Electronic surveillance of patient safety events using natural language processing

Al Ozonoff,
Boston Children’s Hospital, MA, USA
Harvard Medical School, Boston, MA, USA

Carly E Milliren,
Boston Children’s Hospital, MA, USA

Kerri Fournier,
Boston Children’s Hospital, MA, USA

Jennifer Welcher,
Boston Children’s Hospital, MA, USA

Assaf Landschaft,
Boston Children’s Hospital, Boston, MA, USA

Mihail Samnaliev,
Boston Children’s Hospital, MA, USA
Harvard Medical School, Boston, MA, USA

Mehmet Saluvan,
Boston Children’s Hospital, MA, USA

Mark Waltzman,
Boston Children’s Hospital, MA, USA
Harvard Medical School, Boston, MA, USA

Amir A Kimia
Boston Children’s Hospital, MA, USA

Harvard Medical School, Boston, MA, USA

Abstract

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage). Article reuse guidelines: sagepub.com/journals-permissions

Corresponding author: Al Ozonoff, Harvard Medical School, 300 Longwood Ave, Boston, MA 02115, USA.
al.ozonoff@childrens.harvard.edu.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval
The research reported in this journal underwent ethics review and approval as Boston Children’s Hospital protocol number IRB-P00033040. It was deemed to be exempt from requirements for informed consent.
Objective: We describe our approach to surveillance of reportable safety events captured in hospital data including free-text clinical notes. We hypothesize that a) some patient safety events are documented only in the clinical notes and not in any other accessible source; and b) large-scale abstraction of event data from clinical notes is feasible.

Materials and Methods: We use regular expressions to generate a training data set for a machine learning model and apply this model to the full set of clinical notes and conduct further review to identify safety events of interest. We demonstrate this approach on peripheral intravenous (PIV) infiltrations and extravasations (PIVIEs).

Results: During Phase 1, we collected 21,362 clinical notes, of which 2342 were reviewed. We identified 125 PIV events, of which 44 cases (35%) were not captured by other patient safety systems. During Phase 2, we collected 60,735 clinical notes and identified 440 infiltrate events. Our classifier demonstrated accuracy above 90%.

Conclusion: Our method to identify safety events from the free text of clinical documentation offers a feasible and scalable approach to enhance existing patient safety systems. Expert reviewers, using a machine learning model, can conduct routine surveillance of patient safety events.

Keywords
databases and data mining; electronic health records; IT healthcare evaluation; machine learning; text mining; patient safety

Background and significance
Medical errors are serious and a leading cause of death in the United States. Collection of data on these events is a priority for patient safety yet many Electronic Medical Record (EMR) systems do not capture these events adequately. Patient safety surveillance is the ongoing collection, analysis, interpretation, and dissemination of safety outcome data within a healthcare institution or system. As is the case with public health surveillance, patient safety surveillance is greatly enhanced by Information Technology (IT).

The primary source of routinely collected patient safety surveillance data is the electronic patient safety reporting system based upon voluntary provider reports, despite known limitations of these systems. A recent national survey of patient safety culture in U.S. hospitals suggests that up to 33% of near miss safety events are not reported. Another U.S. study reported that traditional electronic patient safety reporting systems captured only 322 of 916 (35%) adverse events documented in the medical record. Ultimately the ‘gold standard’ for safety events is an event reported from any source that is reviewed and confirmed by the appropriate patient safety professional.

The limitations of existing reporting systems have led to enhancements and alternative complementary approaches such as electronic ‘trigger tools’. A typical trigger tool uses specific criteria and algorithms, typically accessing structured data fields within the EMR, to identify events and/or patient records of interest that may require further review. Structured data sources for triggers may be administrative-based (i.e. charges, claims) or
code-based (i.e. ICD-9/10, CPT). The trigger tool approach has been applied successfully for surveillance of sepsis and other conditions. Incidence of many important clinical events are now captured using structured data elements in the EMR, either directly or through the use of triggers. Nonetheless even clearly defined events such as venous thromboembolism (VTEs) are not entirely captured using only structured data sources.

The two data sources used routinely for safety surveillance of peripheral intravenous (PIV) infiltration and extravasation (PIVIE) events are the nursing flowsheets and the voluntary safety reports submitted by clinical staff and reviewed by our institutional patient safety teams. Nursing flowsheets are brief structured data collection forms intended to be completed on a regular basis (e.g. once each shift) for each patient by the nurse on duty. Voluntary safety reports are captured by a dedicated IT system, typically with some limited structured data fields as well as descriptive free text, and these data are managed by the institutional patient safety program. Ideally, they capture critical information from the nursing assessment of each patient, including a record of important patient safety events. Institutional policy prohibits the use of voluntary safety reports for research purposes, so the current work focuses on data available from nursing flowsheets.

There are many clinically important data such as signs and symptoms, symptom severity, or disease status which are not documented within EMR structured data fields. Instead these data are encoded in clinician-generated narrative text e.g. free text clinical notes written by physicians, nurses, or other healthcare providers. The primary purpose of these notes is to document all events and information relevant to clinical care. For this reason, the free text notes may capture clinical events of interest that are not captured in the existing structured data fields of the EMR or capture otherwise unavailable detail about these events. Extracting data reliably from clinical text presents significant informatics challenges. Several solutions use the techniques of Natural Language Processing (NLP) as an essential component of the Learning Healthcare System (LHS) i.e. optimizing health care by understanding and generating knowledge from empirical data. In the context of patient safety, the LHS concept is implemented via monitoring and review of available safety data to inform targeted quality improvement activities. Studying these events is the primary mechanism for institutional learning, yet any system solely relying on voluntarily reported events will be incomplete. Optimal surveillance integrates multiple complementary data sources to identify and capture events of interest.

**Objective**

We aimed to describe and evaluate our approach to surveillance of safety events captured in electronic data sources including structured data fields within the EMR and unstructured data including clinical notes. Successful adoption of our approach would enhance existing patient safety surveillance capabilities and increase opportunities for clinical improvement. We hypothesize that a) some patient safety events are documented only in the clinical notes and not in any other accessible source; and b) large-scale abstraction of event data from clinical notes is feasible. This initial demonstration focused on peripheral intravenous (PIV) infiltrate events which occur when IV fluids or medication enter the surrounding tissue and may range widely in severity and need for intervention.
Materials and methods

We describe our general approach to patient safety event surveillance and our initial pilot study with two key steps: (1) we use software-assisted manual review of retrospective data to develop a robust set of training data; then (2) we train a classifier for prospective surveillance with a specified positive predictive value to conduct quality assurance and estimate necessary time of review. We repeated those 2 steps on quarterly data extracts and trained a machine learning classifier on the entire manually reviewed cohort (Figure 1). We conducted these steps using pilot data from 2016, then simulated a prospective study using data from 2018. For each dataset (2016 pilot data and 2018 comprehensive data), we included nursing shift notes, IV team consultation notes, and plastic surgery notes, all textual sources in which a provider might document a PIVIE event. All documents were inpatient records from a single tertiary care pediatric center. These were obtained from the hospital data research repository which mirrors the clinical health records. To create our model we maximized the use of available variables. These included the words and N-grams within the clinical narrative of the text, regular expressions written by the domain expert, the numeric score assigned to each expression, and finally the cumulative score calculated for each clinical note. We first ran an SVM model using implementation SVM Light. The model hyper-parameters were -c (trade-off between training error): 0.01; -j (cost-factor): 0.5; and -t (kernel): 0 (i.e. linear kernel). Next we ran a random forest model using Python sklearn. The RF model used the SVM score, individual RegEx, and the cumulative RegEx score assigned to each document.

Constructing the Training Set

To establish ongoing surveillance of safety events, we first develop a high-quality training set with both positive and negative cases which display a range of perhaps unmeasured features i.e. key patterns of words or phrases associated with infiltrate events. We assembled a subset of clinical notes that we believed would contain a higher prevalence of event. For this study, we extracted free text notes from hospital admissions with admission date between 1 January 2016 through 30 June 2016, including nursing shift notes, IV team consultation notes, and plastic surgery notes.

We relied on available literature, the clinical experience of our team, and synonym search in the snomed-ct medical ontology to build a list of free-text regular expressions which might indicate documentation of a safety event. We accommodate misspellings and other variations with the use of wildcard characters. We then assign numeric weights to each regular expression, with higher weights for terms that appear to be more suggestive of the event of interest (Table 1). Notes that use one or more of these terms or common variants will tend to have higher scores relative to other notes. Expression weights were based on a clinical judgment of the relative predictive value of each expression, and thus are subjectively based on domain expertise. After a corpus of >400 documents were manually screened and labeled, the frequency of each expression was calculated across the corpus, and weights can be recalculated. The primary function of these weights is to rank-order notes for presentation during manual review and not for prediction purposes. Statistical prediction occurs via the ML document classifier, described below.
We calculate a total score for each note as the sum of the regular expression weights for each expression contained in the note. Thus, a note containing many expressions will generally have a higher score than a note containing few expressions. The relative weights of expressions allow a more flexible scoring regime than one simply based on frequencies. The resulting scores constitute an ordering of the notes based on the subjective expectation that the note describes an event of interest.

We conduct manual chart review of a subset of notes in order to augment the original data set with a new set of gold standard outcomes i.e. human review to determine whether an event meeting the case definition has occurred.\textsuperscript{39,40} We used our homegrown graphic user interface DrT (stands for Document review Tool http://www.documentreviewtools.com/) to reduce the time required for human review of charts with a number of key user interface features.\textsuperscript{40–42} For example, our software highlights each regular expression in the notes to speed manual review, and our list is sorted for efficient review of notes likely to correspond to events or non-events (Figure 2). After completing review of a specified number of notes, our training set is augmented by manual review and ready for input to supervised learning.

**Training the classifier**

We use the training data to fit a machine learning classifier model based on a Support Vector Machine (SVM) approach implemented using n-gram bag-of-words to predict case status.\textsuperscript{43} The model outputs predicted probabilities of case status for all notes in our original data set. We prioritize human review to identify cases with the highest SVM scores as likely cases, thus optimizing the yield of human review and minimizing time spent reviewing cases unlikely to be of interest. This review may improve the list of regular expressions, for example adding new expressions that identify positive cases or negative “look-arounds” to exclude expressions preceded or followed by a specific word or phrase.\textsuperscript{44} By conducting this additional human review to identify notes incorrectly classified by SVM as false positives or false negatives, we refine the list of regular expressions to the specific safety event and local practices of documentation. As events are ascertained during this review, we incorporate these newly determined outcomes into the original training set for future refitting and improvement of the SVM model.

**Machine learning for prospective surveillance**

Once the SVM model is trained on a high-quality data set, we can apply it prospectively. On a periodic (e.g. monthly or quarterly) basis, clinical notes are collected and the SVM classifier is applied to the new input data. Documents classified as likely events undergo human review to ascertain cases or to correct the document classification to false positive, and ascertained cases are then referred to the quality improvement teams responsible for event surveillance as a supplement to their existing data systems. Systematic review of a random sample of negative cases provides further quality assurance and may identify false negatives. Both false positive and false negative cases can be incorporated into the training data for future iterations and improvements of the classifier model. Targets for sensitivity and specificity are established according to the surveillance priority of positive versus negative predictive value, and can be monitored prospectively using a threshold based on the training data. For this study, we extracted free text notes from hospital admissions.
between 1 January 2018 through 31 December 2018, simulating prospective surveillance with four consecutive quarterly datasets.

**Random forest**

Finally we trained a random forest model using the python library sklearn (htt), using available data fields from the training data set: SVM score, different RegEx scoring methods (user-defined vs automated weight), specific regular expression occurrence, and number of words/characters per report.45

**Evaluation and statistical analysis**

We compared the frequency and severity of events captured via the document classifier with existing data sources to capture infiltrate events, including voluntary event reports and nursing data captured in the electronic medical record. We noted when cases were captured across multiple systems. We included 95% confidence intervals (CIs) whenever we calculated proportions such as accuracy, sensitivity, specificity, and positive predictive value. We used kappa chance-corrected measure of agreement to assess inter-rater agreement. To evaluate prospective performance, we created a random forest model using the SVM and RegEx scores and used a split-sample cross-validation with a 70–30 train-test split to estimate sensitivity, specificity, and accuracy. We used these metrics to estimate the potential yield of identified events and total review time necessary to conduct ongoing prospective surveillance.

**Results**

**Training set development (study Phase 1)**

We used a corpus of 21,362 clinical notes from 3150 unique patients admitted to our pediatric hospital during the initial study period. After sorting notes according to subjective ranking based on RegEx weights, we classified a total of 2342 notes via manual review with a binary indicator (yes/no) for the documentation of a new infiltrate event.

The data coordinator retrieved the notes from our hospital data repository, quality assured the data, and verified that reported events were captured accurately. The biostatistician supervised the data process. All clinical notes were exclusively reviewed by an experienced nurse (KF), and a minority by a Pediatric Emergency Medicine Provider (AAK). The cumulative time of all researchers involved was approximately 30 hours total of data processing and review time.

We identified 125 events documented in the clinical notes, of which n = 44 (35%; CI 27–44%) were not previously reported to the safety staff. Patient safety staff did not ascertain the severity of events identified from text notes alone, and so we assigned provisional grades based on information available in the medical record. Briefly, Grade 1 events are characterized by mild redness and swelling, and increasing stage is associated with increasing severity up to Grade 4 events characterized by persistent redness, moderate-severe pain, and necrosis.46 Among these events, we determined n = 24 (55%) were Grade 1, n = 11 (25%) were Grade 2, n = 8 (18%) were Grade 3, and n = 1 (2%) was Grade 4.
Manual review

Nursing handoff notes were generally brief, with a mean number of words of 392 (95% CI 363–422). On a random sample of 250 notes with RegEx highlights manually reviewed, the mean review time per chart was 13 seconds (95% CI 12.1–14.1 seconds).

Evaluation of prospective performance (study Phase 2)

We screened 60,375 clinical notes and identified 1467 notes indicating an infiltration, corresponding to n = 440 unique infiltrate events (because some infiltrates were documented in multiple notes). Not surprisingly, the first clinical note to describe an event was the bedside nurse note (68% of events) followed by the IV team notes (31%) and finally consultant notes (91%). The latter group often identified a known infiltrate as we sorted our documents by date of documentation. The random forest model produced a receiver operating characteristic (ROC) curve with area-under-the-curve (AUC) of 0.96 (Figure 3). Choosing an optimal threshold for binary classification, our model had a sensitivity of 94.7% (95% CI 91–99%), specificity of 90% (95% CI 87–93%) and an accuracy of 91% (95% CI 89–94%). Inter rater agreement with the human reviewer was K = 0.8 (95% CI 0.74–0.86).

Systems-level evaluation

In our evaluation of prospective surveillance, we chose an appropriate threshold to minimize human review since that represents the greatest cost to operationalize ongoing surveillance. Assuming performance metrics above, and with a prevalence of 0.1% of notes documenting an infiltrate event, prospective surveillance of over 60,000 notes annually will yield roughly 360 notes biweekly basis containing 95% of all notes which document an infiltrate. Roughly one-third of these notes will be classified automatically by the model with a positive predictive of 100%, leaving 240 notes for manual review. Using the upper confidence limit of 14.1 seconds of manual review time per note with RegEx highlighted fields, we estimate total review time at our institution of 56 minutes biweekly or just under 24 hours per year of review time to maintain ongoing prospective surveillance.

Discussion

We present a general methodological approach for retrospective or prospective surveillance of patient safety events such as PIV infiltrates that may be difficult to capture completely with existing reporting systems or trigger tools. Although we have validated our approach with surveillance of infiltrate events at a pediatric hospital, our approach is readily generalizable and applicable across many healthcare settings and conditions. We designed our system with low operating costs of IT equipment and personnel time, acknowledging the reality that investments in safety systems must offer high value-to-cost ratio.47,48 We hypothesized that there were infiltrate events that we could identify from free text notes which could not be identified through existing systems. Using our pilot data, we found 44 such events, of which the majority were relatively low severity and therefore less likely to be reported. However, 9 of 44 (21%) events were suspected Grade 3 or 4 events suggesting that
additional surveillance would identify some important safety events not otherwise captured using existing sources.

Our evaluation study demonstrated that when provided a high-quality training set, the SVM model could classify unstructured free-text notes with reasonably high sensitivity and specificity. This is important to reduce the necessary human review that follows classification. There are two common workflows to capture similar events from EHR data. One approach is to select a subsample of medical records for full manual review, but the relatively low incidence of events makes this approach inefficient and costly. Another approach uses electronic ‘triggers’ to identify events using structured data, but these workflows suffer from low sensitivity and/or low specificity unless supplemented by manual review and annotation of events.49 We do not rely entirely on the classifier for identification of events. Instead the machine learning model enhances human intelligence by prioritizing those notes which require ascertainment and/or adjudication using manual chart review.

Our estimate of greater than 90% accuracy when classifying notes establishes that our approach can reduce time needed for human review, thus allowing large scale screening of free-text at a volume which makes screening all inpatient hospital notes a feasible part of ongoing safety event surveillance. Our results strongly support the hypothesis of having multiple data sources to address a problem. In our case, adding clinical notes nearly doubled the cases identified even with a model that captures only 90% of documents describing an infiltrate. Further work should focus on improving case identification by adding structured data elements to a multivariable model such as random forest or recursive partitioning.

There have been repeated calls for improved patient safety surveillance, echoing the words of CDC epidemiologist Stephen Thacker who wrote ‘surveillance is the cornerstone of public health practice’.50 It is widely accepted that complete, accurate, reliable data systems to capture important events of interest are foundational to patient safety efforts,32,33 and generally we have seen corresponding investments in system-wide improvements and research of new methods.51,52 However, no surveillance system is perfect and every system will have blind spots. Fundamental to a robust surveillance approach to patient safety will be the development of complementary systems,53 which may overlap substantially but which mutually address blind spots and thereby provide more complete surveillance coverage. For example, influenza surveillance within the U.S. includes no less than 20 data sources covering different aspects of the disease and its spread.54

We hypothesized that our methods for ongoing prospective surveillance using NLP and machine learning would prove feasible to conduct prospectively as an integrated component of hospital safety event surveillance. Our results establish feasibility, where a small team completed review of 12 months of hospital-wide inpatient data without large investments in resources. Annualized staff time would be less than 60 hours per year and computing costs are negligible. Most of the data capture and data processing can be automated for the purposes of ongoing surveillance. While a formal study of cost effectiveness is outside the scope of the current study, our early results suggest low operational costs.
Adoption of information technology in healthcare institutions is a key strategy to improving quality and patient safety.\textsuperscript{1,2,55} Since publication of the IOM reports on patient safety and quality, adoption of information technology in the U.S. healthcare institutions has increased dramatically.\textsuperscript{56} Despite enhanced IT adoption and sustained efforts to improve healthcare quality and patient safety in the U.S.,\textsuperscript{57,58} further progress remains a national priority. While there have been important advances in medical informatics to improve patient health and healthcare delivery, AHRQ and other federal agencies acknowledge that these advances have not been as rapid in the field of patient safety. Accelerating applications of informatics and machine learning to improve safety remains an important research priority.\textsuperscript{24,59–61}

To meet this call for scalable and effective informatics approaches to improve patient safety, we have designed a system with a focus on low technical requirements, low training costs, and low operational costs in the form of manual review. Future work will include formal cost effectiveness studies, including estimation of the incremental cost to identify each additional event. We believe our semi-supervised approach strikes a balance between fully automated approaches which require up-front investment in technology and expertise versus fully manual active surveillance approaches which may be costly and therefore limited in scale and sustainability. Our process is semi-automated with a goal of scalability while controlling linear growth in costs. We therefore find a happy medium between the extremes of fully intelligent software and fully manual chart review.

We note some limitations with our study. Our approach has been evaluated at a single pediatric academic medical center. Institution-specific findings regarding cost of operation and reporting completeness may not generalize to other institutions. We are integrating the human expertise of the study team with deterministic NLP modeling. Our results depend in part on the subjective inputs to specify the model, as well as the subjective assessment of the expert raters to review events. We note the careful balance that is required when human expertise in the form of domain expert is used to operate a system for surveillance of a medically complex event. Severity grades of infiltrate events were not validated with the hospital patient safety team.

Conclusion

Our method to identify safety events from the free text of clinical documentation offers a feasible, low-cost, transferrable and scalable approach to enhance existing patient safety systems and better inform quality improvement efforts. We have demonstrated that expert reviewers, assisted by NLP and a machine learning model, can conduct cost-effective enhanced routine surveillance of patient safety events. This is an important step to establish a Learning Healthcare System that can sustain ongoing data collection and learning activities integrated with existing operational and quality improvement systems.

Acknowledgements

The authors gratefully acknowledge support for this research funded by AHRQ grant 5R01HS026246. This work was conducted with support from Harvard Catalyst. The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award Award UL1 TR002541) and financial contributions from Harvard University and its affiliated academic healthcare centers. The content is
solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic healthcare centers, or the National Institutes of Health.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Agency for Healthcare Research and Quality (5R01HS026246) and the Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award Award UL1 TR002541).

References

1. IOM. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press, 1999.
2. IOM. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC: National Academy Press, 2001.
3. Nathan M, Gauvreau K, Samnaliev M, et al. Technical Performance Score Predicts Resource Utilization in Congenital Cardiac Procedures. J Am Coll Cardiol 2016; 67(22): 2696–2698. [PubMed: 27256837]
4. Bowman S Impact of electronic health record systems on information integrity: quality and safety implications. Perspect Health Information Management 2013; 10(Fall): 1c.
5. Thacker SB, Qualters JR and Lee LM, Control Cfd Prevention. Public health surveillance in the United States: evolution and challenges. MMWR Surveill Summ 2012; 61(Suppl): 3–9.
6. Porta M A dictionary of epidemiology. Oxford, UK: Oxford University Press, 2008.
7. Langmuir AD. The surveillance of communicable diseases of national importance. The New Engl Journal Medicine 1963; 268: 182–192.
8. Lombardo JS and Buckeridge DL. Disease surveillance: a public health informatics approach. Hoboken, NJ: John Wiley & Sons, 2012.
9. Stavropoulou C, Doherty C and Tosey P. How Effective Are Incident-Reporting Systems for Improving Patient Safety? A Systematic Literature Review. The Milbank Quarterly 2015; 93(4): 826–866. [PubMed: 26626987]
10. Pronovost PJ, Morlock LL, Sexton JB, et al. Improving the Value of Patient Safety Reporting Systems. In: Henriksen K, Battles JB, Keyes MA, et al. (eds). Advances in Patient Safety: New Directions and Alternative Approaches, Vol 1. Rockville MD: Assessment), 2008.
11. Macrae C The problem with incident reporting. BMJ Quality Safety 2015; 25: 71–75.
12. Famolaro T YN and Burns W. Hospital Survey on Patient Safety Culture 2016 User Comparative Database Report. Rockville: Agency for Healthcare Research and Quality, 2016.
13. Melton GB and Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. J Am Med Inform Assoc : JAMIA 2005; 12(4): 448–457. [PubMed: 15802475]
14. Wachter R and Gupta K. Understanding Patient Safety. 3rd Edition. New York, NY: McGraw-Hill Education, 2017.
15. Bhise V, Sittig DF, Vaghani V, et al. An electronic trigger based on care escalation to identify preventable adverse events in hospitalised patients. BMJ Quality Safety 2017; 27: 241–246.
16. Wong BM, Dyal S, Etchells EE, et al. Application of a trigger tool in near real time to inform quality improvement activities: a prospective study in a general medicine ward. BMJ Quality Safety 2015; 24(4): 272–281.
17. Kaafarani HM, Rosen AK, Nebeker JR, et al. Development of trigger tools for surveillance of adverse events in ambulatory surgery. Qual Safety Health Care 2010; 19(5): 425–429.
18. Mitchell I, Schuster A, Smith K, et al. Patient safety incident reporting: a qualitative study of thoughts and perceptions of experts 15 years after “To Err is Human. BMJ Quality Safety 2016; 25(2): 92–99.
19. Stockwell DC, Landrigan CP, Schuster MA, et al. Using a Pediatric Trigger Tool to Estimate Total Harm Burden Hospital-acquired Conditions Represent. Pediatr Qual Saf 2018; 3(3): e081. [PubMed: 30229193]
20. Rhee C, Kadri S, Huang SS, et al. Objective Sepsis Surveillance Using Electronic Clinical Data. Infect Control Hospital Epidemiology 2016; 37(2): 163–171.

21. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. Jama 2011; 306(8): 848–855. [PubMed: 21862746]

22. Collins SA and Vawdrey DK. Reading between the lines” of flow sheet data: nurses’ optional documentation associated with cardiac arrest outcomes. Appl Nursing Research : ANR 2012; 25(4): 251–257. [PubMed: 22079746]

23. Waitman LR, Warren JJ, Manos EL, et al. Expressing observations from electronic medical record flowsheets in an i2b2 based clinical data repository to support research and quality improvement. AMIA Annu Symp Proceedings AMIA Symp 2011; 2011: 1454–1463.

24. Lavin MA, Harper E and Barr N. Health information technology, patient safety, and professional nursing care documentation in acute care settings. Online J Issues Nurs 2015; 20(6).

25. Haller G, Myles PS, Stoelwinder J, et al. Integrating incident reporting into an electronic patient record system. J Am Med Inform Assoc : JAMIA 2007; 14(2): 175–181. [PubMed: 17213499]

26. Gong Y, Kang H, Wu X, et al. Enhancing Patient Safety Event Reporting. A Systematic Review of System Design Features. Appl Clinical Informatics 2017; 8(3): 893–909.

27. Doan S, Conway M, Phuong TM, et al. Natural language processing in biomedicine: a unified system architecture overview. In: Clinical Bioinformatics. Berlin, Germany: Springer, 2014, pp. 275–294.

28. Raghupathi W and Raghupathi V. Big data analytics in healthcare: promise and potential. Health Information Science Systems 2014; 2(1): 3. [PubMed: 25825667]

29. Ohno-Machado L. Realizing the full potential of electronic health records: the role of natural language processing. J Am Med Inform Assoc : JAMIA 2011; 18(5): 539. [PubMed: 21846784]

30. Middleton B, Bloomrosen M, Dente MA, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc : JAMIA 2013; 20(e1): e2–e8. [PubMed: 23355463]

31. Leistikow I, Mulder S, Vesseur J, et al. Learning from incidents in healthcare: the journey. not the arrival, matters. BMJ Qual Saf 2017; 26(3): 252–256.

32. Cullen DJ, Bates DW, Small SD, et al. The incident reporting system does not detect adverse drug events: a problem for quality improvement. The Joint Comm Journal Quality Improvement 1995; 21(10): 541–548.

33. Lee LM, Teutsch SM, Thacker SB, et al. Principles and Practice of Public Health Surveillance. 3rd Edition ed. New York: Oxford University Press, 2010.

34. Joachims T. Making large-Scale SVM Learning Practical. In: Schölkopf B, Burges C and Smola A (eds). Advances in Kernel Methods- Support Vector Learning. Cambridge, MA: MIT-Press, 1999.

35. Scikit. sklearn,ensemble. RandomForestClassifier, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (accessed 7/19/2022).

36. Kimia AA, Savova G, Landschaft A, et al. An Introduction to Natural Language Processing: How You Can Get More From Those Electronic Notes You Are Generating. Pediatr Emergency Care 2015; 31(7): 536–541.

37. Nash SK. Nonsynonymous synonyms: correcting and improving SNOMED CT. AMIA Annu Symp Proc 2003; 2003: 949. [PubMed: 14728454]

38. Friedl JE. Mastering Regular Expressions: Understand Your Data and Be More Productive. Sebastopol, CA: O’Reilly Media, Inc, 2006.

39. Paroubek P, Chaudiron S and Hirschman L. Principles of evaluation in natural language processing. Traitement Automatique des Langues 2007; 48(1): 7–31.

40. Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomedical Informatics 2017; 73: 14–29. [PubMed: 28729030]

41. Kimia A Document Review Tools, 2019, http://www.dокументreviewtools.com/ (accessed 1.28.2019 2019).

42. Friedl JE. Mastering regular expressions. Sebastopol, CA: O’Reilly Media, Inc, 2002.
43. Lebanon G, Mao Y and Dillon J. The Locally Weighted Bag of Words Framework for Document Representation. JMLR 2007; 8: 2405–2441.

44. Bui DD and Zeng-Treitler Q. Learning regular expressions for clinical text classification. J Am Med Inform Assoc 2014; 21(5): 850–857. [PubMed: 24578357]

45. Ho TK. Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition, Montreal, QC, 1995, pp. 278–282.

46. Odom B, Lowe L and Yates C. Peripheral Infiltration and Extravasation Injury Methodology: A Retrospective Study. J Infusion Nursing : The Official Publication Infusion Nurses Soc 2018; 41(4): 247–252. DOI: 10.1097/NAN.0000000000000287.

47. Blanchfield BB, Demehin AA, Cummings CT, et al. The Cost of Quality: An Academic Health Center’s Annual Costs for Its Quality and Patient Safety Infrastructure. The Joint Comm J Qual Patient Saf 2018; 44(10): 583–589.

48. Beauvais B, Richter JP and Kim FS. Doing well by doing good: Evaluating the influence of patient safety performance on hospital financial outcomes. Health Care Management Review 2019; 44(1): 2–9. [PubMed: 2844325]

49. Li Q, Melton K, Lingren T, et al. Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care. J Am Med Inform Assoc JAMIA 2014; 21(5): 776–784. DOI: 10.1136/amiajnl-2013-001914. [PubMed: 24401171]

50. Thacker SB. Public health surveillance and the prevention of injuries in sports: what gets measured gets done. J Athletic Training 2007; 42(2): 171.

51. Lyren A, Brilli R, Bird M, et al. Ohio children’s hospitals’ solutions for patient safety: a framework for pediatric patient safety improvement. J Healthc Qual 2013; 38: 213–222.

52. Lyren A, Brilli RJ, Zieker K, et al. Children’s Hospitals’ Solutions for Patient Safety Collaborative Impact on Hospital-Acquired Harm. Pediatrics 2017; 140(3): e20163494.

53. Erickson SM, Wolcott J, Corrigan JM, et al. Patient safety: achieving a new standard for care. Washington, DC: National Academies Press, 2003.

54. CDC. Overview of Influenza Surveillance in the United States. Atlanta, GA: CDC, 2019. https://www.cdc.gov/flu/weekly/overview.htm (accessed 1/9/2019 2019).

55. Vincent C, Burnett S and Carthey J. Safety measurement and monitoring in healthcare: a framework to guide clinical teams and healthcare organisations in maintaining safety. BMJ Quality Safety 2014; 23(8): 670–677.

56. HealthIT.gov. Health IT Dashboard Quick Stats, 2018, https://dashboard.healthit.gov/quickstats/quickstats.php (accessed 8/1/2018).

57. Chassin MR, Loeb JM, Schmaltz SP, et al. Accountability measures--using measurement to promote quality improvement. The New Engl Journal Medicine 2010; 363(7): 683–688.

58. Schmaltz SP, Williams SC, Chassin MR, et al. Hospital performance trends on national quality measures and the association with Joint Commission accreditation. J Hosp Med 2011; 6(8): 454–461. [PubMed: 21990175]

59. Ortiz E, Meyer G and Burstin H. Clinical Informatics and Patient Safety at the Agency for Healthcare Research and Quality. J Am Med Inform Assoc : JAMIA 2002; 9(6 Suppl 1): s2–s7.

60. AHRQ. AHRQ Announces Interest in Research on Health IT Safety, 2016, https://grants.nih.gov/grants/guide/notice-files/NOT-HS-16-009.html (accessed 5/13/2019).

61. Hoffman JM, Keeling NJ, Forrest CB, et al. Priorities for Pediatric Patient Safety Research. Pediatrics 2019; 143(2): e20180496. [PubMed: 30674609]
Figure 1.
Surveillance workflow begins with data extraction and development of a training set with one or more iterations of SVM and manual review, followed by train-test and evaluation of performance on the full corpus of available data.
Figure 2.
DrT screen shot, showing ease of manually reviewing notes once RegEx is highlighted.
Figure 3.
Area under the curve showing performance of random forest model.
Table 1.

Regular expressions used to identify infiltrate events.

| Expression     | Examples                                         | Clinical instance from notes                           | Weight | Freq (#) | Notes (#) |
|----------------|--------------------------------------------------|--------------------------------------------------------|--------|----------|-----------|
| infiltrat\w*   | Infiltrate, infiltrated, infiltration            | Infiltrate in left arm                                 | 3096   | 321      | 245       |
| \bswell\w*\b   | Swelling                                         | R Leg swelling at baseline                             | 85     | 154      | 135       |
| \bblanch\w*\b  | Blanching, blanchable, blanches, blanched        | 24g PIV placed in R hand, blanches with flush          | 84     | 150      | 142       |
| Xtrav          | Extravasation                                    | Extravasation measures 4.5 cm by 6 cm on the right foot. | 51     | 2        | 1         |
| \bpalpa\w*\b   | Palpation, palpable, palpate, palpated           | Tender to palpation but pt easily distractable         | 51     | 224      | 200       |
| \bpivie\b      | PIVIE                                            | PIVIE swelling 78.6% of extremity.                     | 48     | 18       | 14        |
| \belevat\w*\b  | Elevated, elevating, elevation, elevate           | Non pitting edema to feet bilaterally legs elevated    | 44     | 483      | 441       |
| \bbli\w*\b     | Blister, blisters, blistering, blistered         | Skin was tight but without leaking or blistering.      | 36     | 41       | 36        |
| \bpiv\b        | piv, PIV                                         | Called to 11S MICU to place a PIV on this patient      | 25     | 3616     | 2906      |
| \bphl\w*\b     | Phlebitis, phlebitic                             | Suggest warm pack for phlebitis, arm elevated.         | 23     | 57       | 50        |