ERROR MANAGEMENT

Using trigger phrases to detect adverse drug reactions in ambulatory care notes

Michael N Cantor, Henry J Feldman, Marc M Triola

Qual Saf Health Care 2007;16:132–134. doi: 10.1136/qshc.2006.020073

See end of article for authors' affiliations

Correspondence to:
Dr M N Cantor, New York
University School of
Medicine, 423 East 23rd
Street, #15N-167, New
York, NY 10010, USA;
michael.cantor@bellevue.
nychhc.org

Accepted 19 November
2006

Background: As medical care moves towards an outpatient focus, monitoring systems for ambulatory patients are increasingly important. Because adverse outcomes due to medications are an important problem in outpatients, the authors developed an automated monitoring system for detecting adverse drug reactions (ADRs) in ambulatory patients.

Methods: The authors obtained a set of approximately 110 000 ambulatory care notes from the medicine clinic at Bellevue Hospital Centre for 2003–4, and manually analysed a representative sample of 1250 notes to obtain a gold standard. To detect ADRs in the text of electronic ambulatory notes, the authors used a "trigger phrases" methodology, based on a simple grammar populated with a limited set of keywords.

Results: Under current functionality, this system detected 38 of 54 cases in the authors' gold standard set, of which 17 were true positives, for a sensitivity of 31%, a specificity of 98%, and a positive predictive value of 45%. Their proxy measure correlated with 70% of the ADRs in the gold standard. These values are comparable or superior to other systems described in the literature.

Conclusions: These results show that an automated system can detect ADRs with moderate sensitivity and high specificity, and has the potential to serve as the basis for a larger scale reporting system.

The World Health Organization defines an adverse drug reaction (ADR) as "a response to a medicine which is noxious and unintended, and which occurs at doses normally used in man."¹ In Naranjo, *et al*'s model² for differentiating between ADRs and adverse drug events (ADEs), or injuries due to drugs, ADEs can be thought of as a subset of ADRs. In this model, if a relation between an adverse event and a drug is suspected and plausible, an ADE can be assumed. If causality is established between the ADE and a specific drug, then an ADR can then be assumed.

Outpatient ADRs have been estimated to contribute to 3.2–6.5% of hospital admissions³ and, in one study, to approximately 4% of hospital readmissions.⁶ Among all patients, it is estimated that 5–10% experience an ADR, and a little less than 1% of the general population is sent to the hospital because of ADRs.⁷ Studies of ADRs at academic medical centres have shown that admissions for ADRs had higher charges and significantly higher length of stay than the average medical-surgical admission.^{6, 8}

ADRs are particularly important for pharmacovigilance, or post-marketing surveillance of new or established drugs. Most drugs approved by the FDA average 1500 patient exposures,⁹ so rare and potentially serious side effects often emerge as drugs are prescribed to the general population. Detecting a new type of ADR in a prescription medication may actually lead to action on a larger scale. For example, a single case report started the investigation that led to terfenadine's removal from the market.⁹

Most published surveillance systems either focus on ADEs, rather than ADRs. For example, in the outpatient setting, Honigman, *et al*¹⁰ created an extensive computer monitoring system that used a specialised, commercial medical lexicon in addition to ICD-9 codes, new allergy information, and computer "trigger" rules similar to those described in Jha *et al*¹¹ to detect ADEs. The combined search methods had a sensitivity of 58%, a specificity of 88%, and a positive predictive value of 7.5%. Of the different methods, text searching discovered the largest number of incidents, but also had a

relatively low positive predictive value (7%). Other systems used natural language processing to analyse discharge summaries^{12, 13}; however, we have not found other published examples of ADR surveillance systems that use textual data directly from outpatients. A computerised inpatient ADR detection system using a commercially available drug database reported a sensitivity of 47.5%, specificity of 1.6%, and PPV of 1.8%.¹⁴ The computerised system detected mainly mild ADRs.

We believe that the system used in this study will be useful for patient safety monitoring, and can be effective for monitoring patient tolerance of and adherence to prescription medications. Proving the merit of our approach with this study was a requirement prior to the development of a more complex, wide-ranging surveillance system. We hypothesised that our approach would perform comparably to or better than current systems, and that the tradeoff in performance would be offset by the huge time savings when compared with manual review. The importance of a surveillance system is that it presents a framework for the systematic evaluation of events and data gathering, rather than relying on anecdotal methods such as voluntary reporting,¹⁵ where underreporting of ADRs is the norm.¹⁶

METHODS

Data used

Bellevue Hospital, the nation's oldest public hospital and part of New York City's Health and Hospitals Corporation, uses Misys Computerised Patient Records¹⁷ for its electronic health record system, so documentation in outpatient medicine clinics is completely online. There are approximately 54 000 visits to the outpatient medicine clinic each year, comprising approximately 23 000 unique patients. After approval by the Health and Hospitals Corporation and the New York University Medical School's institutional review boards, we extracted the anonymised free-text of the narrative portion of two years of clinical notes directly from our electronic health record. The free-text

Abbreviations: ADE, adverse drug event; ADR, adverse drug reaction

Table 1 Lexicon of key words

Discontinuation	Causes	Unique phrases
D/C	Because [b/c]	Stopped taking
Discontinued	Due	Not [n't] tolerate
Not [been] tak[ing]	Second[ary] [to]	Unable to tolerate
Stopped		
Switched		

sections have no structured fields or controlled vocabulary requirements.

“Trigger phrase” methodology

In looking for ADRs in outpatients, we modified the traditional inpatient trigger methodology,^{18, 19} substituting patterns in ambulatory care notes for laboratory values and medications. We hypothesised that discontinuation of or non-compliance with a medication would correlate with the presence of an ADR, and used this framework to develop the key words and combinations that we would search for using the automated system.

We used a relatively simple approach to processing the text contained within our corpus of ambulatory care notes, focusing on grammatical structures that defined patients' actions. The basic grammar consisted of Verb|Adverb clause, each with a limited but generalisable lexicon. During preliminary analysis of our algorithm, we found certain unique phrases that appeared to have a high correlation with the presence of ADRs, and added those to the lexicon as well. We looked for patterns in text that matched this grammar, such as “stopped | due to” (the full lexicon can be found in table 1), or contained the unique phrases. To account for the possibility of the adverb clause being used in a separate context in a free-text note, we arbitrarily limited the distance between the verb and adverb clause to 50 characters.

Gold standard

Due to the number of visits in our database, it was impractical to have a human reviewer look at each note. To find a representative sample for our gold standard, we randomly selected 1250 unique visits from the database. Two physician reviewers (MNC, MMT) then each read the notes from all 1250 visits to determine if an ADR was present, and the severity (*significant*, *serious*, *life-threatening*, *fatal*)²⁰ of the ADR. In cases of disagreement, a third reviewer (HJF) acted as a referee.

RESULTS

The two years of data we obtained comprised approximately 103 000 visits to the Bellevue medicine clinics. The set of 1250 reviewed notes represented unique visits by 596 patients. Because of the anonymisation process we were unable to obtain demographic or provider information for specific patients or visits, or the number of active prescriptions at a specific point in time. The median number of prescriptions written per visit over the two year period for the 1250 notes

reviewed (1, range 0–10), however, was equal to the median number written for the entire data set (1, range 0–36).

Within our set of 1250 reviewed notes the review process resulted in 54 unique cases where ADRs were present, for an event rate of 4.3%. If extrapolated to a year of clinic visits, the event rate would translate to 2150 events/year. Interrater reliability for the case reviews before the referee's decision was substantial ($\kappa = 0.69$).

The automated system detected 38 unique ADRs, of which 17 were true positives. Examples of ADRs found by reviewers and the algorithm can be found in table 2. Under its current functionality, our system has a sensitivity of 31% and a specificity of 98%, with a PPV of 45%. Of note, the system analysed the 1250 notes in approximately 3 seconds.

As expected in outpatients, most ADRs were of *significant* severity. The reviewers rated five ADRs as *life-threatening*, 15 as *serious* and 34 as *significant*. Among the 17 ADRs detected by the algorithm, 13 were significant and four were serious. The algorithm did not detect any of the ADRs that were judged to be life-threatening by the reviewers. Of note, medication discontinuation was noted in four out of five of these cases; however, due to the variety with which an intervention may be documented in free-text, the algorithm was unable to detect these events.

Performance of proxy measures

Medication discontinuation correlated well as a proxy measure for ADRs. Of the 54 cases found by reviewers, 35 explicitly mentioned medication discontinuation and three implied the process, for a total of 70% of the cases. Other general categories of actions that correlated with ADRs included non-compliance with prescribed regimens (that is, taking lower doses or with lesser frequency than prescribed), changing doses (reducing ACE inhibitor dose due to hyperkalemia), and continuing medications due to a favourable risk–benefit analysis (that is, insulin in patients with hypoglycaemic episodes). Several notes also had no specific documentation of actions taken concerning the ADRs. Table 3 gives a breakdown of medications and corresponding symptoms by category.

DISCUSSION

Our approach to detecting ADRs had a sensitivity comparable to other detection systems, and a PPV that was comparable to or superior to other, more complex systems. Because of the unstructured nature of the corpus of notes we analysed, as well as the variety of ways in which ADRs may be represented, automated detection is difficult under these circumstances. Using changes in patient behaviours as a proxy for ADRs is an indirect method of measurement, but in the case of this system performs well.

Our system did not detect any ADRs rated as life-threatening by the physician reviewers, though the proxy measure of medication discontinuation would have detected the large majority. This deficiency in our results is illuminating on several fronts. First, it reveals the challenge of analysing free-text documentation; the variety of ways ADRs may be documented requires an algorithm with broad scope.

Table 2 Sample adverse drug reactions

Algorithm—true positives	False negatives
“has not been taking norvasc 5 because it gives her strong headaches”	“has taken hydralazine and has developed rash ... Medication was stopped today”
“started on terazosin but unable to tolerate it because of dry mouth”	“monopril 2.5 (low dose as h/o ARF secondary to ACE in 12/98)”
“ace d/ced sec to hyperkalemia”	“Insomnia is probably secondary to SSRI. Will continue for now”
“zocor 20 (stopped taking due to feeling tired)”	“(FS 30) after juice. Did not take glyburide or Actos since then”

Table 3 Medications and symptoms by category

Endocrine	16	Hypoglycaemia, diarrhoea, psychiatric effects
Cardiovascular	15	Hyperkalemia, dizziness, cough
Pain	7	Gastrointestinal symptoms
Immune	4	Psychiatric, rash, gastrointestinal
Psychiatric	3	Psychiatric, fatigue, insomnia
Neurological	3	Gastrointestinal
Statins	2	Musculoskeletal
GI	2	Gastrointestinal
Diuretics	2	Neurological, genitourinary

Similarly, it shows the need to improve our algorithm to encompass the potential representations of ADRs, either directly or through proxy measures. Finally, the result underlies the utility of medication discontinuation as a proxy measure for ADRs, as discontinuation was reported in the notes for 80% of the most severe events.

As with any automated system, ours is dependent on the quality of data it receives. In this case, the system is dependent on both physician documentation and patient reporting of ADRs during a clinic visit. With this in mind, we chose to focus on the text of the note for two reasons: first, because it is the most challenging from a research perspective, second, because the operating characteristics of these methods are relatively well established, and last, because previous research has shown that analysis of other areas of the chart (ICD-9 codes, new allergy information) do not contribute significantly more to detection.¹⁰

One major limitation to our system is the fact that its use necessitates the presence of an electronic health record, and that only a minority of providers actually use an electronic system for documentation. Existing tools and methodologies facilitate the analysis of traditional triggering events, but analysis of paper progress notes on a large scale would be extremely resource intensive. As technology advances and the standard of medical care changes, one would expect a higher penetrance of electronic health records and thus a wider applicability of a system such as ours.

Ideally, an automated detection system would be able to detect trends in ADRs early, using active short-term surveillance. With improved sensitivity, a system like ours could provide the basis for such a system. Because of its quick response time (3 seconds for 1250 notes), the system could analyse free-text notes in real time to attempt to analyse ADRs and to report on trends. With additional decision support, the system could prompt the provider—if he has not already done so—to change medications based on the probable ADR information contained in the clinical note.

CONCLUSIONS

Our automated system for detecting ADRs performed comparably to existing systems, with moderate sensitivity and positive predictive value and high specificity. Our system was successful at detecting medication discontinuation and other changes to medication regimens, both of which correlated well with ADRs. Because our relatively straightforward methods for detecting ADRs performed comparably to existing systems, we expect that planned improvements to the algorithm, including expanding the lexicon to include abbreviations and non-standard text; more complex grammatical rules; and implementing improved negation detection should improve performance. Improved clinical utility should come through linking symptoms to specific medications and automated classification of ADR severity.

We plan to use this system to populate an ADR database to be used for local pharmacovigilance and other surveillance activities. Detecting adverse outcomes in the outpatient setting will become increasingly important as the bulk of patient care transitions away from the inpatient setting. Detecting patient use patterns with certain drugs, or certain unexpected reactions to newer drugs, is especially important in a large integrated delivery network such as the Health and Hospitals Corporation, where shared formularies mean ADR trends may apply to larger populations, and where single interventions may have better chances of averting future ADRs.

Authors' affiliations

M N Cantor, New York University School of Medicine, New York, New York, USA

H J Feldman, Beth Israel-Deaconess Medical Center, Boston, Massachusetts, USA

M M Triola, New York University School of Medicine, New York, New York, USA

Supported by the Department of Medicine, New York University School of Medicine, and grant 2002-DT-CX-K002 from the US Department of Homeland Security.

REFERENCES

- WHO. Safety of medicines. http://whqlibdoc.who.int/hq/2002/WHO_EDM_QSM_2002.2.pdf (accessed November 2005).
- Hanlon JT, Maher RL, Lindblad CL, et al. Comparison of methods for detecting potential adverse drug events in frail elderly inpatients and outpatients. *Am J Health Syst Pharm* 2001;**58**:1622–6.
- Naranjo CA, Shear NH, Lantot KL. Advances in the diagnosis of adverse drug reactions. *J Clin Pharmacol* 1992;**32**:897–904.
- Camargo AL, Cardoso Ferreira MB, Heineck I. Adverse drug reactions: a cohort study in internal medicine units at a university hospital. *Eur J Clin Pharmacol* 2006;**62**:143–9.
- van der Hooft CS, Sturkenboom MC, van Groothoest K, et al. Adverse drug reaction-related hospitalisations: a nationwide study in The Netherlands. *Drug Saf* 2006;**29**:161–8.
- Dormann H, Neubert A, Criegee-Rieck M, et al. Readmissions and adverse drug reactions in internal medicine: the economic impact. *J Intern Med* 2004;**255**:653–63.
- Moore N, Lecomte D, Noblet C, et al. Frequency and cost of serious adverse drug reactions in a department of general medicine. *Br J Clin Pharmacol* 1998;**45**:301–8.
- Wu WK, Pantaleo N. Evaluation of outpatient adverse drug reactions leading to hospitalization. *Am J Health Syst Pharm* 2003;**60**:253–9.
- US Food and Drug Administration. Preventable adverse drug reactions: a focus on drug interactions. Available at <http://www.fda.gov/cder/drug/drugReactions/default.htm#ADRs> (accessed June 2006).
- Honigman B, Lee J, Rothschild J, et al. Using computerized data to identify adverse drug events in outpatients. *J Am Med Inform Assoc* 2001;**8**:254–66.
- Jha AK, Kuperman GJ, Teich JM, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J Am Med Inform Assoc* 1998;**5**:305–14.
- Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc* 2005;**12**:448–57.
- Murff HJ, Forster AJ, Peterson JF, et al. Electronically screening discharge summaries for adverse medical events. *J Am Med Inform Assoc* 2003;**10**:339–50.
- Egger T, Dormann H, Ahne G, et al. Identification of adverse drug reactions in geriatric inpatients using a computerised drug database. *Drugs Aging* 2003;**20**:769–76.
- Kilbridge PM, Classen DC. Automated surveillance for adverse events in hospitalized patients: back to the future. *Qual Saf Health Care* 2006;**15**:148–9.
- Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf* 2006;**29**:385–96.
- Misys. CPR - Computerized patient record and electronic health record software. Available at <http://www.misyshealthcare.com/Products/Product+Portfolio/Misys+CPR/default.htm> (accessed October 2006).
- Szekendi MK, Sullivan C, Bobb A, et al. Active surveillance using electronic triggers to detect adverse events in hospitalized patients. *Qual Saf Health Care* 2006;**15**:184–90.
- Classen DC, Pestotnik SL, Evans RS, et al. Computerized surveillance of adverse drug events in hospital patients. *JAMA* 1991;**266**:2847–51.
- Gurwitz JH, Field TS, Harrold LR, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *JAMA* 2003;**289**:1107–16.