Automated processing of electronic medical records is a reliable method of determining aspirin use in populations at risk for cardiovascular events

Serguei VS Pakhomov PhD
Associate Professor, Department of Pharmaceutical Care and Health Systems, University of Minnesota, Minnesota, USA

Nilay D Shah PhD
Assistant Professor, Division of Health Care Policy and Research and Knowledge and Encounter Research Unit, Mayo Clinic, Rochester, Minnesota, USA

Penny Hanson
Research Assistant, Division of Endocrinology, Diabetes, Metabolism and Nutrition, Mayo Clinic, Rochester, Minnesota, USA

Saranya C Balasubramaniam BS
Research Assistant, Division of Health Care Policy and Research, Mayo Clinic, Rochester, Minnesota, USA

Steven A Smith MD
Associate Professor, Knowledge and Encounter Research Unit and Division of Endocrinology, Diabetes, Metabolism and Nutrition, Mayo Clinic, Rochester, Minnesota, USA

ABSTRACT

Background Low-dose aspirin reduces cardiovascular risk; however, monitoring over-the-counter medication use relies on the time-consuming and costly manual review of medical records. Our objective is to validate natural language processing (NLP) of the electronic medical record (EMR) for extracting medication exposure and contraindication information.

Methods The text of EMRs for 499 patients with type 2 diabetes was searched using NLP for evidence of aspirin use and its contraindications. The results were compared to a standardised manual records review.

Results Of the 499 patients, 351 (70%) were using aspirin and 148 (30%) were not, according to manual review. NLP correctly identified 346 of the 351 aspirin-positive and 134 of the 148 aspirin-negative patients, indicating a sensitivity of 99% (95% CI 97–100) and specificity of 91% (95% CI 88–97). Of the 148 aspirin-negative patients, 66 (45%) had contraindications and 82 (55%) did not, according to manual review. NLP search for contraindications correctly identified 61 of the 66 patients with contraindications and 58 of the 82 patients without, yielding a sensitivity of 92% (95% CI 84–97) and a specificity of 71% (95% CI 60–80).

Conclusions NLP of the EMR is accurate in ascertaining documented aspirin use and could potentially be used for epidemiological research as a source of cardiovascular risk factor information.

Keywords: aspirin, natural language processing (NLP), quality measurement
Introduction

Diabetes and heart disease are among the leading causes of mortality and morbidity in the United States for which a rich set of national quality of care standards has been developed. Included in these standards are the performance of regular foot examinations, and advising patients on aspirin use and smoking cessation, all of which are critical to reducing the risk of severe complications including ulcers, amputations, myocardial infarctions and heart failure. The assessment of these important processes is a central element in performance measurement and disease management.

National surveys assessing quality have shown very low rates of aspirin use in diabetes patients at high risk, resulting in a lost opportunity to reduce cardiovascular morbidity and mortality at low cost. Generally, performance measures have most often been assessed by using manual charting. However, this manual process is expensive and subject to lack of internal consistency with poor interrater reliability. With increasing adoption of the EMR healthcare providers have turned to structured data within the EMR in order to expedite clinical performance measurement. However, for certain measures (e.g. prescribing medications for specific indications), methods based on structured EMRs are of questionable validity compared to the information reported in the unstructured, free-text part of the EMR. In addition, methods based on automatic examination of the structured part of the EMR for medication prescription are limited in their ability to capture the situations where there are legitimate reasons (e.g. contraindications) for not prescribing a medication otherwise required by the treatment guidelines. The information necessary for a comprehensive and accurate assessment of compliance with quality and safety measures is contained in the unstructured text part of the EMR and requires specialised tools for its extraction and aggregation. In the context of the EMR, natural language processing (NLP) offers a promising method of automating the collection of a rich set of information for quality improvement and safety that would otherwise require manual chart abstraction. In the context of medication use and pharmaco-surveillance, a number of significant advances have been made in automatic identification of disease–drug associations, structured medication event information in discharge summaries and other types of textual clinical reports. The main objective of this study was to determine the validity of NLP for identification from the EMR of documentation indicative of aspirin use and potential contraindications for its use.

Methods

Participants

Eligible patients were 1) seen in one of six primary care practices at the Mayo Clinic; 2) participated in an employer-funded health plan; 3) had the diagnosis of diabetes prior to 1999, and 4) had given prior research authorisation. The Mayo Clinic is a large academic medical centre in Rochester, Olmsted County, Minnesota, USA which provides primary care to local residents, including over 6000 patients with diabetes.

Clinical reports

Mayo Clinic physicians have been documenting patient encounters electronically since 1994. Currently, clinical notes are dictated and stored in electronic format – the Mayo Clinic EMR. This comprises a dataset of over 25 million inpatient and outpatient notes. These notes are in compliance with the American National Standards Institute Clinical Document Architecture, a widely accepted standard for clinical documentation. All clinical notes (12,563 for the year 2006) were available for analysis for all but four of 519 eligible patients (final study population 515).

Reference standards

We compiled three reference standards in stages – one for aspirin exposure, one for cardiovascular vs non-cardiovascular use and one for contraindications to aspirin use. All three reference standards were obtained by manual review of clinical notes dictated in the year 2006 for the participants in the study. The records of all 515 participants were examined to determine if the record contained evidence that the participant was using aspirin. The reference standard for cardiovascular use was obtained by manual examination of the records of those participants for whom there was evidence of aspirin use. Finally, the reference standard for contraindications was obtained by examining only the records that did not contain evidence of aspirin use. Manual examination was performed by a standardised records reviewer (SB). Interrater agreement was determined on a random sample of 50 medical records independently reviewed by two reviewers (SB and PH), with disagreements subsequently resolved by one of the study investigators (NS). We used the National Guideline Clearinghouse as a standard for...
contraindications to aspirin use. These criteria included allergy to aspirin, certain gastrointestinal disorders (e.g. peptic ulcer, bleeding), intracranial bleeding, uncontrolled hypertension and anticoagulant and non-steroidal anti-inflammatory drug (NSAID) therapy. A list of NSAIDs was compiled, using the Federal Drug Administration (FDA)\(^b\) for prescription and the American College of Gastroenterology\(^c\) for non-prescription preparations.

**Automatic text search**

The clinical notes for each patient were searched electronically using a list of key terms representative of generic and brand names for aspirin, dosage information and contraindications. The details of the term lists and the actual algorithm implemented in the Perl programming language are shown in the Appendix and are also available from the corresponding author on request. The algorithms for text processing and the content of the electronic queries were developed solely on the basis of the guidelines that were used for manual classification of the records. No text of the clinical notes was used to define the queries or to design the algorithms. Furthermore, the investigator who did the programming was blinded to the manual classification, while the investigator who did the manual classification and created the reference standard was blinded to the algorithm design and implementation.

**Statistical analysis**

Sensitivity and specificity, standard operating characteristics for binary classification tasks, were used to assess the accuracy of the automated algorithms. Sensitivity is the ratio of correctly classified positive examples to the total number of positive examples in the reference standard. Specificity is the ratio of correctly classified negative examples to the total number of negative examples in the reference standard. A 95% confidence interval was calculated for each estimate based on binomial error distribution. Kappa statistic was used to account for random effects in estimating the interrater agreement.\(^{29}\)

\(^{b}\)www.fda.gov/CDER/Drug/infopage/COX2/NSAIDmedguide.pdf
\(^{c}\)www.gi.org/patients/women/asprin.asp

**Results**

**Manual classification**

**Aspirin exposure**

Of the 515 participants’ records in this study, evidence of aspirin use was unknown for 16 (2%) due to absence of clinical notes for these patients in year 2006. Excluding these participants from further analysis, 351 (70%) of the remaining 499 participants were identified as having evidence of aspirin use, while 148 (31%) did not have such evidence. Patients on aspirin were more likely to be male (\(p=0.0027\)) and elderly (mean age: 58.8 vs 53.2, \(p<0.0001\)).

**Cardiovascular use**

Of the 351 participants with evidence of aspirin use in their record, only 4 (1%) used aspirin for non-cardiovascular purposes (e.g. pain management).

**Contraindications**

Of the 148 participants without evidence of aspirin use in their records, 82 (55%) had no contraindications. Of the remaining 66 (45%) participants with contraindications, 40 (26%) were using NSAIDs, 13 (9%) were on Coumadin, three (2%) on Plavix, one had a diagnosis of peptic ulcer, one had a history of intracranial bleeding, two had a history of gastrointestinal bleeding, two were noted as allergic to aspirin and four had notes in their record advising against the use of NSAIDs for other reasons. The latter category included one participant with a note stating ‘history of papillary necrosis from NSAIDs’, one with ‘NSAID induced gastritis’, one with ‘chronic renal insufficiency – patient cannot use NSAIDs including aspirin’ and one with ‘chronic renal disease – patient advised to stop taking aspirin’.

**Interrater reliability**

The random sample of 50 medical records reviewed by two reviewers contained 500 observations. The interrater agreement measured by the Kappa statistic was 0.93 (95% CI 0.86–1.00). Of the 50 records reviewed, a total of ten had at least one instance of aspirin mention on which the reviewers disagreed. The disagreements were resolved by a study investigator (NS) and the corrected classifications for these ten patients were used in the reference standard.

**Compliance with aspirin use guidelines**

Having 351 participants identified as using aspirin and 66 identified as having contraindications results in 417
out of 499 people (84%). This demonstrates a high level of compliance with guidelines for aspirin and contraindications for its use in this population. A standard chart audit that does not take contraindications into account would have resulted in identifying 351 out of 499 (70%) people, showing a 14% lower level of compliance.

**Automatic text search**

Automatic classification of medical records with NLP compared to the manual reference standard had a sensitivity of 99% (95% CI 97–100; see Table 1). Five records were misclassified as having evidence of aspirin use while no such evidence was found during manual abstraction. Fourteen records were misclassified as containing no evidence of aspirin use resulting in a specificity of 92% (95% CI 88–97).

**Identification of CV use**

All four of the four records with evidence of non-cardiovascular use of aspirin were correctly identified (sensitivity of 100% (95% CI 55–100)); however, the algorithm also produced six false positives resulting in the specificity of 98% (95% CI 96–99).

**Identification of contraindications**

NLP correctly identified 61 out of 66 records with one or more potential contraindications to aspirin use resulting in a sensitivity of 92% (95% CI 84–97). Of the 82 records with no apparent contraindications, 58 were identified correctly by the automated system resulting in the specificity of 71% (95% CI 60–80).

We also applied the NLP mechanism for identifying contraindications to the 351 records that were manually classified as being on aspirin to find that 175, or 49% (95% CI 44–54), of these contained one or more mentions of a potential contraindication. Further manual analysis performed by one of the study investigators (SS) showed that 164 (94%, 95% CI 90–97) of the 175 patients indeed had at least one potentially relevant contraindication including 24 (14%) patients on Coumadin, 19 (11%) on Plavix, 65 (37%) on an as-needed dose of an NSAID, 65 (37%) on a regular dose of NSAID, 13 (7%) with a history of gastrointestinal bleeding and three (2%) with another comorbidity (e.g. renal failure, hypertension). At least half of these 164 patients had two or more (up to five) potential contraindications.

**Compliance with aspirin use guidelines**

The correct identification by the NLP of 346 participants as using aspirin and 61 identified as having contraindications results in 405 out of 499 people (81%), a similar compliance to the manually determined level of 84%.

**Discussion**

We demonstrated that automated NLP methodology for searching the text of the EMR has the potential to offer highly accurate results not only in identifying whether the patient record contains evidence of aspirin use, but whether such use is for cardiovascular or non-cardiovascular purposes and whether the record contains evidence of potential contraindications to

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Sensitivity and specificity of the automatic text classifier as compared to a manual reference standard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference standard</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>On aspirin</td>
<td></td>
</tr>
<tr>
<td>CV use</td>
<td></td>
</tr>
<tr>
<td>Contraindications</td>
<td></td>
</tr>
</tbody>
</table>
These findings demonstrate that the text of the EMR provides a valuable source of information on documentation of medication exposure, particularly for over-the-counter medications for which no other documentation exists either in the pharmacy or in administrative claims databases. Our findings indicate that the automated text processing methodology is as good as manual audit (with only a 3% difference) in determining the level of compliance with guidelines for aspirin use. A standard manual chart audit would have yielded a lower compliance level, underestimating the true quality of care delivery and at the same time involving much more effort and cost. Thus our methodology can be considered another strategy to add to the toolbox for quality improvement.

Similar to our study, previous reports using NLP methodology to detect reliable identification of evidence for foot examinations and tobacco cessation counselling demonstrate the increasing value of EMRs as an alternative to manual records review for difficult to measure data elements required for continuous quality improvement and reporting. Persell et al have demonstrated the value of using the text of clinical notes in reporting on ACE-inhibitor prescribing for patients with heart failure. Similarly, our study provides additional evidence pointing in the same direction, where almost half (45%) of the people with type 2 diabetes who are not using aspirin have documented potential contraindications.

Our current findings also demonstrate that NLP can be used for reliable ascertainment of contraindications for aspirin use. Our automated NLP algorithm identified at least one contraindication for 49% of people on aspirin with half of these people having two or more contraindications. Clearly, these results are indicative rather than conclusive because therapeutic decisions are made based on the values and preferences of individual patients guided by their provider. However, dissemination of NLP-derived reports that include potential contraindications may help encourage patient-provider communication, elucidate potential safety concerns and facilitate risk management for an individual patient.

While it is unlikely that local and national benchmarking organisations will immediately adopt NLP as a technique for documentation of quality measures, nevertheless it is efficient and scalable and can easily provide immediate value for internal reporting of quality and safety metrics by a health system. While the NLP system for determining potential contraindications for aspirin use is highly sensitive (61 of 66 records) it is less specific (24 of 82 records were false positives). However, even if all 85 (61 plus 24) records automatically classified as having potential contraindications required verification by the patient or the healthcare team, this would constitute only 16% of the entire volume of 499 records. Thus our methodology could be used for a dramatic reduction of the workload, making quality assurance and safety monitoring at the individual patient level much more feasible.

Error analysis

To gain a better understanding of the generalisability of our approach, we conducted an informal error analysis by manually examining the records of misclassified patients in each of the three categories: aspirin use, cardiovascular use and contraindications. Of the five records that were manually determined to have evidence of aspirin use and were misclassified by the automated classifier (false negatives), one instance was due to a manual abstraction error, one was due to a misspelling of aspirin as ‘asprin’ and three were due to formatting unanticipated by the algorithm used to query the text. Of the 11 records labelled by the automatic classifier as containing evidence of aspirin use but manually classified as not containing such evidence (false positives), six were due to an error in manual abstraction and the rest were due to a failure in the identification of cues indicative of aspirin mention in a probable or negated context (e.g. ‘Will continue on Plavix as patient’s preference rather than aspirin’).

The examination of misclassification errors in cardiovascular vs non-cardiovascular use revealed that five of the six false positives were due to a manual abstraction error. The latter resulted from examining records with large numbers of notes with predominantly cardiovascular use of aspirin but with one or two notes mentioning additional use of aspirin for pain management. One of the six false positives was truly ambiguous: ‘I recommended to continue Ecotrin daily aspirin and discontinue all other Bayer prn products’. In this example, both daily and pro re nata (as needed) usages of aspirin are mentioned in the same sentence.

Analysis of the errors resulting from automatic identification of contraindications showed that three of the five false negatives were due to a manual abstraction error. For example, one of the records was manually classified as having a diagnosis of gastrointestinal bleeding in one of the notes; however, upon second examination of this patient’s record, a mention of abdominal wall cellulitis and a ruling out of gastrointestinal bleeding were found in a subsequent note. The other two false negatives were due to reasons for non-use of aspirin other than potential contraindications established by the national standards. For example, one of the records referred to NSAID induced gastritis and the other to renal insufficiency as reasons why the treating physician recommended that aspirin should not be used. These examples are rare in our dataset. Overall, our automated NLP method is highly
effective in identifying common contraindications. The examination of the 18 false positives showed that four were due to manual abstraction error, eight were due to a mention of an over-the-counter NSAID in an ‘as needed’ context and six were due to a failure by the regular expression mechanism to identify negation (e.g. ‘He remains off aspirin and NSAID therapy’).

Limitations and strengths

Several limitations must be mentioned to facilitate interpretation of the results. First, the text of the EMR can be used only as a surrogate to identify patient exposure to over-the-counter medications. Arguably, medication reconciliation may be a more complete way to ascertain medication exposure; however, known concerns exist with using patient self-report for medication information due to imperfect recall, particularly for over-the-counter medications. Ideally, a variety of information sources including patient self-report, pharmacy transaction records and the text of the EMR should be used to create an accurate representation of over-the-counter medication use by patients. Second, consistent with national and regional quality reporting and benchmarking efforts and for the sake of feasibility, we limited the manual examination of the text of clinical reports to a single year. While the Mayo Clinic EMR tends to have continuity where important facts from previous visits typically carry over into the documentation of the subsequent visit, this may not be the case with other EMR systems, limiting the generalisability of our findings. This is particularly true of information on contraindications where, for example, a history of intracranial bleeding may date back several years. For continuous EMRs, the scalability of the technique we have described offers a distinct advantage for analysing longitudinal data on large number of patients. Going forward, it will be important to demonstrate the generalisability of the methodology described in this paper to EMR systems other than the one used at the Mayo Clinic.

Our study also has a number of unique strengths. The sample size used for testing the methodology is large (>499 records with 12 563 notes) relative to other studies of automated text processing methods. While using enrollees in an employer-funded health plan limits the generalisability of our findings, it is also a strength as this population receives its care primarily at a single institution and thus represents the model scenario in terms of homogeneity and completeness of the medical records. Finally, the Mayo Clinic EMR is a state of the art system that is fully compliant with national standards for clinical documentation (HL7 CDA) that have also been adopted by most major EMR vendors in the USA.

ACKNOWLEDGEMENTS

We would like to thank Sara Hobbs Kohrt for assistance with manuscript preparation.

REFERENCES

Automated processing of EMRs


FUNDING
Funded by a Mayo Clinic Quality Innovation Program Award.

CONFLICTS OF INTEREST
None.

ADDRESS FOR CORRESPONDENCE
Nilay D Shah
Mayo Clinic
200 First Street SW
Rochester MN 55905
USA
Tel: +1 507.266.5130
Fax: +1 507.284.1731
Email: shah.nilay@mayo.edu

Accepted May 2010
Appendix

Figure 1 Perl regular expressions definitions
Figure 2 Pseudo-code illustrating the search algorithm based on regular expressions