Identification of Inactive Medications in Narrative Medical Text

Eugene M. Breydo[a], PhD, Julia T. Chu[b], Alexander Turchin, MD, MS[a,b,c]

[a]Clinical Informatics Research and Development, Partners HealthCare, Boston, MA
[b]Brigham and Women's Hospital, Boston, MA
[c]Harvard Medical School, Boston, MA

Abstract

Discontinued medications are frequently not removed from EMR medication lists - a patient safety risk. We developed an algorithm to identify inactive medications using in the text of narrative notes in the EMR.

The algorithm was evaluated against manual review of 297 randomly selected notes. One in five notes documented inactive medications. Sensitivity and precision of 87.7% and 80.7%, respectively, on per-note basis and 66.3% and 80.0%, respectively, on per-medication basis. When medication names missing from the dictionary were excluded, the algorithm achieved sensitivity of 91.4%. Using real clinical data, the algorithm identified inactive medications documented in the note but still listed as active on the patient's medication list in more than one in ten notes.

Documentation of inactive medications is common in narrative provider notes and can be computationally extracted. This technology could be employed in real-time patient care as well as for research and quality of care monitoring.

Introduction

Accurate medication information at the point of care is crucial for delivery of high-quality care and prevention of adverse events[1]. Inappropriate administration of medications the patient no longer takes has a particularly high potential for adverse drug events. However, a number of reports have shown that errors of this type are common in electronic medical records[2, 3]. While correct medication regimen can be established through reconciliation of the electronic medical records with other information sources (e.g. patients, insurance claims or pharmacies)[4], these sources may not be available or may be inaccurate themselves[5, 6].

Narrative physician notes are a rich but untapped source of medication information[7]. Physicians are expected to document all patient care they provide, including discontinuation of any medications, in their notes. In recent years many successful natural language processing (NLP) packages have been developed that can accurately identify problems, active medications and other information in narrative medical documents[8-13]. However, the reported accuracy of identification of inactive medications has been low[14]. We have previously established that documentation of inactive medications in narrative documents is expressed in a lexically constrained fashion[15]. We therefore carried out an assessment of feasibility of identification of discontinued medications through analysis of the text of provider notes in the electronic medical record.

Materials and Methods

Algorithm

The aim of the algorithm was to identify documentation of inactive medications in narrative medical documents. Inactive medications were defined as all medications that patient had taken at some point in the past but would not be taking by the end of the visit described in the note. Both medications discontinued before and during the visit documented in the note were included. Medications that were documented only as a part of the allergy list were excluded. Supplements and other substances not regulated by the Food and Drug Administration were excluded as well.

The algorithm was implemented in Caché (Intersystems, Cambridge, MA) – a language commonly used in other healthcare applications. The main components of the algorithm are described in detail below.

I. Identification of Medications

In order to identify medications recorded in the document under analysis the algorithm compares every word in the document to the list of medication names in the internally developed and maintained Master Drug Dictionary (MDD). MDD contains three types of medication names: generic names, brand names and synonyms. The category of synonyms includes words empirically found to be frequently used by the users of the internal electronic medical record system when searching for a medication. Many of the words in this category are homographs (have multiple meanings) and their meaning that represents a medication is not always the one most commonly encountered in narrative medical documents. Therefore in order to increase the algorithm's specificity we empirically removed a number of terms that were felt to more commonly refer to concepts other than medications (e.g. "MS", "thyroid", "fiber").

II. Identification of Inactive Medications

In order to determine which of the medications identified in the previous step were documented as inactive, the algorithm performs semantic analysis of the context of each word recognized as a medication. In the general case the context was defined empirically as a distance of 50 characters (where continuous whitespace characters count as a single character) from the beginning or the end of the medication word. However, depending on the specific semantic key identified in this space, the maximum permitted distance may be shortened. For example, based on our experience the maximum distance between the semantic key "discontinue" and the word representing the medication name was empirically set to 20 characters.

The algorithm analyzes the context space for presence of a semantic key that could potentially identify the medication as inactive. These keys fall into several semantic fields (Table 1). If a semantic key for an inactive medication is found in the context space, the medication is considered inactive if the following constraints are satisfied:

1. Medications section and relative position

In each of the documents the algorithm identifies the Medications section (typically a list starting with the word "Medications"). If a Medications section exists in the note and the medication word being analyzed is outside of the Medications section, the medication will only be considered discontinued if the medication word is located after the Medications section in the note.

2. Medications section and conditionality

If a Medications section exists and the medication word is located in the note after the Medications section, the medication is only considered discontinued if the semantic key indicating discontinuation is non-conditional (e.g. "discontinue", "stop"). If, on the other hand, the semantic key is conditional (e.g. "painful", "unsuccessfully", "not sufficiently"), the medication is not considered discontinued. Conditional semantic keys therefore only indicate that the medication is inactive if the medication is not listed in the Medications section or the Medications section does not exist in the document.

3. Stop List

The space between the semantic marker and the medication word is analyzed for the presence of a Stop List semantic key. These keys (e.g. period, "because", "continue", "added") typically indicate syntactic or semantic discontinuity between the inactivity semantic key and the medication word. If any of these keys are present between the inactivity semantic key and the medication word, the medication is not considered inactive.

Evaluation - Accuracy

We evaluated the accuracy of the software on a dataset of 297 outpatient notes by 270 unique providers. The notes were randomly selected from over eight million documents by providers of all specialties in the electronic medical record system of two large academic hospitals. In order to create a gold standard against which the algorithm was evaluated, each note was manually reviewed by a medical student who was specially trained for this task to identify inactive medications.

The software output was compared to the manual rating to determine sensitivity (recall), specificity and positive predictive value (precision) of the software at the note level and sensitivity and positive predictive value at the token (individual medication) level. Dictionary-based sensitivity was calculated by making the assumption that the dictionary we were using recognized all medication names in the narrative documents. Normal approximation was used to calculate confidence intervals for sensitivity, specificity, and positive predictive value.

Table 1 Semantic Fields for Documentation of Medication Discontinuation

Semantic Field	Examples
Stop	discontinue stop
Does not work	did not respond unsuccessfully
Past	had not used no longer on
Completed	completed finished
Change	change to <med2> switch from <med1> to <med2>

Evaluation – Use Case

In order to determine the usability of the algorithm in real-world clinical settings we analyzed a set of 1,000 outpatient provider notes randomly selected from a single year (2004) in the electronic medical record among patients who had at least one active medication in the structured medication database in the same electronic medical record system. Using audit data we identified all medication records that were active on the day the note was written and stayed active for at least another calendar day (thus excluding the medications that might have been inactivated in the electronic medical record immediately or soon after the note was written). We compared the inactive medications identified by the algorithm in the text of the notes to the medications listed as active (and not deactivated that day) in the structured medication record. We identified the medications that were documented in the note as discontinued (as recognized by the algorithm) but were left as "active" in the structured database.

IRB

The study protocol was reviewed and approved by Partners Human Research Committee.

Results

Manual review identified a total of 65 notes with 109 documented inactive medications among the 297 documents analyzed. We were able to identify 87% of the notes with documented inactive medications at the specificity of over 95% (Table 2). The software successfully recognized two thirds of the documented inactive medications with a positive predictive value of 80% (Table 3). One of the most common reasons for false negatives was absence of the medication name from the dictionary employed by the algorithm. Missing medication names belonged to several categories:

1. The medication name as recorded in the dictionary contained other information besides the name itself (e.g. "Advair 250/50").
2. The medication name used in the note referred to a medication class rather than a specific medication (e.g. "oral corticosteroids", "statins")
3. An abbreviation was used in the note instead of the full medication name (e.g. "d/c of pred shortly")
4. An acronym referring to a combination of medications was used in the note (e.g. CHOP = cyclophosphamide + hydroxydoxorubicin + Oncovin + prednisone)

When false negatives that resulted from the absence of medication names from the dictionary were excluded, the sensitivity of the algorithm rose to 91%. A typical reason for the false negatives not due to the dictionary was that the appropriate semantic key had not been included in the algorithm (e.g. "the episode of bronchitis which she had in early October resolved with doxycycline therapy").

Table 2 Accuracy of Identification of Notes with Documented Inactive Medication

Sensitivity	87.7% (± 6.0%)
Specificity	95.2% (± 3.9%)
Positive Predictive Value	80.7% (±7.2%)

95% confidence interval is given in parentheses.

The most common scenario for incorrect identification of a medication as inactive by the software involved dynamic context where the information about the active status of a particular medication changed from one part of the note to another. For example, past discontinuation of the medication could be documented in the earlier part of the note (e.g. "… and he recently stopped both his Effexor and fluoxetine") followed by a statement that documented re-initiation of the medication ("… he will restart fluoxetine at the previous dose of 40 mg qd"). In this situation the algorithm in its current implementation detects and records the documentation of medication discontinuation without correcting for its subsequent restart in the later part of the note. Modifiers of the semantic keys used to detect documentation of medication discontinuation (e.g. "he is titrating himself off his OxyContin") also led to false positives.

Table 3 Accuracy of Identification of Documented Inactive Medications

Sensitivity	66.3% (± 8.7%)
Dictionary-based Sensitivity	91.4% (± 5.2%)
Positive Predictive Value	80.0% (± 7.3%)

95% confidence interval is given in parentheses.

In order to determine the applicability of our algorithm in the real-world clinical environment we analyzed 1,000 notes of patients with active medications on the structured medication list in the electronic medical record. The algorithm identified 510 instances of documentation of inactive medications in 256 notes with in this dataset. Inactive medications were most commonly documented by a statement that the medication had been or will be stopped or that the medication was not effective (Table 4). Structured medication list in the electronic medical record contained records for 181 of the medications documented in the notes as inactive. Of these, 122 (67.4%) medications were still active one calendar day after their inactive status was documented in the note.

Table 4 Distribution of Semantic Fields Identifying Inactive Medications

SemanticField	Frequency (N, %)
Stop	167 (32.7)
Does not work	156 (30.6)
Past	136 (26.7)
Completed	40 (7.8)
Change	11 (2.2)

Discussion

As the prevalence of chronic illness and the number of available pharmaceuticals grow, individual patients are taking increasing quantities of medications[16]. These medication regimens change frequently because of lack of efficacy, new indications, medication side effects, insurance requirements, etc. Not uncommonly medications of a single patient are managed by multiple physicians making coordination of the patients' treatment even more challenging[17].

Electronic medical records offer a way to facilitate inter-provider communication for many aspects of patient care, including medications[18]. In order to realize these benefits accurate and timely entry of information in the patient's record is essential. Unfortunately, reports indicate that this is not always the case[2]. This problem may be particularly acute for medications that the patient no longer takes. While physicians have an incentive to enter new medications into the electronic medical record to generate prescriptions[19], outdated medication entries are frequently not deleted and studies show that many electronic medication records may be out of date[3].

Providers are expected to document all changes in the patient's treatment plan, including changes in the medication regimen, in narrative notes. Many electronic medical record systems allow entry of notes in digital format, making possible computational identification of medication information from the text of the notes. Until now most text analysis tools have focused on identification of active medications and reported accuracy rates for identification of inactive medications have been low[14]. In this paper we describe the first, to our knowledge, algorithm that focuses on identification of documentation of inactive medications in narrative medical text.

The algorithm has two main structural components that determine its accuracy: medication name dictionary and the set of semantic fields used to recognize documentation of medication discontinuation. We used the standard internal medication dictionary with minimal modifications and have found that it was not optimized for this task. Some medication names in the dictionary were not in the same form as was usually used by providers in the text of the notes. Additionally, the dictionary did not contain many medication name abbreviations / misspellings, names of medication classes or defined medication combinations, such as chemotherapy regimens. We estimate that a comprehensive dictionary would result in c. 50% increase in sensitivity of our algorithm.

In order to be accepted by clinicians, the algorithm must have high positive predictive value. We have found that a common reason for false positives was the need for integrative semantic analysis of the entire note rather than a single spatially co-localized set of tokens. The algorithm already implements this approach by considering all semantic keys in the context of the Medications section in the note (if present). Further expansion of the integrative techniques will be one of the directions for the future development of the algorithm.

This algorithm could have several applications. If used in real time, it could be employed to alert providers that their note documented discontinued medications still listed as active on the structured medication list and prompt them to inactivate the outdated medication entries. An alert like this could be used in conjunction with a training program to encourage providers to delete inactive medications from the electronic medical record. In fact, we found that in our electronic medical record system more than one in ten notes contain information about discontinued medications that are still listed as active on the structured medication list. That fraction could further increase with the improved sensitivity of the algorithm.

If used retrospectively, the algorithm could provide information on inactive medications documented in the notes on multiple patients at a time. These large scale data could be used for either quality control or research. For example, administration of a healthcare organization could use the algorithm to estimate the fraction of active medications in its electronic medical record that patients were actually no longer taking. This information could help the organization determine the scale of the problem and the amount of resources that should be spent to address it. A researcher studying the reasons why medications are discontinued in real-life clinical practice could use the algorithm to identify physician notes that discuss medication discontinuation instead of looking for them manually.

Our study has a number of strengths. To our knowledge, it is the first report on design of an algorithm dedicated to identification of inactive medications – an important constituent of the strategy to minimize medication errors. The algorithm has been validated on the notes created by a large group of providers drawn from both primary care physicians and specialists at two institutions, increasing its applicability in other settings. Finally, the algorithm was tested in a real-life clinical environment and successfully identified inactive medications still listed as active on the patients' medication lists.

Our study had several limitations. It was conducted on the data from two academic medical centers in eastern Massachusetts which could limit its applicability. The sensitivity of the algorithm was lower than expected. However, as discussed above, this could be relatively easily enhanced by using a more comprehensive medication name dictionary.

Conclusion

We have designed and tested a novel algorithm for detection of documentation of inactive medications in the narrative medical documents. We have evaluated the algorithm's accuracy and have determined the potential approaches for its further improvement. We have validated the algorithm on a sample of real-life clinical data and demonstrated that discontinued medications are commonly documented in the notes, even as they are at the same time still listed as active on the structured medication lists in the electronic medical record. The algorithm can potentially be used for both clinical, quality control and research applications.

References

1. Cornish PL, Knowles SR, Marchesano R, et al. Unintended medication discrepancies at the time of hospital admission. *Arch Intern Med.* 2005;165(4):424-429.
2. Manley HJ, Drayer DK, McClaran M, Bender W, Muther RS. Drug record discrepancies in an outpatient electronic medical record: frequency, type, and potential impact on patient care at a hemodialysis center. *Pharmacotherapy.* 2003;23(2):231-239.
3. Turchin A, Gandhi TK, Coley CM, Shubina M, Broverman C. The use of electronic medication reconciliation to establish the predictors of validity of computerized medication records. *Medinfo.* 2007;12(Pt 2):1022-1026.
4. Pronovost P, Weast B, Schwarz M, et al. Medication reconciliation: a practical tool to reduce the risk of medication errors. *J Crit Care.* 2003;18(4):201-205.
5. Vilke GM, Marino A, Iskander J, Chan TC. Emergency department patient knowledge of medications. *J Emerg Med.* 2000;19(4):327-330.
6. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. *J Clin Pharm Ther.* 2001;26(5):331-342.
7. Chen ES, Stetson PD, Lussier YA, Markatou M, Hripcsak G, Friedman C. Detection of Practice Pattern Trends through Natural Language Processing of Clinical Narratives and Biomedical Literature. *Proc AMIA Annu Fall Symp.* 2007:120-124.
8. Friedman C, Hripcsak G, Shagina L, Liu H. Representing Information in Patient Reports Using Natural Language Processing and the Extensible Markup Language. *J Am Med Inform Assoc.* 1999;6(1):76-87.
9. Hahn U, Romacker M, Schulz S. MEDSYNDIKATE - a natural language system for the extraction of medical information from findings reports. *Int J Med Inform.* 2002;67(1-3):63-74.
10. Heinze DT, Morsch ML, Holbrook J. Mining free-text medical records. *Proc AMIA Symp.* 2001:254-258.
11. Hazlehurst B, Frost HR, Sittig DF, Stevens VJ. MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. *J Am Med Inform Assoc.* 2005;12(5):517-529.
12. Meystre S, Haug PJ. Automation of a problem list using natural language processing. *BMC Medical Informatics and Decision Making.* 2005;5(30).
13. Savova GK, Ogren PV, Duffy PH, Buntrock JD, Chute CG. Mayo clinic NLP system for patient smoking status identification. *J Am Med Inform Assoc.* 2008;15(1):25-28.
14. Pakhomov SV, Ruggieri A, Chute CG. Maximum entropy modeling for mining patient medication status from free text. *Proc AMIA Symp.* 2002:587-591.
15. Breydo EM, Chu JT, Turchin A. Identification of Inactive Medications in Narrative Physician Notes. *Proc AMIA Annu Fall Symp.* 2007:882.
16. Muir AJ, Sanders LL, Wilkinson WE, Schmader K. Reducing medication regimen complexity: a controlled trial. *J Gen Intern Med.* 2001;16(2):77-82.
17. Hajjar ER, Hanlon JT, Sloane RJ, et al. Unnecessary drug use in frail older people at hospital discharge. *J Am Geriatr Soc.* 2005;53(9):1518-1523.
18. Hillestad R, Bigelow J, Bower A, et al. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Aff (Millwood).* 2005;24(5):1103-1117.
19. Tamblyn R, Huang A, Kawasumi Y, et al. The development and evaluation of an integrated electronic prescribing and drug management system for primary care. *J Am Med Inform Assoc.* 2006;13(2):148-159.