

Research Paper ■

Electronically Screening Discharge Summaries for Adverse Medical Events

HARVEY J. MURFF, MD, MPH, ALAN J. FORSTER, MD, MSc, JOSH F. PETERSON, MD, MPH, JULIE M. FISKIO, BA, HEATHER L. HEIMAN, MD, DAVID W. BATES, MD, MSc

Abstract Objective: Detecting adverse events is pivotal for measuring and improving medical safety, yet current techniques discourage routine screening. The authors hypothesized that discharge summaries would include information on adverse events, and they developed and evaluated an electronic method for screening medical discharge summaries for adverse events.

Design: A cohort study including 424 randomly selected admissions to the medical services of an academic medical center was conducted between January and July 2000. The authors developed a computerized screening tool that searched free-text discharge summaries for trigger words representing possible adverse events.

Measurements: All discharge summaries with a trigger word present underwent chart review by two independent physician reviewers. The presence of adverse events was assessed using structured implicit judgment. A random sample of discharge summaries without trigger words also was reviewed.

Results: Fifty-nine percent (251 of 424) of the discharge summaries contained trigger words. Based on discharge summary review, 44.8% (327 of 730) of the alerted trigger words indicated a possible adverse event. After medical record review, the tool detected 131 adverse events. The sensitivity and specificity of the screening tool were 69% and 48%, respectively. The positive predictive value of the tool was 52%.

Conclusion: Medical discharge summaries contain information regarding adverse events. Electronic screening of discharge summaries for adverse events using keyword searches is feasible but thus far has poor specificity. Nonetheless, computerized clinical narrative screening methods could potentially offer researchers and quality managers a means to routinely detect adverse events.

■ *J Am Med Inform Assoc.* 2003;10:339–350. DOI 10.1197/jamia.M1201.

Affiliations of the authors: Division of General Internal Medicine, Brigham and Women's Hospital, Boston, Massachusetts. Dr. Murff and Dr. Peterson are currently in the Division of General Internal Medicine, Vanderbilt University Medical Center, Nashville, TN; Ms. Fiskio, Dr. Heiman, and Dr. Bates are currently in the Division of General Internal Medicine, Brigham & Women's Hospital, and Harvard Medical School, Boston, MA; Dr. Forster is currently at the University of Ottawa, Clinical Epidemiology Unit, Ottawa Health Research Unit, Ottawa Hospital, Ottawa, Ontario, Canada.

A part of this material has been presented as a poster at the 2001 American Medical Informatics Association Annual Symposium (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.* 2002;9(6 suppl):S50–1) and the 2002 Society of General Internal Medicine National Meeting (Murff HJ, Forster AJ, Peterson JF, Fiskio JM, Heiman HL, Bates DW. Electronically screening discharge summaries for adverse medical events. *J Gen Intern Med.* 2002;17(suppl 1):A205).

Dr. Murff was supported by a NRSA training grant, 5T321101-12, over the duration of the project.

Correspondence and reprints: David W. Bates, MD, MSc, Chief, Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA 02115; e-mail: <dbates@partners.org>.

Received for publication: 08/08/02; accepted for publication: 01/29/03.

Patient safety has emerged as a highly important issue for health care.^{1–3} Adverse events (AEs)—defined as injuries due to medical management—result in numerous injuries and deaths every year within the United States.^{3,4} Prior studies have found rates of AEs ranging from 2.9% to 16.6% of inpatient admissions.^{5–7} These injury rates have prompted the Institute of Medicine to define patient safety as a key goal in health care quality improvement.²

One of the first laws of quality improvement is that to improve something, one must be able to measure it.⁸ However, most organizations do not have effective approaches to routinely detect and measure AEs. Voluntary incident reports underestimate AE rates, detecting 1.5% of AEs⁹ and 6% of adverse drug events compared with manual chart review.^{10,11} While chart review is effective for research,^{12–14} it is too costly for routine use.

Identification of AEs by searching electronic patient records, rather than manually reviewing paper charts, offers a potential solution. Studies have found that computerized screening for adverse drug events (ADEs) requires 20% of the time and detects 69% as many cases compared with

manual review.¹¹ While computerized screening for ADEs has been successful,^{15–18} electronically screening for AEs remains problematic. In one study, electronic screening of claims data for medical admissions detected potential quality problems in only 15.7% of the alerts triggered.¹⁹

We hypothesized that discharge summaries would include valuable information related to adverse events. Therefore, we developed and assessed an electronic screening method that searched discharge summaries to build a tool that organizations could use routinely to detect AEs. We hoped to develop an approach that would be more efficient than manual screening but with comparable sensitivity.

Methods

Study Site

The Brigham and Women's Hospital is a 726-bed tertiary care teaching hospital. For medical patients, housestaff are responsible for patient management and discharge summary completion. Discharge summaries are dictated, transcribed, and electronically stored within the Brigham Integrated Computer System (BICS).²⁰ Medical record policy requires that physicians complete discharge summaries within 30 days for all patients who have been hospitalized more than 72 hours.

Sample Selection

The patient population consisted of all admissions to the general medicine and medical subspecialty services with discharge dates between January 1 and June 30, 2000. The only exclusion criterion was lack of a completed discharge summary. We randomly selected a smaller subset of admissions from this larger population to pilot test the screening tool (cohort 1). We then applied the tool to the remaining admissions from this cohort and reviewed a random set of these records (cohort 2).

Randomization for cohort 1 was completed using visit numbers. Visit numbers are assigned consecutively for each patient beginning with the initial visit to the institution. Due to an error in the randomization process, we only selected patients with either their first, second, or third visit to the study site occurring during the study period. As a result, all admissions initially identified by the screening tool and reviewed involved patients who were new to the institution. The Human Subjects Committee at the Brigham and Women's Hospital approved the study.

Definitions and Outcomes

The primary outcome of the study was an AE, defined as an injury resulting from medical management rather than the patient's underlying condition.³ We included both injuries that resulted in a disability at discharge and those that resulted in a prolonged hospitalization, the definition used for the Harvard Medical Practice Study (MPS),²¹ as well as injuries that resulted in transient disability or abnormal laboratory studies, which would not have met the MPS

criteria. We applied the more inclusive definition to capture events that would be clinically significant yet might not have met the malpractice-oriented definition of injury used in the MPS. An example of such an event would be an elderly patient who required naloxone because of inappropriately dosed narcotics but who recovered quickly and did not require excess hospital days. These types of events are undesirable and important to detect for improving safety, even if they do not increase length of stay.

Chart review was considered the gold standard for the presence of an AE. Charts were reviewed and information abstracted regarding AEs. Case histories were compiled based on these abstracted data by one of the authors (HJM) and reviewed by two independent, board-certified internists (AJF, JFP). The reviewers were blinded to whether the screening tool had identified the admission. Both physicians reviewed all case histories. Disagreements were settled by consensus and, when an agreement could not be reached, by a third-party review (DWB). Judgments concerning AEs were made using structured data forms based on the Adverse Event Analysis Form developed and used by investigators in the MPS.^{7,21} This data form (Appendix A) is available as an online data supplement at the JAMIA website (www.jamia.org).

Using this methodology, reviewers make implicit judgments regarding whether a patient experienced an injury. If a patient has experienced an injury, the next step is for the reviewer to rate his or her "confidence" that the patient's medical management resulted in the injury. Confidence ratings range from 1 to 6, with 1 being "little or no evidence for management causation" and 6 being "virtually certain evidence for management causation." As with the MPS, confidence scores of 4 or greater (4 = "management causation more likely than not, more than 50–50 but close call") are considered an adverse event. AEs then are judged regarding the severity of the injury, the disability associated with the injury, and the preventability of the injury.

For this study, adverse events were placed into categories using the MPS methodology, while adverse drug events were categorized using review methodology from the ADE Prevention Study.¹⁵ We used the ADE Prevention Study categorization because the MPS used only limited categories for ADE classification.

Objectives and Development of the Screening Tool

The standard methodology used to detect adverse events is structured manual chart review, which was used in the MPS.²¹ In the MPS, medical records underwent a two-step review process. Initially, records were reviewed manually for the presence of one or more predefined screening criteria. Examples of the predefined screening criteria include "transfer from a general care unit to a special care unit" and "cardiac or respiratory arrest." If a chart contained one of these criteria, the chart then was referred for physician review to make the final judgments concerning AE occurrence. The positive predictive value of the preliminary

screening process was determined to be 21%.¹³ Although the time necessary to screen a chart was not reported in that study, similar studies have reported that about half of all charts can be screened in less than 10 minutes, while 11% required more than 20 minutes.²² Because this manual prescreening approach results in numerous false-positive charts, which is both time-consuming and costly, we sought to develop a tool that would automate this process.

After reviewing the 18 adverse event categories used by the MPS,²¹ we eliminated seven because they were either not relevant to adult medical patients or unlikely to be captured from the discharge summary. The remaining 11 adverse event categories were used as a framework to develop our screening trigger words. We mapped multiple standard event concepts, such as iatrogenic hypoglycemia, to these 11 categories. Event concepts were determined based on the authors' own experiences and clinical knowledge. From these event concepts, the authors compiled a list of terms (trigger words) that might be used to represent these concepts within the discharge summary. Multiple trigger words could be mapped to an individual event concept (Table 1). An adverse event was considered to be present if any of the trigger words matched exactly with strings in the text. Ninety-five specific trigger words were compiled and mapped to these event concepts (Fig. 1).

Discharge summaries at the study institution are semistructured using a format that divides different data elements into separate sections. The discharge summary begins with a section entitled "chief complaint" in which the main reason for admission is listed. This section is then followed by several other sections including "history of present illness," "past medical history," "medication," and "hospital course." To ensure that an alerted trigger word represented an event occurring during the index hospitalization, we searched only the "hospital course" section of the discharge summary for trigger words. The "hospital course" section is a free-text narrative that we electronically parsed into single words and searched for matching trigger words. If a trigger word was detected within the clinical narrative, an alert was generated, and the subject's medical record number, date of admission, date of discharge, and trigger word were recorded. Any chart that had a trigger word within its discharge summary was considered a "screened-positive chart," while a medical chart without a trigger word within the discharge summary was considered a "screened-negative chart."

Discharge Summary Review

The overall goal of the project was to create a screening tool that would replace entirely the need for manual chart screening. Ultimately, this electronic tool would search discharge summaries and identify charts with a high likelihood of having an AE. These charts then would go directly to structured physician review. Because it was initially unclear how the identified trigger words would be represented within the discharge summary, and because we had concerns about excessive false-positive alerts, we reviewed the entire text of the discharge summary asso-

ciated with an alert before chart review. This initial rating process was performed to assist the investigators in the development of the tool but was not intended to be a regular part of the screening process. In cohort 1, the results of the discharge summary review did not influence whether the chart was reviewed. This allowed us to determine how well electronic screening would likely perform in the absence of manual discharge review.

After determining the test characteristics of the electronic screening strategy, we performed a second study to evaluate whether a combined screening strategy—electronic screening followed by manual discharge summary review—might improve our overall ability to detect AEs. Using data from cohort 1, we compared the sensitivity and specificity of the combined screening strategy of electronic screening plus manual discharge review with electronic screening alone.

In the combined screening strategy, screened-positive charts underwent a manual review of the discharge summary. Based on information from the discharge summary, judgments were made by a single reviewer concerning the likelihood that an AE had occurred. Screened-positive charts were categorized as either a "possible AE" or "no AE." A "possible AE" occurred when documentation was suggestive of an AE. If no injury was discernible from the discharge summary, the trigger word was considered to represent "no AE" (Fig. 2). To determine the reliability of the discharge summary review process, a second independent reviewer evaluated a random 23% of all screened-positive charts.

Sampling Strategy

In cohort 1, any screened-positive chart underwent a manual chart review. To determine how well the screening tool performed overall, a random 25% sample of screened-negative charts also were reviewed manually. Based on this random sample, we estimated the number of AEs that occurred within the entire set of patients with screened-negative charts (Fig. 3). Of the initial subset of admissions screened, physicians manually reviewed 70% (295 of 424) of the charts.

We applied the screening tool to a second cohort of patients (cohort 2) consisting of the remaining admissions that occurred over the study period. This group was larger than cohort 1, so we reviewed a random 15% sample of all screened-positive charts. For cohort 2, to reduce the sample ultimately requiring physician review, we exclusively used the combined screening strategy of electronic screening followed by manual discharge summary review. To further limit our subset of charts for review, only screened-positive charts judged as having a "possible AE" underwent formal chart reviews. Screened-positive charts judged as having "no AE" on manual discharge summary review were not reviewed. Because we did not review screened-negative charts in cohort 2, we did not have information on false-positive or true-negative alerts and were unable to calculate sensitivity and specificity of the combined screening strategy in this patient sample. In this second set, we reviewed 35% (145 of 413) of the charts.

Table 1 ■ Adverse Event Concepts and Trigger Words

Adverse Event Category	Event Concept	Trigger Word	Number of Times Associated with an AE	Number of Times Fired	Positive Predictive Value	
Previous failure of or untoward result from medical management	Fluid overload/iatrogenic pulmonary edema	"wet"	0	0	NA	
		"overload"	7	10	70%	
		"volume"	4	14	29%	
		"hypervolemia"	0	0	NA	
		"failure"	16	71	23%	
	Delirium		"mental status"	9	24	38%
			"hallucinations"	2	3	67%
			"agitation"	3	3	100%
			"delirium"	5	9	56%
			"lethargic"	1	2	50%
			"oversedation"	1	1	100%
			"sedated"	0	2	0
			"nonresponsive"	0	0	NA
			"unresponsive"	1	8	11%
			"somnolent"	4	1	80%
	Hypotension		"hypotension"	14	30	47%
			"transfusion"	13	27	48%
			"dropped"	6	19	32%
			"fluid resuscitation"	1	2	50%
			"required fluids"	0	0	NA
			"packed RBC"	0	0	NA
			"drop of"	0	0	NA
	Respiratory failure		"desaturation"	3	8	38%
			"hypoxia"	1	9	11%
			"hypoxemia"	0	4	0%
			"shortness of breath"	6	23	26%
			"respiratory distress"	1	7	14%
			"respiratory failure"	1	3	33%
	Error		"error"	1	1	100%
			"accident"	0	5	0%
			"accidentally"	1	2	50%
			"mistake"	0	0	NA
			"complication"	0	0	NA
"mistakenly"			0	0	NA	
"complicated"			28	42	67%	
Iatrogenic hyperglycemia		"DKA"	0	0	NA	
		"hyperglycemia"	0	0	NA	
Iatrogenic hypoglycemia		"D50"	0	0	NA	
		"low sugars"	0	0	NA	
		"hypoglycemia"	0	0	NA	
Acute renal failure		"renal failure"	16	30	53%	
Deep vein thrombosis		"deep vein thrombosis"	0	1	0%	
Hospital-incurred trauma	Fall	"fall"	2	6	25%	
		"slipped"	0	0	NA	
		"fell"	3	8	38%	
		"syncope"	0	0	NA	
	Decubitus ulcer		"decubiti"	0	0	NA
			"decubitus"	2	2	100%
			"bed sore"	0	0	NA
			"pressure sore"	0	0	NA
			"heel ulcer"	1	1	100%
			"skin ulcer"	0	0	NA
	Untoward drug reaction in hospital	Adverse drug event	"allergic"	0	1	0
			"rash"	13	19	68%
			"drug eruption"	0	0	NA

Continued

Table 1 ■ Continued

Adverse Event Category	Event Concept	Trigger Word	Number of Times Associated with an AE	Number of Times Fired	Positive Predictive Value
		"overdose"	1	1	100%
		"subtherapeutic"	0	0	NA
		"supratherapeutic"	0	0	NA
		"OD"	0	0	NA
		"polypharmacy"	0	0	NA
		"discontinued"	36	99	36%
Transfer from another acute care hospital	Hospital-to-hospital transfer	"transfer"	5	33	15%
		"transferred"	16	82	20%
Transfer from general to acute care	Transfer to ICU	"CCU"	1	9	10%
		"ICU"	0	7	0
		"MICU"	1	5	17%
		"NSICU"	0	0	NA
		"SICU"	0	0	NA
		"telemetry"	0	5	0
Return to operating room during index admission	Return to the operating room	"OR"	0	0	NA
		"surgery"	14	43	33%
		"operating room"	6	13	46%
Treatment or surgery for damaged organ subsequent to an invasive procedure	Operative or procedural complication	"perforation"	1	1	100%
		"pneumothorax"	0	8	0
		"PTX"	0	0	NA
		"dissection"	4	11	36%
		"chest tube"	2	6	33%
		"wound infection"	0	0	NA
Acute MI, CVA, or PE during or after an invasive procedure	Postoperative complication	"after surgery"	0	0	NA
		"after the operation"	0	0	NA
		"post-op"	0	0	NA
Death	Unexpected death	"died"	0	0	NA
		"expired"	0	0	NA
		"death"	0	0	NA
Cardiac or respiratory arrest	Cardiopulmonary arrest	"arrest"	0	0	NA
		"code"	0	0	NA
		"intubation"	0	0	NA
		"ACLS"	0	0	NA
Other undesired outcome	Nosocomial infections	"nosocomial"	0	0	NA
		"hospital acquired"	0	0	NA
		"TV sepsis"	0	0	NA
		"line infection"	0	0	NA
		"line sepsis"	0	0	NA
		"clostridium difficile"	0	0	NA
		"TV infection"	0	0	NA

Abbreviations: NA, not applicable; RBC, red blood cell; DKA, diabetic ketoacidosis; D50, 50% dextrose solution; OD, overdose; CCU, cardiac care unit; ICU, intensive care unit; MICU, medical intensive care unit; NSICU, neurosurgical intensive care unit; SICU, surgical intensive care unit; OR, operating room; PTX, pneumothorax; ACLS, Advanced Cardiac Life Support; IV, intravenous; MI, myocardial infarction; CVA, cerebrovascular accident; PE, pulmonary embolism.

Statistical Analysis

For our descriptions of the adverse event category, associated disability, severity, and preventability, we included all adverse events detected during the study. To calculate sensitivity and specificity of the screening tool, we included only one trigger word or AE per admission. For charts with multiple AEs, we selected the AE resulting in the greatest disability. For charts with multiple trigger words, we selected the trigger word that most likely represented an AE based on the manual review. Sensitivity of detection was

calculated by dividing the number of admissions with a trigger word and an AE by the total number of admissions with AEs. Specificity was determined by dividing the number of admissions with no trigger word and no AE by the total number of admissions with no AEs. Positive predictive value for the tool was determined by dividing the number of admissions with a trigger word and an AE by the number of admissions with trigger words.

We used the chi-square test for proportions and the t-test for means to compare the demographic information between

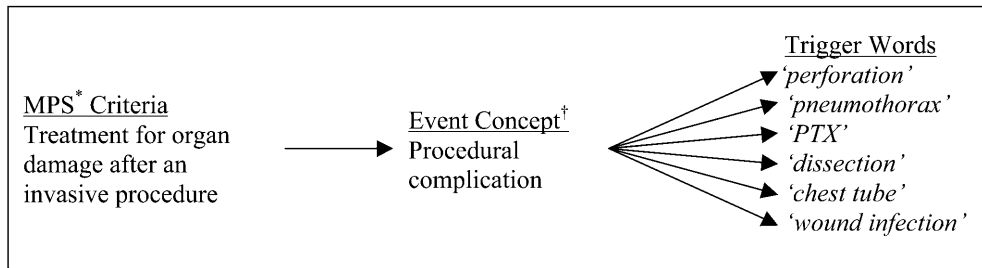


Figure 1. Trigger word mapping to MPS criteria. *MPS, Medical Practice Study. †Standard term represents a specific event concept. PTX, pneumothorax.

our randomly selected cohort 1 and the remaining cohort 2. To test for differences between the screening methodologies in our two cohorts, we used the chi-square test. Kappa statistics were performed to determine interrater reliability for judgments concerning injury causation, disability, severity, preventability, and discharge summary judgments. All statistical calculations were performed using SAS (version 8) statistical software (SAS Institute, Cary, NC).

Results

Sample Selection and Demographic Characteristics

There were 8,109 medical admissions during the study period. Forty percent (3,250 of 8,109) of these admissions

had completed discharge summaries. For patients admitted for less than 72 hours, 13% (629 of 4,918) had completed discharge summaries, while for patients admitted for more than 72 hours, 82% (2,621 of 3,191) had completed discharge summaries. Cohort 1 included 424 admissions involving 416 individual patients. The mean age in cohort 1 was 59.3 years, and the median length of stay was 6.35 days (range, 1 to 56 days). Forty-five percent (189 of 416) of the subjects were female and 79.7% (333 of 416) were white (Table 2).

Total Alerts Detected

The screening tool initially detected 953 alerts in 251 discharge summaries. After adjusting for multiple signals (the identical word appearing more than once in a single discharge summary), there were a total of 730 alerts in 251

Possible AE:

“During the procedure (AICD placement)...the patient’s blood pressure decreased that was found to be secondary to tamponade... patient was *transferred* to the Coronary Care Unit for monitoring.”

“The patient has a sacral *decubitus ulcer* that will need attention and regular dressing changes.”

No AE:

“After completing a seven day course the levofloxacin was *discontinued*.”

“The patient was then *transferred* to the medical floors.”

“...a follow up chest x-ray revealed no *pneumothorax*.”

Figure 2. Examples of trigger word judgments based on discharge summary review. Trigger words are in italics. AE, adverse event.

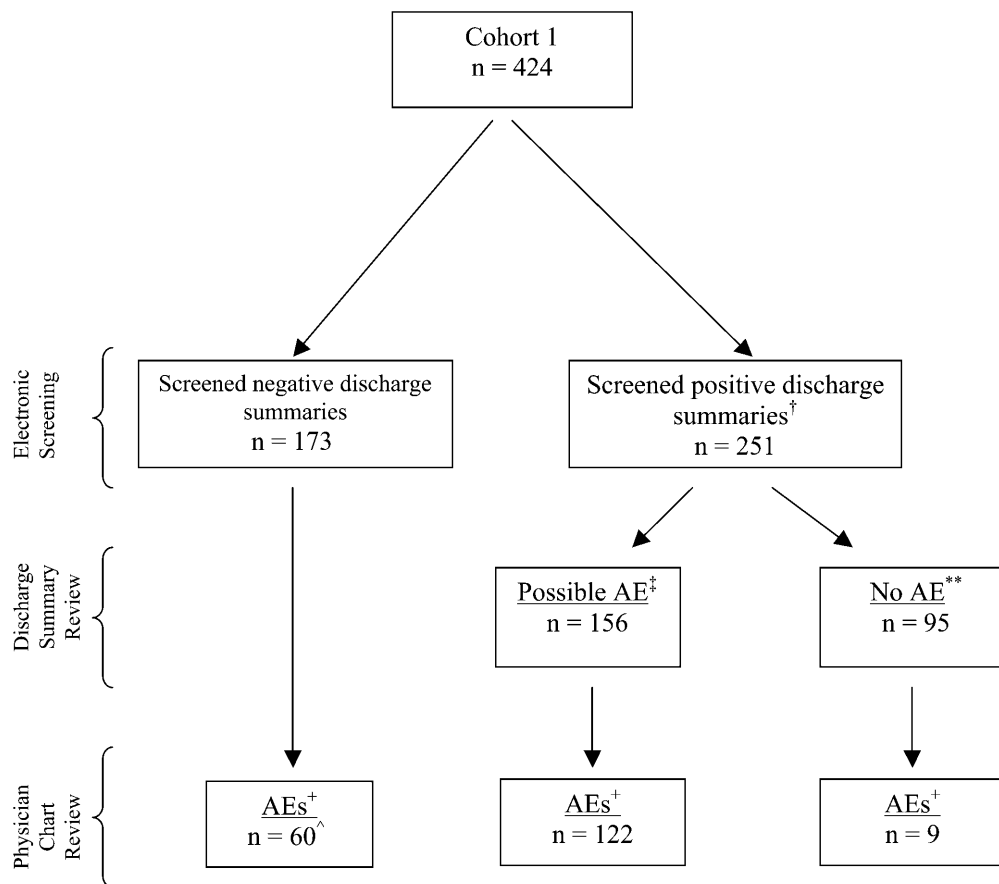


Figure 3. Flow diagram of patients, discharge summary review, and adverse events. ⁺Adverse event (AE) determined by manual review. [†]For this analysis, each discharge summary was allowed only one trigger word. If a discharge summary had both a “possible AE” and a “no AE” trigger word, it was considered a “possible AE.” [^]Number extrapolated from a random 25% sample. In sample 15, adverse events were detected in 43 reviewed charts for an adverse event rate in screened-negative charts of 35%. [‡]Discharge summaries that were reviewed and believed to contain an adverse event were considered “possible AE.” ^{**}Discharge summaries that were reviewed and believed to not contain an adverse event were considered “no AE.”

patient admissions. Forty-five percent (327 of 730) of these trigger word alerts were judged to represent “possible AE,” and 55% (403 of 730) were judged to represent “no AE” (Fig. 3). The kappa statistic judgment concerning the reliability of the rating of AEs based on discharge summary review was 0.50.

Adverse Events Detected

We found a total 204 AEs detected by the trigger words in 131 patient admissions within cohort 1. For descriptions regarding adverse event categories, disability, severity, preventability, and interrater reliability measures, we included all 204 identified AEs. Overall, the physician reviewers had good interrater reliability for the occurrence of an AE (kappa statistic = 0.77).

The most common adverse events detected by the screening tool were adverse drug events (Table 3), representing 52% of the AEs. The most common adverse drug event was neuropsychiatric, such as drug-induced delirium (30%), followed by renal failure (24%) and dermatologic/allergic (13%). The most frequent medications causing an AE were

cardiovascular drugs (18%), followed by anti-infectives and anticoagulants (13% each). The most common operative complication was postoperative bleeding (20%), followed by postoperative pneumonia and pulmonary embolism (15% each). The most common medical procedural complication detected was nonwound infection (30%), followed by bleeding (24%).

Forty-three percent (87 of 204) of all the adverse events detected by the physician reviewers resulted in either a disability at discharge or a prolonged hospitalization (Table 4). Six percent (13 of 204) of these AEs resulted in either permanent disability or death, while 24% (49 of 204) of the AEs detected were associated with either a laboratory abnormality only or less than one day of symptoms (Table 5). The kappa statistics for reviewer judgments of disability and severity were 0.38 and 0.62, respectively. Eighty-four percent (172 of 204) of AEs detected occurred during the hospitalization, 9% (18 of 204) occurred in another hospital before institutional transfer, and the remaining AEs (7%) occurred before hospitalization.

Of the 204 adverse events, 23 (11%) were judged preventable AEs. Forty-eight percent (11 of 23) of these events

Table 2 ■ Patient Demographics of Cohort 1 and Cohort 2

Characteristics	<i>n</i> = 424 (%)	<i>n</i> = 2,826 (%)	p-value
Age, years (median)	62*	64†‡	0.01
Length of stay, days (median)	6	5**	< 0.0001
Gender (female)	189* (45%)	1246†¶ (54%)	0.002
Race			
White	333* (80%)	1698†§ (60%)	0.006
Black	26* (6%)	402†** (14%)	< 0.0001
Unknown	41* (10%)	26†** (0.9%)	< 0.0001
Discharge service			
Oncology/bone marrow transplant	66 (16%)	593§ (21%)	0.006
Medicine, general	79 (19%)	1132** (40%)	< 0.0001
Medicine, subspecialty†† (except cardiology)	57 (13%)	337 (12%)	NS ⁺
Cardiology	184 (43%)	546** (19%)	< 0.0001
Neurology	38 (9%)	218** (8%)	< 0.0001

* = *n* of 416 (only earliest admission considered for patients with multiple admissions during the study period).

† = *n* of 2,317 (only earliest admission considered for patients with multiple admissions during the study period).

‡p-value of 0.01.

§p-value of 0.006.

¶p-value of 0.002.

**p-value < 0.0001.

††Subspecialty services included pulmonology, gastroenterology, rheumatology, nephrology, and endocrinology.

+Not significant.

resulted in a disability at discharge or a prolonged hospitalization. The kappa statistic for preventability judgments was 0.52.

To determine the performance of our screening tool on cohort 1, we adjusted for multiple AEs within a single admission. Physicians judged 191 admissions to have AEs. The screening tool detected 131 of the actual AEs and had a sensitivity of 69% (95% confidence interval [CI] 62, 76) and a specificity of 48% (95% CI 42, 54) in this population (Table 6). The positive predictive value of the screening tool was 52% (95% CI 46, 58).

When the electronic screening tool was couple with a manual discharge summary review, 122 admissions with an AE were detected. This combined screening methodology had a sensitivity of 64% (95% CI 61, 67), a specificity of 85%

Table 3 ■ Categories of All Adverse Events Detected within Cohort 1

Category	<i>n</i> = 204* (%)
Diagnostic error	15 (7.4%)
Operative complication	20 (9.8%)
Medical procedural complication	54 (26.5%)
Drug-related	105 (51.5%)
Therapeutic error	7 (3.4%)
Falls	3 (1.5%)

*Some admissions had more than one AE within the chart; for this analysis all AEs are included.

Table 4 ■ Disability Score for All Adverse Events Detected within Cohort 1

Disability	<i>n</i> = 204* (%)
Disability at discharge or prolonged hospitalization	87 (43%)
Neither disability at discharge nor prolonged hospitalization	117 (57%)

*Some admissions had more than one AE within the chart; for this analysis all AEs are included.

(95% CI 80, 90), and a positive predictive value of 78% (95% CI 71, 85).

Second Cohort Results

In cohort 2, we applied the tool to 2,826 patients and detected 4,594 trigger words. We found 112 adverse events within the random sample of 413 patient admissions reviewed. The kappa statistic regarding judgments concerning the occurrence of an AE in this cohort was 0.57. The overall positive predictive value of electronic screening followed by manual discharge summary review to detect an AE within cohort 2 was 84% (95% CI 81, 87). The difference in the positive predictive value between the electronic-only screening method in cohort 1 and the electronic-plus-manual screening method in the cohort 2 was statistically significant ($p < 0.0001$). There were no statistically significant differences in the electronic-plus-manual screening methodology between cohort 1 and cohort 2 ($p = 0.19$).

False-negative Hits

In cohort 1, there were 15 charts that physicians identified as having an AE that did not have a trigger word within the discharge summary. On review of the accompanying discharge summaries, we identified potential trigger words that could have been used to identify the AE as well as the false-negative alerts that had resulted from lexical variants (Table 7).

Discussion

Medical discharge summaries at the study institution contain useful information for the identification of adverse events. Electronically screening discharge summaries is a feasible and potentially efficient means of detecting adverse medical events. While screening charts electronically is advantageous in that it requires less time compared

Table 5 ■ Severity of All Adverse Events Detected within Cohort 1

Severity	<i>n</i> = 204* (%)
Laboratory abnormality or one day of symptoms	49 (24)
More days of symptoms or nonpermanent disability	142 (70)
Permanent disability or death	13 (6)

*Some admissions had more than one AE within the chart; for this analysis all AEs are included.

Table 6 ■ Performance Characteristics of Varying Detection Methods for the Electronic Screening Tool

	Electronic Only* (Cohort 1)			Electronic + Manual [^] (Cohort 1)			Electronic + Manual [^] (Cohort 2)		
	AE+	AE-	Total	AE+	AE-	Total	AE+	AE-	Total
Screened-positive summary	131	120	251	122	34	156	122	23	145
Screened-negative summary†	60	113	173	69	199	268	ND	ND	268
Total	191	233	424	191	233	424	ND	ND	413
Sensitivity (95% CI)			69% (62, 76)			64% (61, 67)			ND
Specificity (95% CI)			48% (42, 54)			85% (80, 90)			ND
Positive predictive value (95% CI)			52% (46, 58)			78% (71, 85)			84% (81, 87)

Abbreviations: ND, not determined in study; AE, adverse event.

*Screened-positive summary includes discharge summaries with a trigger word present, and screened-negative summary includes discharge summaries without trigger words.

[^]Screened-positive summary includes only discharge summaries with a trigger word that is considered a "possible AE" present, and screened-negative summary includes discharge summaries without trigger words or with trigger words considered "no AE."

†For the screened-negative row, we reviewed a random 25% of these records. We detected 15 adverse events in 43 medical records. We extrapolated these numbers to the entire sample.

with manual reviews, our current tool was not very specific and resulted in numerous false-positive results. Thus, improvement must be made in the electronic tool itself, or an intermediate manual review process must be utilized to make the tool practical on an organizational level. Despite its poor specificity and low positive predictive value, our screening tool still performed better than many other electronic tools used for detecting adverse events.

Most prior studies using electronically stored information to search for medical complications have utilized administrative data such as trigger alerts. For example, Bates et al.²³ studied a subset of generic screens that were present using billing data. With five screens, they were able to detect 50% of the events determined through manual review with a positive predictive value of 20%. By eliminating a poorly performing billing screen, the positive predictive value of the tool increased to 30%, but this approach detected only 26% of the events detectable by manual review. Both strategies cost significantly less per admission screened than manual review.

Another attempt to identify complications through administrative data was the Complications Screening Program (CSP)—a computer algorithm that searches medical and surgical claims data for potential complications.^{24,25} Such

claims are available for all discharges, making this approach potentially very powerful. However, while the tool performed well for surgical admissions, it did not perform as well for medical admissions.²⁶ In medical admissions, only 15.7% of the alerts had potential quality problems.¹⁹ A limitation of this tool has been its reliance on ICD-9-CM codes as the searching criteria. For medical patients in the CSP study, 30% had no objective evidence supporting the ICD-9-CM codes within the medical record.²⁷ Because of these concerns, the authors concluded that the CSP was not valid as a "stand-alone" test to search for complications in medical patients. However, ICD-9-CM codes may still be valuable if combined with other data.

Other investigators have utilized billing or generic screening criteria to search for quality problems within medical admissions and have identified weaknesses including low sensitivities and concerns about validity.²⁸⁻³⁰ Despite these concerns, the need for an efficient and inexpensive means to detect quality of care concerns and patient injuries makes an electronic screening approach attractive. This is especially true because electronic detection approaches already have been used successfully to detect adverse drug events^{11,16-18} and nosocomial infections.³¹ In one study, electronic surveillance for nosocomial infections had a sensitivity of 90%, while manual surveillance had a sensitivity of 76%.³²

Table 7 ■ Adverse Events Not Detected by the Screening Tool and Potential Trigger Word

Event Description	Potential Trigger Word
Gastrointestinal bleed on heparin	"bleeding"
Postoperative stroke	"infarct"
Transaminitis secondary to phenytoin	"elevated LFTs"
Mental status change on medications resulting in a fall	"falls" (plural variant)
Rash on ceftazidime	"skin biopsy"
Mental status change as a result of lithium	"over sedation" (variant)
Procedural complication resulting in shock	"pressor support"
Intracranial bleed after thrombolysis for myocardial infarction	"intracranial"
Flank hematoma and gross hematuria on anticoagulants	"hematoma"
Femoral artery thrombus after catheterization procedure	"thrombus"

A strength of our study results from the use of clinical narrative rather than administrative data to screen for AEs. Narratives, such as discharge summaries, contain important data, yet represent a largely untapped source of information. Searching clinical narratives has several advantages compared with searching billing data. First, events documented within discharge summaries are recorded for clinical communication and not billing reasons. This is advantageous in that it helps avoid the potential inaccuracies of ICD-9-CM coding.³³ Second, administrative screening frequently detects events that occurred before admission.²⁸ Studies with the CSP found that 58% of conditions detected in medical patients were present on admission.²⁶ Eighty-four percent of the events detected by our tool occurred during the hospitalization.

An important finding of our study is that discharge summaries contain useful information for detecting adverse events. Adverse events are difficult to detect because of the reluctance of clinicians to report them, and medical record review has been an important source of information on adverse events. We are unaware of any prior studies that have evaluated whether the discharge summary, a concise summary of the medical record, would also contain information regarding AEs.

In this study, 45% (191 of 424) of patients experienced an AE. We believe there are two reasons that this proportion is substantially higher than most previously reported figures (for example, 3.7% in the MPS⁷ and 2.9% in the Colorado-Utah study³⁴). First, we used a more inclusive definition for adverse events, which effectively increased our numerator. If we exclude events that would not have met the MPS definition (72 adverse events) our number of detected events decreases from 191 to 119, and the event rate becomes 28.1% (119 of 424). Second, our approach searched only the records of patients with completed discharge summaries. At our institution, 59.7% of patient admissions do not have a discharge summary, predominantly because of a length of stay less than 72 hours. In the pilot set, if patients without discharge summaries had been included, the actual sample size of our denominator would be approximately 1,052 patient admissions rather than 424. With this sample size, the rate of AEs per admission occurring in our initial sample would be 11% (119 of 1,052). This percentage is within the range of AEs supported by the literature, especially at teaching institutions, which have a high rate of adverse events, with a low proportion that are preventable.^{7,37}

Although the screening tool performed moderately well, its poor specificity resulted in numerous false-positive hits, which reduces the overall effectiveness of the tool. Several changes might prove beneficial. The overall performance of the screening tool could be improved through the removal of poorly performing trigger words, such as *shortness of breath* or *transfer*. While we identified trigger words that could detect the missed AEs from our cohort (Table 7), the addition of these trigger words could have an overall detrimental effect on the screening tool by decreasing specificity. A major improvement to the performance of the tool could result from developing a rule base that incorporates negative modifiers into the alert algorithm. For instance, in

our study, if the computer searched for the trigger word *pneumothorax* and came across the sentence "the patient had no pneumothorax," this would have been alerted as a positive screen. Relatively simple electronic tools have been developed that could be used to detect these trigger words with negative modifiers.^{38,39} Once detected, these false-positive alerts could be reduced, improving the tools overall specificity and reducing unnecessary chart review. In the initial cohort, approximately 9% of the trigger words included a negative modifier indicating that no AE was present.

More sophisticated parsing could be performed using natural language processing (NLP). Natural language processors are programs that can parse free-text narrative and, through using basic grammatical rules and knowledge about the domain, extract concepts represented by the narrative.^{40,42} Thus, NLP has several advantages over free-text searching in that NLP addresses negation and lexical variation of terms and incorporates context from the sentence in an attempt to understand the meaning of a term.

Electronically screening for AEs has important implications. Adverse events are difficult to study; thus, our knowledge concerning AEs is incomplete. Currently, researchers and quality managers interested in adverse events must perform many manual chart reviews to find AEs. The cost and time required for these studies make them difficult to perform. An automated method of detecting charts with a high likelihood of having an AE could significantly reduce the cost and time associated with these studies.

Because of the high number of false-positive results, many of the time gains acquired from electronic screening using only keyword queries could be reduced by unnecessary chart reviews. Adding a manual discharge summary review step into the review process markedly improved the test's specificity and positive predictive value in our study and would probably be required using a keyword-only searching strategy. Advances in the electronic screening methods, however, could probably result in a tool that could completely eliminate manual screening.

This study has several limitations. First, we applied the tool on discharge summaries from only one teaching hospital. The tool might not perform as well in community hospitals. Second, we were able to search only charts with an accompanying discharge summary; thus, the screening tool did not evaluate patients with shorter stays. Prior studies have indicated that longer length of stay is associated with the presence of adverse events.^{35,36} Thus, we would expect our group to be at higher risk for adverse events. This limitation could be addressed potentially through screening other documents that are present regardless of length of stay (operation notes, progress notes, nursing notes). Third, we relied on implicit physician review to make judgments concerning AEs. Several studies have documented the limitations of this method.⁴³⁻⁴⁵ In this study, physician agreement was good for the presence of an AE (kappa statistic = 0.77) but only fair for judgments concerning preventability (kappa statistic = 0.52).

Another limitation of our study was the difference between our two cohorts. Differences in the populations selected for cohort 1 and cohort 2 likely resulted from our randomization process. In cohort 1, we sampled subjects who were new to the study institution. As a result, a larger percentage of derivation set patients was admitted to the cardiology service. In cohort 1, 53% (69 of 131) of patients with AEs were discharged from the cardiology service as opposed to 17% (20 of 118) in cohort 2 ($p < 0.0001$). This discrepancy likely influenced the types of AEs detected within the cohorts. Furthermore, procedural complications were responsible for 26% of the total AE rate in cohort 1, versus 10% in cohort 2 ($p = 0.0005$). Despite these differences, the combination of electronic screening and manual review produced similar results in both cohorts.

Conclusions

Electronic screening discharge summaries for the presence of potential adverse medical events is feasible, but simple keyword queries are not specific enough to be practical. A combination of electronic screening with manual review would improve specificity and reduce false-positive results but be impractical to implement. However, implementing a more sophisticated keyword search that incorporates natural language processing techniques may improve specificity enough not to require manual review. Electronically screening narrative text within medical discharge summaries is a promising routine method for detecting a wide range of adverse events.

References ■

- Shojania KG, Duncan BW, McDonald K, Wachterm RM. Making HealthCare Safe: A Critical Analysis of Patient Safety Practices: Evidence Report/Technical Assessment #43; Rockville, MD: Agency for Healthcare Research and Quality, 2001.
- Corrigan JM, Kohn LT, Donaldson MS, Maguire SK, Pike KC. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press, 2001.
- Kohn LT, Corrigan JM, Donaldson MS, McKay T, Pike KC. To Err Is Human. Washington, DC: National Academy Press, 2000.
- Weingart SN, Wilson RM, Gibberd RW, Harrison B. Epidemiology of medical error. *BMJ*. 2000;320:774-7.
- 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.
- 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.
- 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.
- Langley GJ, Nolan KM, Nolan TW, Norman CL, Provost LP. The Improvement Guide: A Practical Approach to Enhance Organization Performance. San Francisco: Jossey-Bass, 1996.
- O'Neil AC, Petersen LA, Cook EF, Bates DW, Lee TH, Brennan TA. Physician reporting compared with medical-record review to identify adverse medical events. *Ann Intern Med*. 1993;119:370-6.
- 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. *Jt Comm J Qual Improv*. 1995;21:541-8.
- 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.
- Goldman R. Development of a Veterans Administration occurrence screening program. *QRB Qual Rev Bull*. 1989;15:315-9.
- Brennan TA, Localio RJ, Laird NL. Reliability and validity of judgments concerning adverse events suffered by hospitalized patients. *Med Care*. 1989;27:1148-58.
- Wolff AM. Limited adverse occurrence screening. A medical quality control system for medium sized hospitals. *Med J Aust*. 1992;156:449-52.
- Bates DW, Leape LL, Cullen DJ, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA*. 1998;280:1311-6.
- Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. *JAMA*. 1991;266:2847-51.
- Evans RS, Pestotnik SL, Classen DC, Bass SB, Burke JP. Prevention of adverse drug events through computerized surveillance. *Proc Annu Symp Comput Appl Med Care*. 1992:437-41.
- Raschke RA, Gollihare B, Wunderlich TA, et al. A computer alert system to prevent injury from adverse drug events: development and evaluation in a community teaching hospital. *JAMA*. 1998;280:1317-20.
- 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.
- Teich JM, Glaser JP, Beckley RF, et al. The Brigham Integrated Computing System (BICS): advanced clinical systems in an academic hospital environment. *Int J Med Inform*. 1999;54:197-208.
- Hiatt HH, Barnes BA, Brennan TA, et al. A study of medical injury and medical malpractice. *N Engl J Med*. 1989;321:480-4.
- Stevens G, Bennett J. Clinical audit—occurrence screening for QA. *Health Serv Manage*. 1989;85:178-81.
- Bates DW, O'Neil AC, Petersen LA, Lee TH, Brennan TA. Evaluation of screening criteria for adverse events in medical patients. *Med Care*. 1995;33:452-62.
- Iezzoni LI, Foley SM, Heeren T, et al. A method for screening the quality of hospital care using administrative data: Preliminary validation results. *QRB Qual Rev Bull*. 1992;18:361-71.
- Iezzoni LI, Daley J, Heeren T, et al. Using administrative data to screen hospitals for high complication rates. *Inquiry*. 1994;31:40-55.
- Lawthers AG, McCarthy EP, Davis RB, Peterson LE, Palmer RH, Iezzoni LI. Identification of in-hospital complications from claims data. Is it valid? *Med Care*. 2000;38:785-95.
- McCarthy EP, Iezzoni LI, Davis RB, et al. Does clinical evidence support ICD-9-CM diagnosis coding of complications? *Med Care*. 2000;38:868-76.
- Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med*. 1997;127(8 Pt 2):666-74.
- Hayward RA, Bernard AM, Rosevear JS, Anderson JE, McMahon LF, Jr. An evaluation of generic screens for poor quality of hospital care on a general medicine service. *Med Care*. 1993;31:394-402.
- Evans RL, Connis RT. Risk screening for adverse outcomes in subacute care. *Psychol Rep*. 1996;78(3 Pt 1):1043-8.
- Evans RS, Gardner RM, Bush AR, et al. Development of a computerized infectious disease monitor (CIDM). *Comput Biomed Res*. 1985;18:103-13.

32. Evans RS, Larsen RA, Burke JP, et al. Computer surveillance of hospital-acquired infections and antibiotic use. *JAMA*. 1986;256:1007-11.
33. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. *Ann Intern Med*. 1993;119:844-50.
34. 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.
35. Andrews LB, Stocking C, Krizek T, et al. An alternative strategy for studying adverse events in medical care. *Lancet*. 1997;349:309-13.
36. Bates DW, Pappius E, Kuperman GJ, et al. Using information systems to measure and improve quality. *Int J Med Inform*. 1999;53(2-3):115-24.
37. Localio AR, Lawthers AG, Brennan TA, et al. Relation between malpractice claims and adverse events due to negligence. Results of the Harvard Medical Practice Study III. *N Engl J Med*. 1991;325:245-51.
38. Mutalik PG, Deshpande A, Nadkarni PM. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. *J Am Med Inform Assoc*. 2001;8:598-609.
39. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*. 2001;34:301-10.
40. 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.
41. Friedman C, Hripcsak G. Natural language processing and its future in medicine. *Acad Med*. 1999;74:890-5.
42. Friedman C. Towards a comprehensive medical language processing system: methods and issues. *Proc AMIA Annu Fall Symp*. 1997:595-9.
43. Rubin HR, Rogers WH, Kahn KL, Rubenstein LV, Brook RH. Watching the doctor-watchers. How well do peer review organization methods detect hospital care quality problems? *JAMA*. 1992;267:2349-54.
44. Localio AR, Weaver SL, Landis JR, et al. Identifying adverse events caused by medical care: Degree of physician agreement in a retrospective chart review. *Ann Intern Med*. 1996;125:457-64.
45. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA*. 2001;286:415-20.