

Research Paper ■

Automated Detection of Adverse Events Using Natural Language Processing of Discharge Summaries

GENEVIEVE B. MELTON, MD, GEORGE HRIPCSAK, MD, MS

Abstract **Objective:** To determine whether natural language processing (NLP) can effectively detect adverse events defined in the New York Patient Occurrence Reporting and Tracking System (NYPORTS) using discharge summaries.

Design: An adverse event detection system for discharge summaries using the NLP system MedLEE was constructed to identify 45 NYPORTS event types. The system was first applied to a random sample of 1,000 manually reviewed charts. The system then processed all inpatient cases with electronic discharge summaries for two years. All system-identified events were reviewed, and performance was compared with traditional reporting.

Measurements: System sensitivity, specificity, and predictive value, with manual review serving as the gold standard.

Results: The system correctly identified 16 of 65 events in 1,000 charts. Of 57,452 total electronic discharge summaries, the system identified 1,590 events in 1,461 cases, and manual review verified 704 events in 652 cases, resulting in an overall sensitivity of 0.28 (95% confidence interval [CI]: 0.17–0.42), specificity of 0.985 (CI: 0.984–0.986), and positive predictive value of 0.45 (CI: 0.42–0.47) for detecting cases with events and an average specificity of 0.9996 (CI: 0.9996–0.9997) per event type. Traditional event reporting detected 322 events during the period (sensitivity 0.09), of which the system identified 110 as well as 594 additional events missed by traditional methods.

Conclusion: NLP is an effective technique for detecting a broad range of adverse events in text documents and outperformed traditional and previous automated adverse event detection methods.

■ *J Am Med Inform Assoc.* 2005;12:448–457. DOI 10.1197/jamia.M1794.

Adverse event prevention and detection are national health care priorities.¹ Detection of adverse events represents an opportunity to learn from events via a cognitive perspective so that inciting factors surrounding events can be identified and improved.² Voluntary reporting of adverse events by most institutions, however, remains a largely unsuccessful prac-

tice.^{3–7} While chart review is effective,⁸ it is too costly for routine use.

Health care policy makers and practitioners will need information technology coupled with improved data collection to improve patient safety.⁹ Computerized systems for event detection rely on signals suggestive of adverse events both in the case of impending events (for prevention) and of events that have occurred (for management).¹⁰ For example, a discharge diagnosis of myocardial infarction in a patient with an unrelated surgical admission diagnosis might indicate an adverse event. Event detection systems reduce the cost of chart review by identifying those cases that are most appropriate for review.⁶ Successful systems require sufficient positive predictive value to avoid needless chart review and sufficient sensitivity to gather a meaningful number of events.

Most adverse event detection systems exploit numeric or coded data derived from patient registration, pharmacy orders, admission and discharge diagnoses, clinical laboratory results, and ancillary information systems.^{11–16} Investigators have studied adverse event detection from the perspective of adverse drug events, dangerous laboratory values, failure to follow critical paths, and other events. Although these adverse detection systems often perform well, they are limited because they require clinical data that are in coded format.

Unfortunately, most institutions lack a detailed record of their patients' care in coded electronic format. Symptoms, physical findings, and clinical reasoning are recorded as narrative text in notes but are unavailable in coded form. The lack of coded

Affiliations of the authors: Department of Biomedical Informatics, Columbia University (GBM, GH); and Medical Informatics Services, NewYork-Presbyterian Hospital (GH), New York, NY.

Supported by grants from the Agency for Healthcare Research and Quality (R18 HS11806) "Mining Complex Clinical Data for Patient Safety Research" and National Library of Medicine (R01 LM06910) "Discovering and Applying Knowledge in Clinical Databases." Dr. Melton was supported by the National Library of Medicine Training Grant (5T15LM007079-12).

The authors thank Carol Friedman for the use of the natural language processor MedLEE (National Library of Medicine grant support R01 LM06274 and R01 LM07659), Sue West for her assistance with institutional NYPORTS reporting, and Karina Tulipano for serving as a case reviewer.

Correspondence and reprints: George Hripcsak, MD, MS, Department of Biomedical Informatics, Columbia University, 622 West 168th Street, Vanderbilt Clinic, 5th Floor, New York, NY 10032; e-mail: <hripcsak@columbia.edu>.

Received for publication: 01/19/05; accepted for publication: 03/20/05.

information limits the performance of event detection systems and limits the breadth of events that they can detect.

Narrative clinical notes such as discharge summaries, operative reports, clinic notes, and nursing notes are increasingly available in electronic form either through transcription or direct data entry. Investigators have begun to exploit these documents for event detection by looking for notes with relevant words ("trigger words") such as "iatrogenic," "error," or "perforation."^{17,18} This technique helps, but its predictive value remains low, largely because it is difficult to distinguish whether a clinician is saying that a condition is present, is absent, or was present in the past. Natural language processing is an automated technique that converts narrative documents into a coded form that is appropriate for computer-based analysis.

Natural language processing has been used successfully for several specific domains of medicine^{19–24} and for the detection of specific adverse events, such as falls and nosocomial infections.^{10,25} It is unclear, however, whether natural language processors can detect a wide range of complex adverse events accurately enough to assist health care institutions meaningfully. In this study, we built an event detection system for electronic discharge summaries using an existing, noncommercial natural language processor, MedLEE,²⁶ in an effort to detect a broad range of adverse events.

Background

The natural language processor MedLEE employs a vocabulary and a grammar to extract information from narrative text. MedLEE was initially developed to process radiographic reports²¹ but has been expanded to process a wide range of medical texts.²⁶ MedLEE also handles negation (denial), uncertainty, timing, synonyms, and abbreviations. For example, the sentence "The patient may have a history of MI" is coded as follows:

problem: myocardial infarction
certainty: moderate
status: past history

The certainty and status fields indicate that the diagnosis is unsure ("moderate" certainty) and that if the myocardial infarction did occur, it occurred in the past. A detailed overview of MedLEE has been published.^{21,26}

The New York Patient Occurrence Reporting and Tracking System (NYPORTS) is a mandatory adverse event reporting framework instituted in 1996 for all health care institutions in New York State.²⁷ We used the criteria for each of the 45 patient-related hospital-based adverse event types defined in NYPORTS (Appendix 1); they represent a broad range of adverse events.

Many NYPORTS adverse event types are complex. For example, NYPORTS event type 751 includes falls in the hospital resulting in an x-ray-proven fracture, a subdural or epidural hematoma, cerebral contusion, traumatic subarachnoid hemorrhage, or internal organ trauma. The event type excludes falls that occur outside of the institution or that result in only soft tissue injuries. NYPORTS event type 604 includes perioperative myocardial infarction within 48 hours of an operative procedure. The procedure must not be cardiac related, birth related, an abdominal aortic aneurysm rupture, or a multiple trauma.

Methods

We developed and tested our adverse event system at NewYork-Presbyterian Hospital–Columbia University Medical Center, an urban, tertiary health care institution. There were 107,305 inpatient cases for the years 1996 and 2000. The target population of the study comprised all 57,452 inpatient cases at our institution with electronic discharge summaries during this period.

The adverse event detection system²⁸ comprised the MedLEE natural language processor^{21,26} and a set of criteria that mapped each MedLEE-coded discharge summary to the adverse events that occurred during the admission. The inclusion and exclusion criteria for each event were implemented as a computer query, which is a short program that includes logic and terms from MedLEE's vocabulary. MedLEE converted each discharge summary to a coded form, and the 45 computer queries converted that coded form to a list of events that appeared to have occurred during each admission. The computer queries were developed iteratively; we tested them on discharge summaries from the years 1990 to 1995 (before implementation of NYPORTS), modified the queries to improve performance, and retested them on the cohort.

System Evaluation

Manual chart review served as the gold standard. We assessed the reliability of the reviewers on 100 cases as follows. Two reviewers, a physician coauthor (GBM) and an informatician independent of this study, identified NYPORTS events in 100 cases selected randomly but stratified so that about 40% had events. The reviewers' raw agreement was 0.97, and chance-corrected agreement (kappa) was 0.94. This high agreement justified the use of a single reviewer per case.

Reliability of the data sources was assessed on 1,000 randomly selected cases in which the physician identified NYPORTS events using (1) the discharge summaries alone, (2) the full electronic chart, and (3) for a subset of 100, the combined electronic and paper charts. Electronic charts included discharge summaries, operative reports, pathology reports, laboratory results, radiology results, registration data including coded diagnoses and procedures, residents' transfer of service notes, and other ancillary notes, but they contained few admission notes, progress notes, or nursing notes. The paper chart supplied the latter missing notes. We calculated the agreement among the three data sources.

Performance of the system was assessed with the same 1,000 random cases from 1996 and 2000 used for the full data reliability dataset. These cases were used to obtain an unbiased and direct estimate of sensitivity and specificity of the system for identifying cases that had NYPORTS events. The system identified apparent events based on discharge summaries. The physician manually reviewed the electronic chart for each case and determined which NYPORTS events had clearly occurred in the case.

System performance was then assessed using all electronic discharge summaries from 1996 and 2000 to get a more precise estimate of the positive predictive value and performance on individual event types. The physician reviewed those discharge summaries that the system identified as having events. An identification was considered correct only if the system selected the correct NYPORTS event type.

Finally, to assess how the system might work in practice, we compared the events that were detected by the system and confirmed by the physician reviewer with the events that were actually detected during those years using traditional event detection techniques. In 1996 and 2000, hospital personnel reported candidate NYPORTS events in one of three ways: (1) direct phone calls from practitioners, patients, and other hospital personnel; (2) incident reports from practitioners; and (3) report forms completed by case management personnel in conjunction with utilization review. Hospital personnel then determined the veracity of candidate NYPORTS events by manual screening of the electronic chart and, if needed, the paper chart.

The institutional review board approved the study and waived informed consent for this retrospective review.

Results

Data Reliability

In the 100 cases with both electronic chart review and combined paper-electronic chart review, there was complete agreement on all 39 events. This high agreement justified the use of electronic charts as the gold standard for the 1,000 case set. Manual review of discharge summaries agreed with manual review of the electronic chart in all but five of 1,000 cases, resulting in a raw agreement of 0.995 and kappa of 0.96. This high agreement demonstrates that discharge summaries contain most of the information needed to detect NYPORTS adverse events, so a system based on discharge summaries has the potential for accurate identification.

System Performance on 1,000 Cases

Table 1 shows the performance of the system for detecting cases with at least one adverse event, based on the 1,000 case set. "True events" are those identified by manual review of the electronic chart, and "apparent events" are those identified by the system. The system correctly identified 15 of 53 cases with events. Table 2 shows the performance of the system for detecting individual events, based on the 1,000 case set. The system correctly identified 16 of 65 true events and incorrectly identified 49. Event specificity (0.9996 in Table 2) exceeds case specificity (0.982 in Table 1) because case specificity is subject to the sum of the false-positive rates of all the event types, whereas event specificity represents the average specificity expected for an investigator interested in a single NYPORTS event type.

System Performance on Full Cohort of 57,452 Cases

Table 3 shows the number of events that the system identified in the full cohort of 57,452 cases, the number of those events that manual review verified, and the overall positive predictive value for the system calculated by case and by event. Appendix 1 contains the positive predictive value of the system by each event type. In sum, the system identified 1,590 events in 1,461 cases, and manual review verified 704 of the events in 652 cases.

"Best Estimate" System Performance

Table 4 summarizes the event prevalence and "best estimates" of system performance using a combination of the data obtained from the 1,000 case set and the full cohort of 57,452 cases. The specificity for specific event types can also be estimated from Appendix 1. The range is 0.998 (95% con-

Table 1 ■ Automated Adverse Event Detection System Versus Manual Review for 1,000 Charts, Aggregated by Case

	Automated Detection System		
	Cases with Apparent Events	Cases without Apparent Events	Total
Manual review			
Cases with true events	15	38	53
Cases without true events	17	930	947
Total	32	968	1,000
value (95% CI)			
Sensitivity	0.28 (0.17–0.42)		
Specificity	0.982 (0.971–0.990)		
Positive predictive value	0.47 (0.29–0.65)		
Negative predictive value	0.96 (0.95–0.97)		

CI = confidence interval.

Table 2 ■ Automated Adverse Event Detection System Versus Manual Review for 1,000 Charts, Aggregated by Event

	Value (95% CI)
Events identified by manual review	65
Events identified by the system	32
Events identified by the system and verified by manual review	16
Sensitivity	0.25 (0.15–0.37)
Specificity	0.9996 (0.9994–0.9998)
Positive predictive value	0.50 (0.32–0.68)
Negative predictive value	0.9989 (0.9986–0.9992)

CI = confidence interval.

fidence interval [CI]: 0.997–0.998) for event 803 to 1 (95% CI: 0.9999–1.0) for event 852.

System Performance Compared with Traditional Reporting

The last two columns of Appendix 1 tally traditional NYPORTS detection and its overlap with the automated system. Table 5 compares traditional detection to the automated system followed by manual verification. The sensitivity of traditional detection can be approximated as 322 of 3,734 ($0.065 \times 57,452$) or about 0.086. The system identified 110 of 322 traditionally detected events (0.34; 95% CI: 0.29–0.40). The system identified 594 events that were missed by traditional detection methods, increasing the total number of events detected from 322 for traditional detection alone to 916 using a combined approach.

Discussion

In this study at a large, tertiary care medical center, our automated adverse event detection system with natural language processing achieved excellent performance and provided effective screening for NYPORTS adverse events contained within a large corpus of electronic discharge summaries over a two-year period. The sensitivity to detect events was only fair (0.25 by event and 0.28 by case) but far higher than that found for traditional reporting in this study

Table 3 ■ System Performance with 57,452 Electronic Discharge Summaries Aggregated by Case and by Event

	Value by case (95% CI)	Value by event (95% CI)
Identified by the system	1,461	1,590
Identified by the system and verified by manual review	652	704
Positive predictive value	0.45 (0.42–0.47)	0.44 (0.42–0.47)

CI = confidence interval.

(0.086) or in previous studies.^{3–7} The system achieved very high specificity.

The current system, when compared with other adverse event detection systems using text documents, is unique in its ability to both recognize a broad range of events and identify the specific event type in each case. Thus, it enables highly focused manual review to detect a significant fraction of events at minimal cost.

Most previous studies of automated adverse event detection from narrative documents used simple text search techniques and achieved limited success. In two studies of adverse drug event detection in the outpatient setting using automated text searching in clinic notes, the text search method performed well compared with other automated methods but achieved positive predictive values of only 7%¹³ and 12%.⁵ In a different study, text searching in discharge summaries, residents' transfer of service notes, and outpatient visit notes using the search terms "mistake," "error," "incorrect," and "iatrogenic" to find medical errors identified a broad range of medical errors and had positive predictive values ranging from 3.4% to 24.4%.¹⁷ The system did not distinguish among the event types, however, and its sensitivity was less than 4%. In a study of text searching on discharge summaries to identify a broad range of events, the system returned 59% of discharge summaries with a predictive value of 52%.¹⁸ Because the prevalence of these nonspecific events in the underlying sample was 45%, however, the predictive value was only moderately higher than would be achieved by random sampling. Our system identified specific event types, with average prevalence per event type of less than 1%, and it still achieved a positive predictive value of 44% per event.

In addition, a recent report by Forster et al.²⁹ described the validation of an adverse event detection instrument for discharge summaries using term searching. In contrast to the current study, which contains a direct reliability study, that report used an established instrument. The authors reported a positive predictive value of 0.41, a sensitivity of 0.23, and a specificity of 0.92. The predictive value of 0.41 must be interpreted in light of the high underlying prevalence of adverse events, which was 20% (48 of 245) in the reported case sample using a broad definition of adverse events. In addition to achieving a comparable predictive value with rare and specific events, our system achieved a better specificity and identified the exact event type.

Our reliability studies, which were conducted to verify the rater and data sources, revealed that NYPORTS events were straightforward for clinicians to identify with manual review

and that discharge summaries contain most NYPORTS adverse events. Although the raters had little difficulty with manual review, query development for these events was a long and intricate task for system developers. Queries were developed in an iterative manner with many rounds often necessary to decrease both false negatives and false positives. Because of the large amount of complexity surrounding these adverse event definitions with respect to inclusion and exclusion criteria, however, mimicking the natural reasoning of a clinician within an automated query was difficult.

For example, an area being actively investigated by others,³⁰ which was particularly difficult in this project, was reasoning with respect to time. While MedLEE does have some time representations for dates and other simple time structures, its current capabilities with respect to these issues are limited. Certain time reasoning could be inferred, such as an event occurring after another event using collocation information in the text. Many other time-reasoning issues, however, were not easily modeled in the queries. For instance, five postoperative NYPORTS events require that the event occurred *within* 48 hours of the procedure (events 601 to 605, see [Appendix 1](#)). Modeling a time difference of 48 hours with the coded data from MedLEE was difficult. The addition of other data sources, in addition to other text documents, to augment the system could potentially improve time reasoning as well as improve overall data modeling for the event detection system.

Although the system was successful in detecting NYPORTS events, there are important adverse event types that the NYPORTS structure does not include or sometimes explicitly excludes. For instance, the NYPORTS adverse event criteria for iatrogenic pneumothorax include solely those pneumothoraces due to an intravascular catheter and exclude other iatrogenic causes, including thoracentesis or lung biopsy. For this reason, the system would need modification if the goal were to obtain all possible adverse events of potential interest.

While the overall performance of the system was excellent compared with that of other text-processing adverse event detection systems, system performance at the event or query level varied somewhat by event type. Many event types had a low event prevalence ([Appendix 1](#)), so the performance for individual event types could not be determined accurately. Nevertheless, certain queries were more difficult to implement in an automated fashion than others, resulting in variable system performance. Another central issue, in addition to issues with time reasoning, was handling event criteria not typically contained explicitly in the discharge summary. This required indirect modeling in the query (e.g., the use of conscious sedation was indirectly modeled by detecting procedures that typically use conscious sedation). The addition of other data sources could potentially enhance system performance by directly supplying this inferred information.

One potential source of bias in this study was that only patients with electronic discharge summaries were included. Patients who stayed less than 48 hours did not require a discharge summary, and sometimes summaries were simply missing from the record. This group may have had a different event rate than those included in the study.

An important aspect of this technology is its straightforward transferability to other institutions. Previous experience using

Table 4 ■ “Best Estimate” Event Prevalence and System Performance

Metric*	Derivation	Value (95% CI)
Prevalence		
Case rate: proportion of cases with one or more true events	$53 \div 1,000$	0.053 (0.040–0.069)
Event rate: true events per case	$65 \div 1,000$	0.065 (0.051–0.082)
System performance for detecting cases with events		
Sensitivity: proportion of cases with true events that had apparent events	$15 \div 53$	0.28 (0.17–0.42)
Specificity: proportion of cases with no true events that had no apparent events	$1 - \frac{1461-652}{(1-0.053) \times 57,452}$	0.985 (0.984–0.986)†
Positive predictive value: proportion of cases with apparent events that had true events	$652 \div 1,461$	0.45 (0.42–0.47)
Negative predictive value: proportion of cases with no apparent events that had no true events	$930 \div 968$	0.96 (0.95–0.97)
System performance for detecting individual events		
Sensitivity: proportion of true events that were identified by the system	$16 \div 65$	0.25 (0.15–0.37)
Specificity: proportion of cases without true events of a given type that the system did not identify	$1 - \frac{(1590-704) \div 45}{(1-0.065 \div 45) \times 57,452}$	0.9996 (0.9996–0.9997)†
Positive predictive value: proportion of apparent events that were true	$704 \div 1,590$	0.44 (0.42–0.47)
Negative predictive value: proportion of cases without true events of a given type that had no true event	$\frac{1,000 - (81 \div 45)}{1,000 - (32 \div 45)}$	0.9989 (0.9986–0.9992)

CI = confidence interval.

*A true event was detected by manual review; an apparent event was identified by the system.

†See the text for an explanation of the difference between case specificity and event specificity.

Table 5 ■ Automated Detection with Manual Verification Versus Traditional Detection

	Automated Detection System with Manual Verification		
	Event Detected	Event Missed	
Traditional detection			
Event detected	110	212	322
Event missed	594	*	
Total	704		

*The number of events missed by both systems is unknown but can be estimated as $0.065 \times 57,452 - (110 + 212 + 594) = 2,818$.

the MedLEE natural language processor at other institutions suggests that performance should be comparable and that adjusting the computer queries should reduce any loss of performance.³¹ For patients with electronic discharge summaries, the overhead of using the system should be minimal. There are minor formatting requirements, and standardized section headings are helpful but not mandatory. Transferability is limited in two ways: (1) not all patients have discharge summaries, typically due to short hospital stays or lack of clinician compliance, and (2) some institutions do not currently have discharge summaries in electronic form. The MedLEE natural language processing component can process a broad range of documents, and extending the adverse event detection system to progress notes, operative reports, consult notes, and ancillary reports would likely result in the detection of additional adverse events.

Moreover, system specificity is high enough to make nationwide screening feasible. For example, if electronic discharge summaries were available for all inpatients, then an investigator interested in wound dehiscence (event 805) could run the system on the 30 million admissions expected per year³²

and produce about 11,000 cases with about 11,000 false positives (from Appendix 1, event positive predictive value of 0.51 with approximately one case returned by the system for every 1,350 discharge summaries).

Natural language processing may revolutionize adverse event reporting and may play a significant role in adverse event prevention and other forms of intervention. The described system tripled the number of detected events without impeding or increasing the clinicians' workflow, as the operation of our system on discharge summaries was completely automated and transparent to clinicians. As health care moves from simple detection to actual intervention and prevention, the system may become even more important. Processing takes only about a second per document, and MedLEE processes documents at our institution as they are created. In contrast to retrospective manual detection and to voluntary reporting in which clinicians must know about and decide to report an event, natural language processing can provide immediate feedback to clinicians for issues of which they may be unaware. For example, MedLEE processing of chest radiograph reports reduced the rate of erroneously assigning patients with active tuberculosis to nonprivate rooms by almost one half.³³

Conclusion

Natural language processing was an effective method for automated adverse event detection, with the reported system outperforming traditional and previous automated adverse event detection methods. In contrast to previously reported techniques, the system detected a broad range of complex adverse events and identified the specific event type with high specificity, although only fair sensitivity. Ultimately, this study demonstrates the potential of natural language processing to facilitate health care processes. Automated diagnosis coding,

real-time clinical guidance, computer-assisted documentation, and improved clinical trial recruitment are some of the far-reaching applications of this important technique.

References ■

- Kohn LT, Corrigan JM, Donaldson MS (eds). *To err is human: building a safer health system*. Washington, DC: National Academy Press; 2000.
- Zapf D, Reason JT. Introduction to error handling. *Appl Psychol*. 1994;43:427–32.
- Benson M, Junger A, Fuchs C, Quinzio L, Bottger S, Jost A, et al. Using an anesthesia information management system to prove a deficit in voluntary reporting of adverse events in a quality assurance program. *J Clin Monit Comput*. 2000;16:211–7.
- Cullen DJ, Sweitzer BJ, Bates DW, Burdick E, Edmondson A, Leape LL. Preventable adverse drug events in hospitalized patients: a comparative study of intensive care and general care units. *Crit Care Med*. 1997;25:1289–97.
- Field TS, Gurwitz JH, Harrold LR, Rothschild JM, Debellis K, Seger AC, et al. Strategies for detecting adverse drug events among older persons in the ambulatory setting. *J Am Med Inform Assoc*. 2004;11:25–80.
- Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Rittenberg E, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and simulated voluntary report. *J Am Med Inform Assoc*. 1998;5:305–14.
- von Laue NC, Schwappach DL, Koeck CM. The epidemiology of medical errors: a review of the literature. *Wien Klin Wochenschr*. 2003;115:318–25.
- Brennan TA, Localio AR, Leape LL, Laird NM, Peterson L, Hiatt HH, et al. Identification of adverse events occurring during hospitalization. A cross-sectional study of litigation, quality assurance, and medical records at two teaching hospitals. *Ann Intern Med*. 1990;112:221–6.
- Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academy Press; 2001.
- Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. *J Am Med Inform Assoc*. 2003;10:115–28.
- Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. *JAMA*. 1995;274:29–34.
- Benson M, Junger A, Michel A, Sciuk G, Quinzio L, Marquardt K, et al. Comparison of manual and automated documentation of adverse events with an Anesthesia Information Management System (AIMS). *Stud Health Technol Inform*. 2000;77:925–9.
- Honigman B, Lee J, Rothschild J, Light P, Pulling RM, Yu T, 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, Rittenberg E, Teich JM, Bates DW. Identifying hospital admissions due to adverse drug events using a computer-based monitor. *Pharmacoepidemiol Drug Saf*. 2001;10:113–9.
- Murff HJ, Patel VL, Hripcsak G, Bates DW. Detecting adverse events for patient safety research: a review of current methodologies. *J Biomed Inform*. 2003;36:131–43.
- Samore MH, Evans RS, Lassen A, Gould P, Lloyd J, Gardner RM, et al. Surveillance of medical device-related hazards and adverse events in hospitalized patients. *JAMA*. 2004;291:325–34.
- Cao H, Stetson P, Hripcsak G. Assessing explicit error reporting in the narrative electronic medical record using keyword searching. *J Biomed Inform*. 2003;36:99–105.
- Murff HJ, Forster AJ, Peterson JF, Fiskio JM, Heiman HL, Bates DW. Electronically screening discharge summaries for adverse medical events. *J Am Med Inform Assoc*. 2003;10:339–50.
- Baud RH, Rassinoux AM, Wagner JC, Lovis C, Juge C, Alpay LL, et al. Representing clinical narratives using conceptual graphs. *Methods Inf Med*. 1995;34:176–86.
- Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc*. 2000;7:593–604.
- Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc*. 1994;1:161–74.
- Gundersen ML, Haug PJ, Pryor TA, van Bree R, Koehler S, Bauer K, et al. Development and evaluation of a computerized admission diagnoses encoding system. *Comput Biomed Res*. 1996;29:351–72.
- Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. *Ann Intern Med*. 1995;122:681–8.
- Sager N, Lyman M, Nhan NT, Tick LJ. Medical language processing: applications to patient data representation and automatic encoding. *Methods Inf Med*. 1995;34:140–6.
- Mendonca EA, Haas J, Shagina L, Larson E, Friedman C. Extracting information on pneumonia in infants using natural language processing of radiology reports. *J Biomed Inform*. In press.
- Friedman C. A broad-coverage natural language processing system. *Proc AMIA Symp*. 2000;270–4.
- Novello AC. *NYPORIS: The New York Patient Occurrences and Tracking System Annual Report 2000/2001*. Albany, NY: New York State Health Department; 2003.
- Hripcsak G, Bakken S, Stetson PD, Patel VL. Mining complex clinical data for patient safety research: a framework for event discovery. *J Biomed Inform*. 2003;36:120–30.
- Forster AJ, Andrade J, van Walraven C. Validation of a discharge summary term search method to detect adverse events. *J Am Med Inform Assoc*. 2005;12:200–6.
- Hripcsak G, Zhou L, Parsons S, Das AK, Johnson SB. Modeling electronic discharge summaries as a simple temporal constraint satisfaction problem. *J Am Med Inform Assoc*. 2005;12:55–63.
- Hripcsak G, Kuperman GJ, Friedman C. Extracting findings from narrative reports: software transferability and sources of physician disagreement. *Methods Inf Med*. 1998;37:1–7.
- MEPS HC-067D: 2002 Hospital inpatient stays. Rockville, MD: Agency for Healthcare Research and Quality; 2004.
- Knirsch CA, Jain NL, Pablos-Mendez A, Friedman C, Hripcsak G. Respiratory isolation of tuberculosis patients using clinical guidelines and an automated clinical decision support system. *Infect Control Hosp Epidemiol*. 1998;19:94–100.

Appendix 1 ■ Events Identified by the Automated Adverse Event Detection System and by Traditional Event Detection on 1,000 Cases and on 57,452 Cases

Adverse Event Type with Description	1,000 Case Set			57,452 Case Set					Proportion of Traditionally Detected Events Detected by the System
	Events Identified by Manual Review	Events Identified by the System	Events Identified by the System and Verified by Manual Review	Events Identified by the System	Events Identified by the System and Verified by Manual Review	Positive Predictive Value of the System	Events Identified by Traditional Event Detection	Events Identified by Both the System and by Traditional Event Detection	
803. Hemorrhage or hematoma requiring drainage, evacuation, or other procedural intervention*†	1	4	1	230	107	0.47	52	28	0.54
402. New documented deep venous thrombosis (excludes superficial thrombophlebitis)	9	6	5	204	105	0.51	28	10	0.37
819. Unplanned operation or return to reoperation (excludes nonanesthesia procedures, procedures commonly sequential or repeated)*†	8	2	0	172	80	0.47	54	18	0.33
808. Postoperative wound infection requiring drainage (excludes contaminated or dirty case)*†‡	9	2	1	137	47	0.34	70	8	0.11
801. Procedure-related injury requiring repair, removal of an organ, or other procedural intervention (excludes intended injuries based on disease process, anatomical structure, lack of alternative approach)*†	2	1	0	128	39	0.30	22	14	0.64
302. Intravascular catheter-related volume overload leading to pulmonary edema (excludes secondary to acute myocardial infarction or patients with predisposing conditions; congestive heart failure, cardiac disease, renal failure/insufficiency, hemodynamic instability, or critically ill)	3	1	0	96	54	0.56	0	0	—
401. New, acute pulmonary embolism, confirmed or suspected and treated (excludes suspected cause of sudden death with no autopsy to confirm)‡	2	1	1	95	61	0.64	12	5	0.42
938. Malfunction of equipment during treatment or diagnosis or a defective product with death or serious injury§	1	3	1	94	44	0.47	2	1	0.50
806. Displacement, migration, or breakage of an implant, device, graft, or drain*†	3	4	2	70	24	0.34	14	1	0.07
604. Perioperative/periprocedural acute myocardial infarction (excludes multiple trauma, abdominal aortic aneurysm rupture)*‡	3	3	1	66	26	0.39	9	3	0.33
605. Perioperative/periprocedural death (excludes multiple trauma, abdominal aortic aneurysm rupture)*‡	0	0	0	58	12	0.21	4	1	0.25

Continued

Appendix 1 ■ Continued

Adverse Event Type with Description	1,000 Case Set			57,452 Case Set					Proportion of Traditionally Detected Events Detected by the System
	Events Identified by Manual Review	Events Identified by the System	Events Identified by the System and Verified by Manual Review	Events Identified by the System	Events Identified by the System and Verified by Manual Review	Positive Predictive Value of the System	Events Identified by Traditional Event Detection	Events Identified by Both the System and by Traditional Event Detection	
805. Wound dehiscence requiring repair*‡	1	4	1	43	22	0.51	8	5	0.63
301. Intravascular catheter-related necrosis or infection requiring repair, regardless of location of repair (excludes exclusive treatment with packs, intravenous catheter change, medications, wound irrigation)	0	0	0	29	10	0.34	5	4	0.80
603. Perioperative/periprocedural cardiac arrest with successful resuscitation (excludes intentional arrest during cardiopulmonary procedures, multiple trauma, abdominal aortic aneurysm rupture)*‡	0	2	0	28	6	0.21	0	0	—
751. Falls resulting in x-ray-proven fractures, subdural or epidural hematoma, cerebral contusion, traumatic subarachnoid hemorrhage and/or internal trauma (excludes soft tissue injuries)	1	0	0	26	15	0.58	22	5	0.23
601. Perioperative/periprocedural new central nervous system deficit (excludes direct central nervous system procedure)*‡	5	1	1	21	5	0.24	2	0	0.00
807. Thrombosed distal bypass graft requiring repair (excludes arteriovenous grafts and fistulas used for dialysis)*‡	2	1	1	15	7	0.47	3	1	0.33
602. Perioperative/periprocedural new peripheral nervous system deficit with motor weakness (excludes direct procedures on specific nerve or sensory symptoms without weakness)*‡	0	1	0	10	8	0.80	2	0	0.00
804. Anastomatic leakage requiring repair*†	1	0	0	10	8	0.80	4	3	0.75
501. Laparoscopic unplanned conversion to open procedure because of injury and/or bleeding (excludes diagnostic laparoscopy with planned conversion or conversion based on findings, conversions based on difficulty identifying anatomy)	0	0	0	10	4	0.40	1	1	1.00
303. Intravascular catheter-correlated pneumothorax, regardless of size or treatment (excludes nonintravascular related such as those resulting from lung biopsy, thoracentesis, permanent pacemaker insertion, and others)	0	0	0	8	6	0.75	3	1	0.33

Continued

Appendix 1 ■ Continued

Adverse Event Type with Description	1,000 Case Set			57,452 Case Set					Proportion of Traditionally Detected Events by the System
	Events Identified by Manual Review	Events Identified by the System	Events Identified by the System and Verified by Manual Review	Events Identified by the System	Events Identified by the System and Verified by Manual Review	Positive Predictive Value of the System	Events Identified by Traditional Event Detection	Events Identified by Both the System and by Traditional Event Detection	
913. Unintentionally retained foreign body due to inaccurate surgical count or technique break	0	0	0	8	2	0.25	0	0	—
853. Ruptured uterus†	1	1	1	6	4	0.67	0	0	—
201. Aspiration pneumonia/pneumonitis in a nonintubated patient related to conscious sedation (excludes patients intubated/on a ventilator or with a known history of chronic aspiration)	0	0	0	5	1	0.20	0	0	—
701. Second- or third-degree burn in the hospital (excludes first-degree burn)	1	0	0	4	2	0.50	0	0	—
851. Hysterectomy in a pregnant woman†	0	0	0	3	2	0.67	1	1	1.00
917. Unexpected loss of limb or organ#	0	0	0	3	1	0.33	0	0	—
108. A medication error that resulted in permanent patient harm¶	0	0	0	3	0	0.00	0	0	—
109. A medication error that resulted in a near-death event (anaphylaxis, cardiac arrest)¶	0	0	0	3	0	0.00	0	0	—
921. Crime resulting in death or serious injury§	0	0	0	2	1	0.50	1	0	0.00
852. Inverted uterus†	0	0	0	1	1	1.00	0	0	—
911. Wrong patient, wrong site (surgical procedure)	0	0	0	1	0	0.00	0	0	—
915. Unexpected death§	2	0	0	1	0	0.00	0	0	—
110. A medication error that resulted in a patient death¶	0	0	0	0	0	—	0	0	—
854. Circumcision requiring repair	0	0	0	0	0	—	0	0	—
912. Incorrect procedure of treatment (invasive)	2	0	0	0	0	—	1	0	0.00
916. Cardiac and/or respiratory arrest requiring BLS/ACLS intervention#	2	0	0	0	0	—	1	0	0.00
918. Impairment of limb (limb unable to function at same level before occurrence)***	2	0	0	0	0	—	1	0	0.00
919. Loss or impairment of bodily functions (sensory, motor, communication or physiologic function diminished from level before occurrence)***	4	0	0	0	0	—	0	0	—
920. Errors of omission (related to patient's underlying condition) with death or serious injury§	0	0	0	0	0	—	0	0	—
922. Suicide and attempted suicides with death or serious injury§	0	0	0	0	0	—	0	0	—
923. Elopement from hospital death or serious injury§	0	0	0	0	0	—	0	0	—
961. Infant abduction	0	0	0	0	0	—	0	0	—

Continued

Appendix 1 ■ Continued

Adverse Event Type with Description	1,000 Case Set			57,452 Case Set					Proportion of Traditionally Detected Events Detected by the System
	Events Identified by Manual Review	Events Identified by the System	Events Identified by the System and Verified by Manual Review	Events Identified by the System	Events Identified by the System and Verified by Manual Review	Positive Predictive Value of the System	Events Identified by Traditional Event Detection	Events Identified by Both the System and by Traditional Event Detection	
962. Infant discharged to wrong family	0	0	0	0	0	—	0	0	—
963. Rape by another patient or staff	0	0	0	0	0	—	0	0	—
Total	65	32	16	1,590	704		322	110	

*Exclude cardiac related.

†Exclude birth related.

‡Include readmissions.

§Serious injury includes arrest, impairment of limb, loss of limb, or impairment of bodily functions.

||Within 48 hours.

¶Exclude adverse drug reaction that was not the result of a medication error.

#Exclude unexpected adverse occurrence directly related to the natural course of patient's condition (i.e., terminal or severe illness present on admission).

**Present at discharge or for at least two weeks if patient not discharged.



Automated Detection of Adverse Events Using Natural Language Processing of Discharge Summaries

Genevieve B Melton and George Hripcsak

J Am Med Inform Assoc 2005 12: 448-457

doi: 10.1197/jamia.M1794

Updated information and services can be found at:

<http://jamia.bmj.com/content/12/4/448.full.html>

These include:

References

This article cites 27 articles, 4 of which can be accessed free at:

<http://jamia.bmj.com/content/12/4/448.full.html#ref-list-1>

Article cited in:

<http://jamia.bmj.com/content/12/4/448.full.html#related-urls>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>