

Research Paper ■

Validation of a Discharge Summary Term Search Method to Detect Adverse Events

ALAN J. FORSTER, MD, FRCPC, MSc, JASON ANDRADE, MD, CARL VAN WALRAVEN, MD, FRCPC, MSc

Abstract **Objective:** Adverse events are poor health outcomes caused by medical care. Measuring them is necessary for quality improvements, but current detection methods are inadequate. We performed this study to validate a previously derived method of adverse event detection using term searching in physician-dictated discharge summaries.

Design: This was a retrospective, chart review study of a random sample of 245 adult medicine and surgery patients admitted to a multicampus academic medical center in 2002.

Measurements: The authors used a commercially available search engine to scan discharge summaries for the presence of 104 terms that potentially indicate an adverse event. Summaries with any of these terms were reviewed by a physician to determine the term's context. Screen-positive summaries had a term that was contextually indicative of an adverse event. We used a two-stage chart review as the gold standard to determine the true presence or absence of an adverse event.

Results: The average patient age was 62 years (standard deviation 18.6) and 55% were admitted to a medical service. By gold standard criteria, 48 of 245 patients had an adverse event. Term searching classified 27 cases with an adverse event, with 11 true positives; 218 cases were classified as not having an adverse event, with 181 true negatives. The sensitivity, specificity, and positive and negative predictive values were 0.23 (95% confidence interval [CI] = 0.11–0.35), 0.92 (95% CI = 0.88–0.96), 0.41 (95% CI = 0.25–0.59), and 0.83 (95% CI = 0.77–0.97), respectively.

Conclusion: Although the sensitivity of the method is low, its high specificity means that the method could be used to replace expensive manual chart reviews by nurses.

■ *J Am Med Inform Assoc.* 2005;12:200–206. DOI 10.1197/jamia.M1653.

Introduction

Adverse events are poor health outcomes caused by medical care.¹ Their detection is a very important aspect of improving patient safety. Although many hospitals use incident reporting or spontaneous voluntary reporting to identify safety problems, this method is very insensitive and nonspecific for identifying adverse events.² To avoid this problem, researchers have commonly used a two-stage chart review methodology. This requires a nurse to screen medical records for adverse events using objective criteria.^{3–6} Then, screen-positive records are reviewed by one or two physicians to de-

termine whether an adverse event truly occurred. This approach is desirable because it is more sensitive for adverse events. However, it is extremely costly.

A main contributor to the expense of the chart review method is the low rate of adverse events, which implies that most charts are reviewed unnecessarily. For example, in the major adverse event studies, nurse screening found that between 19% and 63%^{4,5,7,8} of charts were screen positive, meaning that as many as 81% of the records are reviewed by a nurse unnecessarily. In turn, only 20% of the screen-positive charts are judged to contain adverse events when reviewed by a physician.^{4,8,9} Thus, as most charts are reviewed needlessly, the process is very inefficient.

Adverse event detection could be made more cost effective if it were possible to identify charts with a low probability of an adverse event. These charts could then be excluded from subsequent review. One approach is to automatically scan electronic hospital data.¹⁰ This method uses database queries and search engines to find “signals” suggesting adverse event occurrence. Various types of data sources have been used for this purpose including discharge abstracts, pharmacy databases, and laboratory information systems. Although information from each of these data sources could identify adverse events, the signals often have very low positive predictive values. These range from 16% for International Classification of Diseases, Ninth Edition, Clinical Modification (ICD-9-CM) codes,¹¹ to 20% for adverse drug reactions identified from a pharmacy information system.^{12,13} The negative predictive

Affiliations of the authors: Department of Medicine University of Ottawa (AJF, CVW), Ottawa Health Research Institute (AJF, CVW), Institute for Clinical Evaluative Sciences (CVW), Ottawa, ON, Canada; University of British Columbia, Vancouver, BC, Canada (JA).

Supported by an internal research award by the Ottawa Hospital and by the Ottawa Internists Research Group.

Dr. Forster is the PSI Foundation Fellow for Innovative Health Services Research. Dr. van Walraven is an Ontario Ministry of Health Career Scientist.

Correspondence and reprints: Alan J. Forster, MD, FRCPC, MSc, C406-1053 Carling Avenue, Ottawa, ON, Canada K1Y 4E9; e-mail: <aforster@ohri.ca>.

Received for publication: 07/16/04; accepted for publication: 10/16/04.

values of these methods range between 76% and 91%^{12,14} and suggest that scanning electronic data could be useful as a screening tool. However, most of these studies do not report negative predictive values for adverse event identification.

A promising subcategory of electronic data screening for adverse events involves term searching of discharge summaries.¹⁵ Discharge summaries are physician-generated textual reports that describe a hospital encounter. Because these reports summarize patient hospitalizations, they should contain information about complications experienced during the hospital stay. Therefore, using a search engine to scan a discharge summary for words suggestive of complications may be an efficient way of identifying adverse events. Recently, Murff et al.¹⁵ evaluated this approach in a cohort of medical patients and found the technique to have a positive predictive value of 0.78, a negative predictive value of 0.74, a sensitivity of 0.64, and a specificity of 0.85 for identifying adverse events.

Given these test characteristics and the method's simplicity, the term search method could be used to replace the nurse reviewer in the traditional method of adverse event detection. However, because the Murff et al. study included only a single population within a single hospital, the method needs to be validated elsewhere. One reason for this concerns variation of discharge summary content across hospitals due to different discharge summary generation practices by physicians and different perceptions about whether documentation of complications could be prejudicial. In addition, surgeons may be more or less descriptive than internists in describing complications. To validate the term search method, we performed the current study in a diverse mix of patients admitted to a Canadian teaching hospital.

Methods

Overview

This research was a substudy of the Ottawa Hospital Patient Safety Study, which was designed to measure the risk of adverse events in adult patients admitted to a Canadian teaching hospital.⁷ The Ottawa Hospital Patient Safety Study used a two-stage chart review process to identify adverse events. Adverse events were defined as poor patient outcomes that resulted from medical care rather than the natural history of disease. For the present study, we performed term searching within discharge summaries using a computerized text-word search engine (dtSearch Desktop available at www.dtSearch.com). This was done independently of the chart review methodology or the results of the chart review. The chart review was the gold standard for evaluating the test characteristics of the term search method. The protocol was approved by the Ottawa Hospital Research Ethics Board.

Setting and Patients

The Ottawa Hospital is a multicampus academic hospital. The patients included in the Ottawa Hospital Patient Safety Study consisted of a random sample of nonpsychiatric patients admitted to the hospital during fiscal 2002. Patients were eligible for the substudy if they were admitted to medical or surgical services and if their physician completed a discharge summary.

Term Search Detection Method

We created a set of machine-readable files containing discharge summaries. At our institution, discharge summaries are dictated as free-text narrative reports by physicians.

They are then transcribed by a central transcription service and stored as text files within an internal database. These files are sent to our online clinical data repository, and hard copies are placed on the paper chart and mailed to physicians indicated by the dictating physician.

We reviewed the electronic medical records of the patients in the original sample to determine whether a discharge summary was created. If a discharge summary existed, we created a text file using a unique identifier as its name and stored it in our study database. In addition, we abstracted the level of training of the dictating physician (attending, resident, medical student), and the delay between patient discharge and summary transcription (in days).

Our second step was to establish an adverse event lexicon (see Appendix 1). We used a set of 104 terms that were used by Murff et al.¹⁵ These terms were originally generated by mapping adverse event types to specific complications. For example, the adverse event procedural complication was mapped to several terms including perforation, pneumothorax, and wound infection.

Next, we used a computer search engine to scan the discharge summaries for the presence of our terms. We used dtSearch Desktop as our search engine. We used this software because it is commercially available, simple to use, and can be programmed to search text across a network as well as a desktop computer. It does not require any specialized knowledge or training.

The discharge summaries containing at least one of our terms were then reviewed manually by a board-certified general internist. This step was required to determine whether the context of the term in the discharge summary was suggestive of an adverse event. For example, consider the context of the term *delirium* in the following two discharge summaries:

Example 1: "The delirium with hallucinations and delusions was thought to be secondary to recent use of narcotics..."

Example 2: "It was suspected that he had delirium from a urinary tract infection."

The context of the term in Example 1 is suggestive of an adverse event, whereas in Example 2, it is not. If the context was considered to be consistent with an adverse event, then the patient was rated as having an adverse event. To complete this task, it took the physician approximately 2 minutes per discharge summary.

Gold Standard Detection Method

Our two-stage chart review process was similar to previously described methods.³⁻⁶ Briefly, a nurse who was trained in adverse event detection methodology reviewed charts to identify hospitalizations in which at least one of 16 screening criteria occurred. Charts that were screened positive by the nurse reviewer were reviewed by a physician to determine whether an adverse outcome that was significant enough to cause prolongation of the hospitalization, temporary or permanent disability, or death had occurred. The physician then used a 6-point scale to rate his or her confidence that an adverse outcome was due to medical care. A score of 4 or higher indicated that the outcome was an adverse event. All hospitalizations that were rated as "adverse event positive" and a sample of "adverse event negative" charts

were re-reviewed by a second physician. All adverse events were then rated for preventability, severity, type, and location, according to standard definitions.³⁻⁵

Analysis

SAS 8.2 was used for all analyses. Continuous and ordinal variables were compared using Student's t-test and Wilcoxon rank test, respectively. A chi-square test was used to compare categorical variables. Confidence intervals (CIs) (95%) for likelihood ratios (LRs) were calculated using methodology described by Simel et al.¹⁶ When categorizing continuous variables, we selected variable quartiles.

Results

The original study included 364 medical and surgical patients or 1.3% of all such patients admitted to the Ottawa Hospital in fiscal 2002. Of these patients, 245 (67.3%) had a discharge summary and were included in this analysis (Fig. 1). Patients with a discharge summary were significantly older, had a longer hospital length of stay, and were more likely to have been admitted urgently or emergently and to a medical service (Table 1).

Figure 1 demonstrates the results of our term search method with reference to adverse events determined by the gold standard method. Overall, 48 of 245 (20%, 95% CI = 15%–25%) study patients had an adverse event. Terms indicative of a possible adverse event were present in the discharge summaries of 168 patients and absent in 77. On physician review of the discharge summary, the term's context was not suggestive of an adverse event in 141 patients. Of the remaining 27 patients, 11 had adverse events.

Table 2 describes the study patients by whether an adverse event occurred during the admission. Study patients with an adverse event were significantly older and had a longer hospitalization. The overall median delay to discharge summary dictation was 14 days (interquartile range, 1–63), and 88.2% of summaries were dictated by residents or staff. The timing, severity, preventability, and type of adverse events are presented in Table 3. Most adverse events occurred prior to the index hospitalization. The characteristics of the subset of adverse events identified using our term search method were similar to the entire sample.

The operating characteristics of the term search method for adverse event detection are presented in Table 4. The term search method was very insensitive for adverse events (23%, 95% CI = 11%–35%) but was adequately specific (92%, 95% CI = 88%–96%). The positive and negative predictive values for the term search method were, respectively, 42% (95% CI = 25%–59%) and 83% (95% CI = 77%–87%).

The only patient, hospitalization, and discharge summary characteristics that influenced these operating characteristics were admission urgency and dictating physician (Fig. 2). The specificity of term search method for adverse events was significantly higher for nonelective admissions ($p = 0.03$) and discharge summaries dictated by residents or staff ($p = 0.01$). The operating characteristics of term search method improved when the analysis was limited to nonelective admissions and summaries dictated by residents or staff (specificity, 96.3%; 95% CI = 93.1%–99.5%; + LR 6.55, 95% CI = 2.3–18.7).

We reviewed the discharge summaries of the 37 false-negative patients to determine why the term search method failed

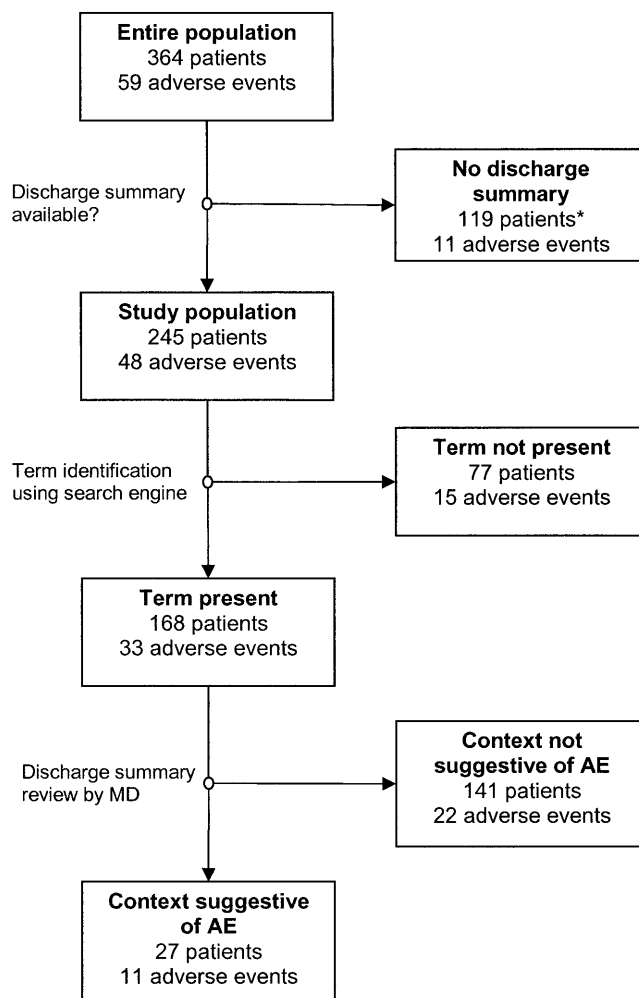


Figure 1. Study flow. The number of adverse events was identified using the gold standard method. The gold standard was a two-stage chart review. *Excluded from the study.

Table 1 ■ Comparison of Original Study Patients Who Were and Were Not Included in This Study

	Discharge Summary Dictated?		p-Value
	No (n = 119)	Yes (n = 245)	
Mean age, yr (SD)	57.1 (19.1)	62.3 (18.6)	0.014
Median length of stay, days (IQR)	6.5 (15.5)	10.4 (12.5)	0.01
Admission type			
Elective	58 (54.2%)	49 (45.8%)	<0.001
Urgent	56 (22.7%)	191 (77.3%)	
Emergent	5 (50%)	5 (50%)	
Service			
Medicine	23 (14.6%)	135 (85.4%)	<0.001
Surgery	96 (46.6%)	110 (53.4%)	
Campus			
Civic	60 (36.8%)	103 (63.2%)	0.13
General	59 (29.4%)	142 (70.6%)	

Only patients with a discharge summary (n = 245) are included in this study. Unless otherwise indicated, counts and percentages are provided.

IQR = interquartile range.

Table 2 ■ Description of Study Patients by Whether an Adverse Event Was Detected by Chart Review

	Adverse Event*		p-Value
	No (n = 197)	Yes (n = 48)	
Mean patient age, yr (SD)	60.9 (18.8)	68.1 (16.9)	0.02
Age quartiles			
<50	55 (27.9%)	7 (14.6%)	0.07
50–66	51 (25.9%)	9 (18.8%)	
67–77	45 (22.8%)	17 (35.4%)	
>77	46 (23.4%)	15 (31.3%)	
Median length of stay (IQR)	5 (4–10)	10.5 (5–18)	0.02
Length of stay quartiles			
0–4	67 (34.0%)	9 (18.8%)	0.004
5–6	50 (25.4%)	6 (12.5%)	
7–12	41 (20.8%)	14 (29.2%)	
>12	39 (19.8%)	19 (39.6%)	
Median days from DC to summary	15 (1–59)	13.5 (0–74)	>0.2
Days-DC to summary quartiles			
0–1	57 (28.9%)	14 (29.2%)	>0.2
2–14	41 (20.8%)	11 (22.9%)	
15–63	51 (25.9%)	10 (20.8%)	
>63	48 (24.4%)	13 (27.1%)	
Admission type			
Elective	43 (21.8%)	6 (12.5%)	>0.2
Urgent	151 (76.6%)	40 (83.3%)	
Emergent	3 (1.5%)	2 (4.2%)	
Service			
Medicine	103 (52.3%)	32 (66.7%)	0.07
Surgery	94 (47.7%)	16 (33.3%)	
Campus			
Civic	83 (42.1%)	20 (41.7%)	>0.2
General	114 (57.9%)	28 (58.3%)	
Training level of dictating MD			
Student	20 (10.2%)	9 (18.8%)	>0.2
Resident	97 (49.2%)	20 (41.7%)	
Staff	80 (40.6%)	19 (39.6%)	

Adverse events were poor outcomes that were judged to be due to medical management rather than the disease. Unless otherwise indicated, counts and percentages are provided.

DC = discharge.

*As detected by chart review.

to identify the adverse event. Twelve of these discharge summaries did not mention the adverse event. Another six alluded to the outcome, but did not mention its etiology. For example, several prehospital adverse events, which were caused by care in other facilities, contained information pertaining to the previous health care encounter but contained no explicit link between the encounter and the outcome. The remaining 19 discharge summaries explicitly mentioned the adverse events but used terms or expressions to describe them that we did not include in our search strategy. Five of these could have been identified by including verbs that relate outcomes to treatments (e.g., caused by, induced by, secondary to). Two discharge summaries contained one of our terms but were spelled differently. The remainder consisted of terms that were not a part of our search strategy, including toxicity, pneumonia, febrile neutropenia, dehiscence, and nausea.

Table 3 ■ Timing, Severity, and Preventability of All Adverse Events (AEs) and Those Detected by Term Searching

AE Type	Gold Standard* (%)	Identified Using Term Search Method (%)
All	48 (100%)	11 (100%)
AE occurred prior to admission	31 (65%)	5 (46%)
Severe (resulting in death or permanent disability)	9 (19%)	3 (27%)
Preventable	18 (38%)	3 (27%)
Adverse drug event	26 (54%)	5 (46%)
Operative complication	11 (23%)	4 (36%)
Nosocomial infection	10 (20%)	2 (18%)
Diagnostic error	4 (8%)	1 (9%)
Procedural complication	5 (10%)	2 (18%)
System errors†	4 (8%)	0 (0%)

Although the number of events is small, the characteristics of the adverse events identified using the gold standard are similar to the characteristics of the subset identified by the screening method.

*The gold standard was a two-stage chart review.

†System errors were due to poorly designed systems such as inadequate training, poorly designed protocols, or inadequate or defective equipment.

Table 4 ■ Operating Characteristics of Discharge Summary Screening for Adverse Events

	Adverse Event		Total
	Yes	No	
DC summary screen positive			
Yes	11	16	27
No	37	181	218
Total	48	197	245

Sensitivity: 0.23 (0.11–0.35); Specificity: 0.92 (0.88–0.96); +LR: 2.82 (1.40–5.68); –LR: 0.83 (0.72–0.98); positive predictive value: 0.41 (0.25–0.59); negative predictive value: 0.83 (0.77–0.87).

Patients were classified as having an adverse event if physicians judged their poor outcome to be due to medical treatment.

We also reviewed the discharge summaries of the 16 false-positive cases; ten of these cases were major postoperative events, including death in two instances. Presumably, there was information contained in the medical record that led the physician reviewer to rate the outcome as due to the disease as opposed to treatment in these instances. An additional three cases were indeed adverse events but were of minor consequence so that they did not meet the strict severity definition required for adverse event classification by the gold standard method. For example, one discharge summary commented on hyperglycemia caused by prednisone. Another three cases were preadmission intentional overdoses, which were simply misclassified by our physician reviewer during the discharge summary review.

Discussion

The term search method had a specificity of 0.92 and a negative predictive value of 0.83 for identifying adverse events in

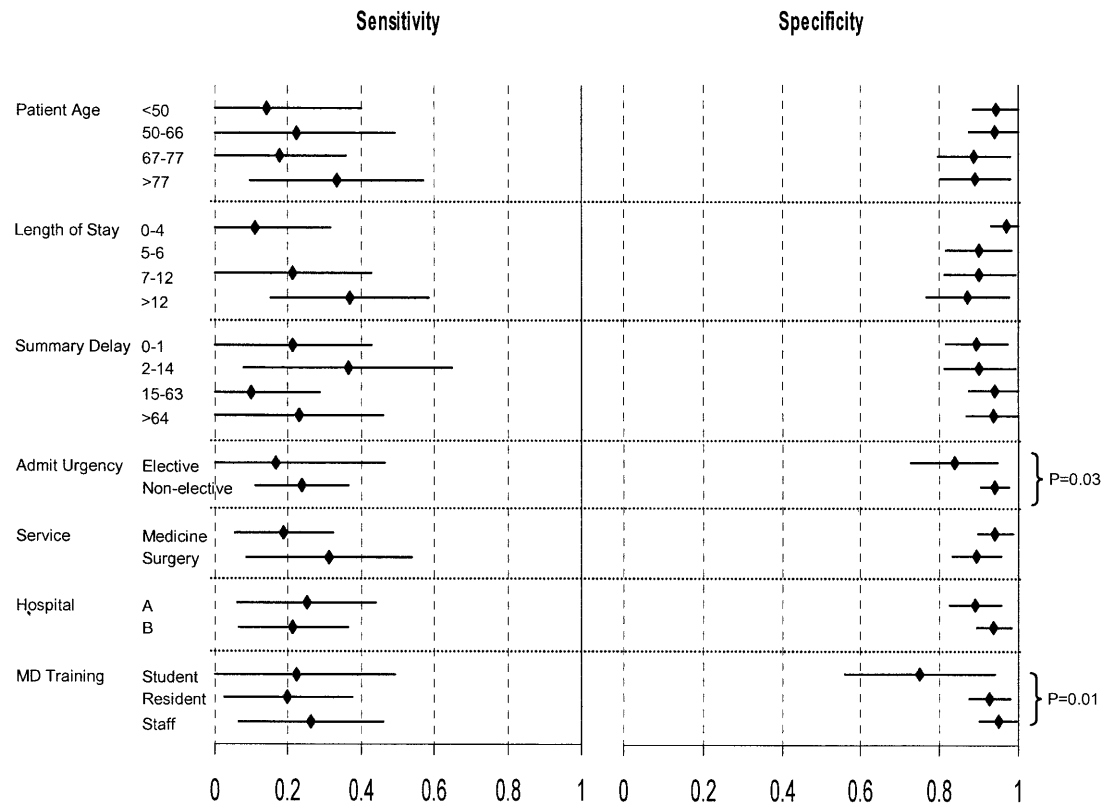


Figure 2. Effect of patient, hospitalization, and discharge summary characteristics on the sensitivity and specificity of text words for adverse hospital events. Patients were classified with an adverse event if physicians judged their poor outcome to be due to medical treatment.

hospitalized patients. However, the sensitivity and the positive predictive value were relatively low at 0.23 and 0.41, respectively. The method's overall usefulness was also diminished because almost a third of patients did not have a discharge summary, including 78% of patients discharged before 48 hours. The method's operating characteristics improved slightly if we excluded discharge summaries dictated by medical students and discharge summaries for elective cases. Although unproven, minor modifications in our search strategy could potentially improve the test characteristics further.

Although the sensitivity and positive predictive value of the term search method in our population appears poor by standard clinical criteria, they compare favorably with other methods of adverse event detection. For example, studies of adverse event detection using billing codes¹⁴ or ICD-9-CM¹¹ codes had positive predictive values of 0.3 and 0.16, respectively. In addition, incident reporting, a process that involves voluntary reporting of errors and adverse events by providers and is used in most hospitals, has been shown to identify less than 6% of adverse drug events.² Furthermore, the high negative predictive value that we found suggests that this method could definitely be used to exclude charts from further review.

The derivation study of Murff et al. of the term search method found a sensitivity of 0.64, a positive predictive value of 0.78, and a negative predictive value of 0.74.¹⁵ Our results show that the term search method does not perform as well in a different hospital. We can only speculate on the reasons for this.

Most likely, it relates to different details included in discharge summaries. Alternatively, the better positive predictive value in the Murff et al. study may be due to the inclusion of less severe adverse events, which then led to a much higher prevalence. As well, we scanned the entire discharge summary rather than using the approach of Murff et al. to scan simply the course in the hospital section. This should have led to an increased sensitivity, given the high proportion of events occurring before the hospital encounter in our study (Table 3). The fact that sensitivity is lower despite this suggests that a different approach must be used to identify prehospital events. Finally, we evaluated both medical and surgical patients and found that admission service did not affect the usefulness of the method (Fig. 2).

Our study was strengthened by using a gold standard that we applied in a blinded fashion. It is interesting to note that only seven of 25 studies identified by a recent systematic review evaluating the use of information technology to detect adverse events used a gold standard.¹⁰ In addition, some investigators limit their evaluation to the positive signals, thereby calculating a positive predictive value only. In general, when evaluating any diagnostic test, it is fundamentally important to assess patients without the condition as well as the cases because outcome prevalence greatly influences the test's positive predictive value. In addition, if one is to use a technology to exclude patients from subsequent evaluation, it is imperative to know the test's negative predictive value.

The gold standard that we used is an accepted method of adverse event detection and has been used in every major

epidemiological study of patient safety.^{3-6,8} It has also been used as a gold standard by many other investigators evaluating the utility of various methods of adverse event detection.¹⁰ However, despite its widespread use, the chart review method may not be the optimal gold standard. Some investigators have used active surveillance to identify adverse events and identified considerably more events than chart review methods.^{17,18} The purported reason for this is that the chart review, like the discharge summary, may not contain descriptions of all adverse events if providers fail to include them in their documentation. We did not use prospective surveillance because it has not been generally accepted as an adverse event detection method.

Another limitation of the method is its dependency on the quality of the discharge summaries. A third of patients in our original cohort (119/364) did not even have a discharge summary generated. Most of these patients did not require one by law because they stayed in the hospital less than 48 hours (89/119, 75%). Apart from actually dictating a summary, the content was also an important deficiency. In 12 false-negative cases, the adverse event was not mentioned in the discharge summary even though it was clinically important and contributed to the hospital length of stay.

The term search method could be applied by any hospital whose discharge summaries are stored in machine-readable format. We suggest that it could replace the nurse screener in the two-phase chart review methodology. In our original study, it took the nurse reviewer 121 hours (20 minutes/chart) to complete the initial screening. Subsequently, it took the physician reviewer 60 hours (15 minutes/chart) to review the 238 charts that were screen positive. In contrast, to perform the term search method, it took 5.7 hours of physician time to review the 168 discharge summaries (2 minutes per discharge summary), followed by another 6.8 hours of physician time to review the 27 charts identified as possibly containing an adverse event. Thus, using term searching, we reduced the physician time to one-fifth of what was required to complete the original method. More important, the method obviates using a nurse reviewer. Although only 23% of the adverse events were identified, it appears as if the types of adverse events found were representative of all adverse events (Table 3), thus making it possible to set priorities and develop rational quality improvement strategies. The sensitivity of the method could be increased to 69% by skipping the manual discharge summary review. However, this decreases the positive predictive value to 20%, implying that the physician will have to manually review more charts in their entirety.

In conclusion, the method of adverse event detection using term searching is promising. Future research should use search engines in a more sophisticated manner than we used. For example, the search engine should be programmed to allow for variations in spelling, incorporate Boolean logic, and account for negative modifiers. The list of terms should also be expanded, for example, using terms like naloxone or toxicity. Finally, natural language processing has been used to aid in classification of textual reports for a variety of conditions, such as chest x-ray reports for pneumonia. Application of this technology to adverse event detection may be similarly beneficial. Finally, as more hospitals integrate their information systems, it will be possible to combine different detection

strategies as other investigators have done in outpatients for adverse drug events.¹⁹

References ■

1. The Institute of Medicine. To err is human: building a safer health system. Washington, DC: National Academy Press; 2000.
2. Cullen DJ, Bates DW, Small SD, Cooper JB, Nemeskal AR, Leape LL. The incident reporting system does not detect adverse drug events: a problem for quality improvement. *Joint Commission J Qual Improve.* 1995;21:541-8.
3. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med.* 1991;324:370-6.
4. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care.* 2000;38:261-71.
5. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The Quality in Australian Health Care Study. *Med J Aust.* 1995;163:458-71.
6. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ.* 2001;322:517-9.
7. Forster AJ, Asmis TR, Clark HD, et al. Ottawa Hospital Patient Safety Study: incidence and timing of adverse events in patients admitted to a Canadian teaching hospital. *CMAJ.* 2004;170:1235-40.
8. Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ.* 2004;170:1678-86.
9. Forster AJ, Asmis TR, Clark HD, et al. Ottawa Hospital Patient Safety Study: incidence and timing of adverse events in patients admitted to a Canadian teaching hospital. *CMAJ.* 2004;170:1235-40.
10. 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.
11. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. *Med Care.* 2000;38:796-806.
12. 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.
13. Levy M, Azaz-Livshits T, Sadan B, Shalit M, Geisslinger G, Brune K. Computerized surveillance of adverse drug reactions in hospital: implementation. *Eur J Clin Pharmacol.* 1999;54:887-92.
14. Bates D, O'Neil AC, Petersen L, Lee T, Brennan T. Evaluation of screening criteria for adverse events in medical patients. *Med Care.* 1995;33:452-62.
15. 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.
16. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol.* 1991;44:763-70.
17. Bates DW, Cullen DJ, Laird N, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. *JAMA.* 1995;274:29-34.
18. Forster AJ, Halil RB, Tierney MG. Pharmacist surveillance of adverse drug events. *Am J Health-Syst Pharm.* 2004;61:1466-72, 2004.
19. 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.

Appendix 1 ■ Terms

Accident (2, 0, 0)	Fluid resuscitation (2, 0, 0)	Perforation (0, 0, 0)
Accidentally (1, 1, 0)	Hallucinations (1, 1, 0)	Pneumothorax (2, 1, 0)
Acute renal failure (3, 0, 0)	Heel ulcer (0, 0, 0)	Polypharmacy (0, 0, 0)
After surgery (0, 0, 0)	Hematoma (3, 0, 0)	Postop (0, 0, 0)
After the operation (1, 0, 0)	Hemorrhage (2, 0, 0)	Postoperative (33, 6, 1)
Agitation (0, 0, 0)	Hospital acquired (0, 0, 0)	Pressure sore (0, 0, 0)
Allergic (1, 0, 0)	Hyperglycemia (3, 1, 0)	Pressure ulcer (0, 0, 0)
Allergic reaction (0, 0, 0)	Hypervolemia (0, 0, 0)	PTX (0, 0, 0)
Aspiration pneumonia (0, 0, 0)	Hypoglycemia (0, 0, 0)	Rash (2, 0, 0)
Bed sore (0, 0, 0)	Hypotension (1, 1, 0)	Renal failure (10, 1, 1)
CCU (1, 0, 0)	Hypoxemia (1, 1, 1)	Reopen (0, 0, 0)
Chest tube (3, 0, 0)	Hypoxia (5, 2, 0)	Reopening (0, 0, 0)
<i>Clostridium difficile</i> (1, 1, 1)	ICU (2, 1, 0)	Required fluids (0, 0, 0)
Complicated (4, 1, 0)	IV infection (0, 0, 0)	Respiratory distress (3, 2, 0)
Complication (0, 0, 0)	IV sepsis (0, 0, 0)	Respiratory failure (3, 3, 0)
D50 (0, 0, 0)	Laceration (3, 0, 0)	Rupture (0, 0, 0)
Death (4, 1, 0)	Lethargic (1, 1, 0)	Sedated (0, 0, 0)
Decubiti (0, 0, 0)	Line infection (0, 0, 0)	Shortness of breath (15, 0, 0)
Decubitus (0, 0, 0)	Line sepsis (0, 0, 0)	SICU (0, 0, 0)
Deep vein thrombosis (1, 0, 0)	Low sugars (0, 0, 0)	Skin ulcer (0, 0, 0)
Delirium (7, 3, 3)	Medicine consult (0, 0, 0)	Slipped (1, 0, 0)
Desaturation (0, 0, 0)	Mental status (1, 1, 0)	Somnolent (0, 0, 0)
Died (5, 1, 0)	MICU (0, 0, 0)	Subtherapeutic (0, 0, 0)
Discontinued (13, 3, 1)	Mistake (1, 1, 0)	Supratherapeutic (0, 0, 0)
Dissection (4, 0, 0)	Mistakenly (0, 0, 0)	Surgery (38, 4, 3)
DKA (0, 0, 0)	Nonresponsive (0, 0, 0)	Syncopy (0, 0, 0)
Drop of (0, 0, 0)	Nosocomial (0, 0, 0)	Telemetry (1, 0, 0)
Dropped (4, 1, 0)	NSICU (0, 0, 0)	Transfer (11, 0, 0)
Drug eruption (0, 0, 0)	OD (2, 0, 0)	Transferred (34, 3, 0)
Drug rash (0, 0, 0)	Operating room (32, 0, 0)	Transfusion (4, 2, 0)
Error (0, 0, 0)	OR (55, 0, 0)	Unresponsive (4, 1, 0)
Expired (2, 1, 0)	Overdose (4, 3, 0)	Volume (11, 1, 0)
Failure (35, 4, 0)	Overload (1, 0, 0)	Wet (1, 0, 0)
Fall (3, 0, 0)	Oversedation (0, 0, 0)	Wound infection (1, 1, 0)
Fell (7, 1, 0)	Packed RBC (0, 0, 0)	

These are the 104 terms used. The search engine was not case sensitive. The numbers in parentheses represent, respectively, the number of times that the term was identified in a discharge summary, the number of times that the context of the term was suggestive of an adverse event, and the number of times that the term was associated with a true adverse event, as determined by the gold standard method.