587. [PubMed: 22580356]

[23]. Altman DG, Bland JM, Diagnostic tests 2: predictive values, Bmj 309 (1994) 102. [PubMed: 8038641]

[24]. Diagnostic test evaluation calculator, MedCalc Software.

[25]. Grimes DA, Schulz KF, Refining clinical diagnosis with likelihood ratios, Lancet 365 (2005) 1500–1505. [PubMed: 15850636]

[26]. Haley RW, Schaberg DR, McClish DK, Quade D, Crossley KB, Culver DH, Morgan WM, McGowan JE Jr., Shachtman RH, The accuracy of retrospective chart review in measuring nosocomial infection rates. Results of validation studies in pilot hospitals, Am. J. Epidemiol 111 (1980) 516–533. [PubMed: 7377196]

[27]. Biese KJ, Forbach CR, Medlin RP, Platts-Mills TF, Scholer MJ, McCall B, Shofer FS, LaMantia M, Hobgood C, Kizer JS, Busby-Whitehead J, Cairns CB, Computer-facilitated review of electronic medical records reliably identifies emergency department interventions in older adults, Acad. Emerg. Med 20 (2013) 621–628. [PubMed: 23758310]

[28]. Botsis T, Buttolph T, Nguyen MD, Winiecki S, Woo EJ, Ball R, Vaccine adverse event text mining system for extracting features from vaccine safety reports, J. Am. Med. Inform. Assoc 19 (2012) 1011–1018. [PubMed: 22922172]

[29]. Cullen DJ, Bates DW, Small SD, Cooper JB, Nemeskal AR, Leape LL, The incident reporting system does not detect adverse drug events: a problem for quality improvement, Comm. J. Qual. Improv 21 (1995) 541–548.

[30]. Singleton JA, Lloyd JC, Mootrey GT, Salive ME, Chen RT, An overview of the vaccine adverse event reporting system (VAERS) as a surveillance system. VAERS Working Group, Vaccine 17 (1999) 2908–2917. [PubMed: 10438063]

[31]. Hazlehurst B, Naleway A, Mullooly J, Detecting possible vaccine adverse events in clinical notes of the electronic medical record, Vaccine 27 (2009) 2077–2083. [PubMed: 19428833]

[32]. Mullooly JP, Crane B, Chun C, Trivalent inactivated influenza vaccine safety in children: assessing the contribution of telephone encounters, Vaccine 24 (2006) 2256–2263. [PubMed: 16375995]

[33]. Adacel, (Tetanus Toxoid, Reduced Diphtheria Toxoid and Acellular Pertussis Vaccine Adsorbed) [package insert], Sanofi Pasteur Inc., Swiftwater, PA, 2005.

[34]. BOOSTRIX, (Tetanus Toxoid, Reduced Diphtheria Toxoid and Acellular Pertussis Vaccine, Adsorbed) [package insert], GlaxoSmithKline, Research Triangle Park, NC, 2005.

[35]. Larson HJ, Jarrett C, Eckersberger E, Smith DM, Paterson P, Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012, Vaccine 32 (2014) 2150–2159. [PubMed: 24598724]

[36]. Greene SK, Kulldorff M, Lewis EM, Li R, Yin R, Weintraub ES, Fireman BH, Lieu TA, Nordin JD, Glanz JM, Baxter R, Jacobsen SJ, Broder KR, Lee GM, Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project, Am. J. Epidemiol 171 (2010) 177–188. [PubMed: 19965887]

[37]. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, Pacheco JA, Boomershine CS, Lasko TA, Xu H, Karlson EW, Perez RG, Gainer VS, Murphy SN, Ruderman EM, Pope RM, Plenge RM, Kho AN, Liao KP, Denny JC, Portability of an algorithm to identify rheumatoid arthritis in electronic health records, J. Am. Med. Inform. Assoc 19 (2012) e162–169. [PubMed: 22374935]

[38]. Zheng C, Rashid N, Wu YL, Koblick R, Lin AT, Levy GD, Cheetham TC, Using natural language processing and machine learning to identify gout flares from electronic clinical notes, Arthritis Care Res. (Hoboken) 66 (2014) 1740–1748. [PubMed: 24664671]

[39]. Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, Yin R, Brown JS, Platt R, Vaccine T Safety Datalink Rapid Cycle Analysis, Real-time vaccine safety surveillance for the early detection of adverse events, Med. Care 45 (2007) S89–95. [PubMed: 17909389]

[40]. Leite A, Andrews NJ, Thomas SL, Near real-time vaccine safety surveillance using electronic health records-a systematic review of the application of statistical methods, Pharmacoepidemiol. Drug Saf 25 (2016) 225–237. [PubMed: 26817940]

[41]. Shimabukuro TT, Nguyen M, Martin D, DeStefano F, Safety monitoring in the vaccine adverse event reporting system (VAERS), Vaccine 33 (2015) 4398–4405. [PubMed: 26209838]

[42]. Tse A, Tseng HF, Greene SK, Vellozzi C, Lee GM, Signal identification and evaluation for risk of febrile seizures in children following trivalent inactivated influenza vaccine in the Vaccine Safety Datalink Project, 2010–2011, Vaccine 30 (2012) 2024–2031. [PubMed: 22361304]

[43]. Glanz JM, Newcomer SR, Jackson ML, Omer SB, Bednarczyk RA, Shoup JA, DeStefano F, Daley MF, White Paper on studying the safety of the childhood immunization schedule in the Vaccine Safety Datalink, Vaccine 34 (Suppl. 1) (2016) A1–A29. [PubMed: 26830300]

[44]. IOM, Adverse Effects of Vaccines: Evidence and Causality, The National Academies Press, Washington (DC), 2012.

[45]. Zhou L, Hripcsak G, Temporal reasoning with medical data–a review with emphasis on medical natural language processing, J. Biomed. Inform 40 (2007) 183–202. [PubMed: 17317332]

[46]. Sun W, Rumshisky A, Uzuner O, Evaluating temporal relations in clinical text: 2012 i2b2 Challenge, J. Am. Med. Inform. Assoc. JAMIA 20 (2013) 806–813. [PubMed: 23564629]

Summary table

What was already known on this topic:

- The Vaccine Safety Datalink (VSD) plays a critical role in monitoring adverse events after vaccinations by utilizing the electronic health records.

- Most studies performed in the VSD rely on diagnosis codes and manual chart review for outcome identification and confirmation.

What this study added to our knowledge:

- A natural language processing (NLP) system that was developed at one institution and deployed and executed at multiple institutions achieved reasonable accuracy in identifying a specific vaccine-related adverse event.

- This study demonstrates the feasibility of using NLP to reduce the potential burden of conducting manual chart review in future vaccine safety studies.

- "False negatives" of diagnosis codes are not commonly investigated in vaccine safety studies. NLP can identify cases missed by diagnosis codes.

- NLP has many potential applications in future vaccine safety studies based on the considerations of the pros and cons of NLP and the specific requirements of the study.

Table 1

International Classification of Disease 9th Revision, Clinical Modification (ICD-9-CM) codes used to identify presumptive cases of local reaction.

ICD-9-CM codes and descriptions
289.3 Lymphadenitis
682.3 Cellulitis and abscess, upper arm and forearm
682.9 Cellulitis and abscess, unspecified site
683 Acute lymphadenitis
709.8 Other unspecified disorder of skin
709.9 Other unspecified disorder of skin and subcutaneous tissue
729.5 Pain in limb
729.81 Swelling of limb
785.6 Lymphadenopathy
995.2 Other and unspecified adverse effect of drug, medicinal and biological substance
995.3 Allergy, unspecified
999.3 Other infection after infusion, injection, transfusion, or vaccination
999.5 Serum reaction
999.9 Complications of medical care, not elsewhere classified

Table 2

Types of evidence identified by NLP from each clinical note.

Group	Evidence type	Type of information	Percentage (%)[*]
A	Sign or symptom	Outcome	100.0
B	Tdap	Cause	65.6
C	Vaccine or vaccination	Cause	64.0
D	Injection site or injection without non-Tdap terms	Spatial (location) and cause	60.0
E	Local reaction (explicitly stated)	Spatial (location) and cause	10.4
F	Vaccine site	Spatial (location) and cause	8.8
G	Tdap site such as "Tdap location"	Spatial (location) and cause	35.2
H	Sign or symptom occurred with the vaccine site	Spatial (location) and outcome	89.6
I	Vaccine reaction	Cause and outcome	31.2
J	Sign or symptom was caused by something other than Tdap	Cause	0

[*] Percentages in NLP positive cases among the validation data.

Table 3

Levels of output from the NLP system for combination of evidence types.

Combination of evidence types[*]	Output level	Percentage %[**]	FP rate %[***]
A and B and C and D and (E or F or G or H)	1	33.6	0
A and B and C and (E or F or G or H)	2	18.9	11.5
A and B and D and (E or F or G or H)	3	19.6	10.7
A and B and (E or F or G or H)	4	11.2	6.3
(A and C and D and (E or F or G or H)) or (A and H and I)	5	3.5	0
A and C and (E or F or G or H)	6	5.6	37.5
A and D and H	7	0.7	100
A and H and (not J)	8	7.7	81.9

[*]
The algorithm checks the combination of evidence types from top to bottom, produces an output level, and stops once it finds a match. The combination of evidence types and output levels were manually created based on the performance on the training data until the NLP algorithm achieved 100% sensitivity and specificity.

[**]
Percentages in NLP positive cases among the validation data.

[***]
False positive rate in the validation data by NLP output level.

Table 4

Accuracy measurements of NLP in identifying Tdap-related local reaction, as compared with chart-confirmed validation data.

Site	Reference Standard (n/N)	Chart confirmation rate (%)	TP	TN	FN	FP	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR +	LR −
KPSC	75/250	30.0 24.4–36.1	68	162	7	13	90.7 81.7–96.2	92.6 87.6–96.0	84.0 75.5–89.9	95.9 92.0–97.9	12.2 7.2–20.7	0.10 0.05–0.20
KP site 2	16/69	23.2 13.9–34.9	15	45	1	8	93.8 69.8–99.8	84.9 72.4–93.3	65.2 49.4–78.2	97.8 87.1–99.7	6.2 3.2–11.9	0.07 0.01–0.49
KP site 3	21/79	26.6 17.3–37.3	17	53	4	5	81.0 58.1–94.6	91.4 81.0–97.1	77.3 58.9–89.0	93.0 84.5–97.0	9.4 4.0–22.3	0.21 0.09–0.51
Non-KP site 1	14/75	18.7 10.6–29.3	11	59	3	2	78.6 49.2–95.3	96.7 88.7–99.6	84.6 57.8–95.7	95.2 87.8–98.2	24.0 6.0–96.2	0.22 0.08–0.60
Non-KP site 2	8/27	29.6 13.8–50.2	5	19	3	0	62.5 24.5–91.5	100 82.4–100	100 (NA)	86.4 72.1–93.9	NA	0.38 0.15–0.92
Overall[a]	134/500	26.8 23.0–30.9	116	338	18	28	86.6 79.6–91.8	92.3 89.1–94.9	80.6 74.3–85.6	94.9 92.4–96.6	11.3 7.9–16.3	0.15 0.09–0.22
Overall[b]	141/500	28.2 24.3–32.4	124	333	17	26	87.9 81.4–92.8	92.8 89.6–95.2	82.7 76.6–87.4	95.1 92.6–96.8	12.1 8.3–17.7	0.13 0.08–0.20

Note: 95% confidence intervals are displayed beneath point estimates.

N: number of cases reviewed; FN: false negative; FP: false positive; TN: true negative; TP: true positive.

PPV: positive predictive value; NPV: negative predictive value.

LR+/LR−: positive and negative likelihood ratio. A LR + value between 5–10 indicates a moderate increase, while a value larger than 10 indicates a strong and often conclusive increase in the probability. A LR-value between 0.1 – 0.2 indicates a moderate decrease, while a value less than 0.1 indicates a strong and often conclusive decrease of the probability.

[a]Unweighted results are reported for the entire validation sample (n = 500).

[b]Weighted overall measurements are based on the number of patients with selected diagnosis codes at each site.

Table 5

Error analysis of NLP false positives and false negatives.

Type of NLP Error	# of cases	Description
False positive		
Non-vaccine cause	7	Allergic reaction to antibiotic, cat/puppy scratch, infection, injury
Not local reaction	6	Lymphadenitis (chronic or near clavicle), neck pain radiated to arm, skin problem, foot cellulitis
Non-Tdap vaccine cause	5	Pneumococcal or shingles vaccine
Conflicting information	2	Both 'left' and 'right' arms were associated with the sign/symptom
False negative		
Missing punctuation	3	Multiple sentences were not separated which caused negation error
Cause misclassification	3	Tdap was given due to wound which had similar signs/symptoms as Tdap-related local reaction
Missed sign/symptom	3	Sign/symptom was documented in separate sentence apart from the body location
Missing information	2	Tdap or sign/symptom location was not documented
Misspelling	1	'site' was misspelled as 'sight': "at sight of tetanus booster"
Total	32	

Table 6

Total number of Tdap-related local reaction cases identified and estimated incidence rates for three different methods of case ascertainment.

Site	Study population	With codes [a]	Median age of patients with codes	With codes & notes [b]	Code-based incidence rate [c]	NLP positive [d]	NLP-based incidence rate [c]
KPSC	1,240,594	5271	42.9	5268	42.5	1770	14.3
KP site 2	151,998	486	45.5	483	32.0	162	10.7
KP site 3	158,574	489	52.1	489	30.8	145	9.1
Non-KP site 1	281,903	768	55.9	626	27.2	150	5.3
Non-KP site 2	71,526	301	54.6	257	42.1	40	5.6
Overall	1,904,595	7315	NA	7123	38.4	2267	11.9

[a] Number of patients identified using ICD-9-CM local reaction codes documented in all care settings in the 1–6 days following vaccination with Tdap.

[b] For the patients identified by ICD-9-CM codes, number of patients who had at least one chart note available for NLP in the 1–6 days after vaccination.

[c] Incidence rates represent the number of identified cases per 10,000 Tdap vaccinees as defined in the "Study population" column.

[d] Number of positive cases identified by NLP among patients with ICD-9-CM codes in the 1–6 days following vaccination with Tdap.