Automated Identification of Postoperative Complications Within an Electronic Medical Record Using Natural Language Processing

Harvey J. Murff, MD, MPH
Fern FitzHenry, RN, PhD
Michael E. Matheny, MD, MSc, MPH
Nancy Gentry, RN
Kristen L. Kotter, MS
Kimberly Crimin, PhD
Robert S. Dittus, MD, MPH
Amy K. Rosen, PhD
Peter L. Elkin, MD
Steven H. Brown, MD, MSc
Theodore Speroff, PhD

Context Currently most automated methods to identify patient safety occurrences rely on administrative data codes; however, free-text searches of electronic medical records could represent an additional surveillance approach.

Objective To evaluate a natural language processing search–approach to identify postoperative surgical complications within a comprehensive electronic medical record.

Design, Setting, and Patients Cross-sectional study involving 2974 patients undergoing inpatient surgical procedures at 6 Veterans Health Administration (VHA) medical centers from 1999 to 2006.

Main Outcome Measures Postoperative occurrences of acute renal failure requiring dialysis, deep vein thrombosis, pulmonary embolism, sepsis, pneumonia, or myocardial infarction identified through medical record review as part of the VA Surgical Quality Improvement Program. We determined the sensitivity and specificity of the natural language processing approach to identify these complications and compared its performance with patient safety indicators that use discharge coding information.

Results The proportion of postoperative events for each sample was 2% (39 of 1924) for acute renal failure requiring dialysis, 0.7% (18 of 2327) for pulmonary embolism, 1% (29 of 2327) for deep vein thrombosis, 7% (61 of 866) for sepsis, 16% (222 of 1405) for pneumonia, and 2% (35 of 1822) for myocardial infarction. Natural language processing correctly identified 82% (95% confidence interval [CI], 67%-91%) of acute renal failure cases compared with 38% (95% CI, 25%-54%) for patient safety indicators. Similar results were obtained for venous thromboembolism (59%, 95% CI, 44%-72% vs 46%, 95% CI, 32%-60%), pneumonia (64%, 95% CI, 58%-70% vs 5%, 95% CI, 3%-9%), sepsis (89%, 95% CI, 78%-94% vs 34%, 95% CI, 24%-47%), and postoperative myocardial infarction (91%, 95% CI, 78%-97%) vs 89%, 95% CI, 74%-96%). Both natural language processing and patient safety indicators were highly specific for these diagnoses.

Conclusion Among patients undergoing inpatient surgical procedures at VA medical centers, natural language processing analysis of electronic medical records to identify postoperative complications had higher sensitivity and lower specificity compared with patient safety indicators based on discharge coding.

JAMA. 2011;306(8):848-855

©2011 American Medical Association. All rights reserved.
ity surveillance tool. First, administrative data are readily available, easily accessible, and inexpensively captured. However, they are not without limitations. Concerns exist about the validity of administrative codes,7,9 and it can be difficult to determine from discharge diagnostic codes whether a disease entity existed before the patient was hospitalized or occurred during the hospital admission.10,11

With the rapid expansion of electronic medical record (EMR) use, along with increased federal support for health care information technology, a far richer source of clinical information regarding hospital-related safety events has emerged.12 The development of automated approaches, such as natural language processing, that extract specific medical concepts from textual medical documents that do not rely on discharge codes offers a powerful alternative to either unreliable administrative data or labor-intensive, expensive manual chart reviews.13 Nevertheless, there have been few studies investigating natural language processing tools for the detection of adverse events.14,15 It is not known whether a surveillance approach based on language processing searches of free-text documents will perform better than currently used tools based on administrative data. The purpose of this study was to evaluate a language-processing-based approach to identify postoperative complications within a multihospital health care network using the same EMR. We hypothesized that the language processing searches would better detect surgical complications than the patient safety indicators identified from administrative discharge information.

METHODS

Setting

The study population included a randomly selected sample of Veterans Affairs Surgical Quality Improvement Program (VASQIP)–reviewed surgical inpatient admissions to 6 Veterans Health Administration medical centers across 3 states between fiscal years 1999 and 2006. At each study site the institutional review board approved the study and granted a waiver for the need to obtain informed consent for the use of patient data. Self-reported patient race/ethnicity information was obtained from demographic files.

Database

We linked VASQIP cases to the Veterans Affairs (VA) patient treatment file, an administrative database containing records on all veterans discharged from VA facilities. Linkage was based on the patients’ identifier code and having the surgical procedure date fall between a patient treatment file admission and discharge date. Narrative clinical notes such as discharge summaries, progress notes, operative notes, microbiology reports, imaging reports, and outpatient visit notes were obtained from the Veterans Health Information System and Technology Architecture. In addition we acquired structured data tables including demographic data, vital sign information, pharmacy data files, and laboratory results.

Measures

VASQIP. As part of the VASQIP protocol during the time of this study, only major noncardiac surgical procedures were eligible for review. In addition, surgical procedures were excluded if performed under local nerve blocks, were low volume, or were low risk. The VASQIP nurse reviewers underwent extensive training to prospectively collect clinical information on surgical cases.16 These nurse reviewers tracked eligible surgical cases for 30 days after surgery and recorded the occurrence of 1 of 20 prespecified postoperative complications. We focused on the 6 postoperative complications (acute renal failure requiring dialysis, sepsis, deep vein thrombosis, pulmonary embolism, myocardial infarction, and pneumonia) that are also included as patient safety indicator events (TABLE 1). At the time of this study, the patient safety indicators for myocardial infarction and pneumonia were considered experimental. The nurse reviewer interrater reliability on postoperative occurrences has been estimated at 0.73 for acute renal failure requiring dialysis, 0.65 for pneumonia, 0.60 for myocardial infarction, 0.81 for deep vein thrombosis, and 0.89 for pulmonary embolism.17

Patient Safety Indicators. To simultaneously evaluate the approaches of natural language processing and patient safety indicators, we applied additional exclusion rules to the VASQIP cohort in order to match a previously published method for applying patient safety indicator software to VA databases.9,18 First, the patient safety indicators are hospital-based, whereas outcomes for VASQIP are surgery-based (ie, may include inpatient and outpatient outcomes). As such, we matched individual surgeries to specific hospitalizations and limited our sample to patients with a surgical episode occurring during their hospitalization. Because the patient safety indicators were designed to detect potentially preventable adverse occurrences, specific exclusion criteria were created to eliminate patients who were at high risk or had a greater likelihood of an adverse event due to preexisting comorbid illnesses or other circumstances. Thus, we applied specific safety indicator numerator and denominator exclusion rules for each safety indicator event to the VASQIP database as originally described by the Agency for Healthcare Research and Quality to obtain our analytic sample for each study outcome9 (FIGURE 1). In addition, as the safety indicators were developed to identify complications occurring during hospitalizations, we excluded any postoperative complications identified by VASQIP that had occurred after the patients’ hospital discharge. Finally, the safety indicator combined both deep vein thrombosis and pulmonary embolism into a single category (venous thromboembolism), and as such we combined these outcomes.

Natural Language Processor. The Multi-threaded Clinical Vocabulary Server natural language processor system19 was used to index the free text rec-
Acute Pneumonia

Pneumonia must meet 1 of the following 2 criteria:

**Criterion 1:** Rales or dullness to percussion on physical examination of chest and any of the following: (1) new onset of purulent sputum or change in character of sputum, (2) organism isolated from blood culture, or (3) isolation of pathogen from specimen obtained by transtracheal aspirate, bronchial brushing, or biopsy

**Criterion 2:** Chest radiographic examination shows new or progressive infiltrate, consolidation, cavitation, or pleural effusion plus any of the following: (1) new onset of purulent sputum or change in character of sputum; (2) organism isolated from blood culture; (3) isolation of pathogen from specimen obtained by transtracheal aspirate, bronchial brushing, or biopsy; (4) isolate of virus or detection of viral antigen in respiratory secretions; (5) diagnostic single antibody titer or 4-fold increase in paired serum samples for pathogen; or (6) histopathologic evidence of pneumonia

**EMR Measures.** Source documents for this study included narrative clinical notes, such as progress notes, consultant notes, imaging reports, microbiology reports, and discharge summaries. Some documents such as electrocardiograms and some types of imaging reports were not in a machine-readable format within the EMR and were thus unable to be processed by the language processing tool. This occurred when documents were scanned.

---

**Table 1.** Patient Safety Indicator and Veterans Health Administration Surgical Quality Improvement Program Outcomes Definitions

<table>
<thead>
<tr>
<th>Postoperative Outcome</th>
<th>VASQUIP Definition</th>
<th>Patient Safety Indicator Definition</th>
</tr>
</thead>
</table>
| Acute renal failure   | A patient who did not require dialysis preoperatively experienced worsening of renal dysfunction postoperatively requiring hemodialysis, peritoneal dialysis, hemofiltration, or ultrafiltration | Discharged cases meeting the inclusion and exclusion rules with ICD-9-CM codes for acute renal failure and accompanied by a procedure code for dialysis (5965.0, 5496)

Exclusions: Preexisting (principal diagnosis or secondary diagnosis present on admission, if known) acute renal failure or chronic renal failure; or acute renal failure for which a procedure for dialysis occurs before or on the same day as the first operating room procedure; or both with a secondary diagnosis code for acute renal failure and a principle diagnosis of acute myocardial infarction, cardiac arrhythmia, cardiac arrest, shock, hemorrhage, or gastrointestinal hemorrhage

| Pulmonary embolism or deep vein thrombosis | Pulmonary embolism is diagnosed by a V/Q scan interpreted as high probability of pulmonary embolism or a positive pulmonary arteriogram or positive spiral CT examination. The identification of a new blood clot or thrombus within the venous system, which may be coupled with inflammation. This diagnosis is confirmed by a duplex venogram or CT scan | Discharges among cases meeting the inclusion and exclusion rules with ICD-9-CM codes for deep vein thrombosis or pulmonary embolism in any secondary diagnosis field

Exclusions: Preexisting (principal diagnosis or secondary diagnosis present on admission, if known) deep vein thrombosis or pulmonary embolism for which a procedure for interruption of venous access occurs before or on the same date as the first operating room procedure

| Sepsis | Sepsis is clinically recognized by the presence of ≥2 of the following: (1) temperature >38°C or <36°C; (2) heart rate >90/min; (3) respiratory rate >20/min or PaCO2 <32 mm Hg (<4.3 kPa); (4) white blood cell count >12,000/µL, <4000/µL; or >10% immature forms; and (5) anion gap acidosis and 1 of the following: positive blood culture, clinical documentation of purulence, or positive culture from any site thought to be causative | Discharges among cases meeting the inclusion and exclusion rules with ICD-9-CM codes sepsis in any secondary diagnosis field

Exclusions: Preexisting (principal diagnosis or secondary diagnosis present on admission, if known) sepsis or infection; or any code for immunocompromised state or cancer, or with a length of stay <4 d

| Pneumonia | Pneumonia must meet 1 of the following 2 criteria:

**Criterion 1:** Rules or dullness to percussion on physical examination of chest and any of the following: (1) new onset of purulent sputum or change in character of sputum, (2) organism isolated from blood culture, or (3) isolation of pathogen from specimen obtained by transtracheal aspirate, bronchial brushing, or biopsy

**Criterion 2:** Chest radiographic examination shows new or progressive infiltrate, consolidation, cavitation, or pleural effusion plus any of the following: (1) new onset of purulent sputum or change in character of sputum; (2) organism isolated from blood culture; (3) isolation of pathogen from specimen obtained by transtracheal aspirate, bronchial brushing, or biopsy; (4) isolate of virus or detection of viral antigen in respiratory secretions; (5) diagnostic single antibody titer or 4-fold increase in paired serum samples for pathogen; or (6) histopathologic evidence of pneumonia | Discharges among cases meeting the inclusion and exclusion rules with ICD-9-CM codes aspiration pneumonia in any secondary diagnosis field

Exclusions: Preexisting (principal diagnosis present on admission, if known) seizure, trauma, drug overdose, or poisoning

| Acute myocardial infarction | A new transmural acute myocardial infarction occurring during surgery or within 30 d after surgery as manifested by new Q waves on ECG | Discharges among cases meeting the inclusion and exclusion rules with ICD-9-CM codes acute myocardial infarction in any secondary diagnosis field

Exclusions: Patients undergoing cardiac surgery

Abbreviations: CT, computed tomography; ECG, electrocardiograph; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; PaCO2, partial pressure of carbon dioxide; VASQUIP, Veterans Health Administration Surgical Quality Improvement Program; V/Q, ventilation and perfusion ratio.
in from outside sources or were rendered internally into PDF documents. Microbiology reports were transformed into structured data using regular expressions that recognized strings of text, in this case bacterial or fungal organisms but did not account for the syntax in which the term was identified. Search queries were also constructed using structured data from laboratory, pharmacy, and vital-sign databases. Because we were interested in postoperative occurrences, we only applied our search queries to documents and structured data with dates occurring after the date of the surgical procedure. In addition, to facilitate our comparison with the patient safety indicator, we applied our language processing queries to clinical narratives that occurred only within the inpatient stay and were directly associated with the surgical procedure.

Narrative clinical notes were initially processed by parsing each note and then by electronically identifying specific medical concepts and mapping these concepts to SNOMED-CT concepts. Text documents were also mapped to phrase and sentence strings allowing inclusion in the rules string searches of colloquial terms or ordering of expressions not yet recognized by the language processing tool vocabulary. The rule-building process involved clinical teams working from the VASQIP criteria to create specific search criteria (Table 1). These initial queries were tested on a training set of 6 randomly selected cases for each condition and 94 randomly selected controls. Search query development was an iterative 2-stage process in which the training documents were evaluated with individual queries followed by various combinations of queries. We also tested sequential testing strategies, where an initially highly sensitive single query would be applied to the analytic samples followed by a second round of more specific queries applied to all positive hits generated from the initial query. This process was repeated for each of the 6 postoperative complications. These rules were then applied to our patient sample, which excluded any cases or controls included within the training set.

**Analyses**

VASQIP-identified postoperative complications were considered the referent standard. We applied the natural language processing software and development query rule sets to determine the rate of language processing-detected complications and ran the patient safety indicator software version 3.1 to determine the rate of safety indicator events. We calculated sensitivity and specificity for the 6 adverse outcomes of interest. Because the safety indicators combined deep vein thrombosis and pulmonary embolism into a single event, we presented the results of our search algorithms for the 2 events both separately and combined. Sensitivity was defined as the proportion of the 6 postoperative events that were
POSTOPERATIVE COMPLICATIONS WITH ELECTRONIC MEDICAL RECORD

Table 2. Characteristics of Veterans Health Administration Surgical Quality Improvement Program Sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>VASQIP Hospitalizations, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR), y</td>
<td>64.5 (55-74)</td>
</tr>
<tr>
<td>Men</td>
<td>2837 (95)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>2450 (82)</td>
</tr>
<tr>
<td>Black</td>
<td>446 (15)</td>
</tr>
<tr>
<td>Other or unknown</td>
<td>78 (3)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>1228 (41)</td>
</tr>
<tr>
<td>Length of stay, median (IQR), d</td>
<td>8 (4-18)</td>
</tr>
<tr>
<td>American Society of Anesthesiologists classification</td>
<td></td>
</tr>
<tr>
<td>1. Normal healthy</td>
<td>23 (1)</td>
</tr>
<tr>
<td>2. Mild systemic disease</td>
<td>531 (18)</td>
</tr>
<tr>
<td>3. Severe systemic disease</td>
<td>1713 (58)</td>
</tr>
<tr>
<td>4. Severe systemic disease that is a constant threat to life</td>
<td>646 (22)</td>
</tr>
<tr>
<td>5. Moribund and not expected to survive without surgery</td>
<td>61 (2)</td>
</tr>
<tr>
<td>Category of surgical procedure</td>
<td></td>
</tr>
<tr>
<td>Cardiothoracic</td>
<td>212 (7)</td>
</tr>
<tr>
<td>General</td>
<td>1116 (38)</td>
</tr>
<tr>
<td>Gynecology</td>
<td>30 (1)</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>176 (6)</td>
</tr>
<tr>
<td>Orofacial maxillary surgery</td>
<td>1 (&lt;1)</td>
</tr>
<tr>
<td>Orthopedics</td>
<td>616 (21)</td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>44 (1)</td>
</tr>
<tr>
<td>Plastic</td>
<td>20 (1)</td>
</tr>
<tr>
<td>Urology</td>
<td>274 (9)</td>
</tr>
<tr>
<td>Vascular</td>
<td>423 (14)</td>
</tr>
<tr>
<td>Nonspecifica*</td>
<td>62 (2)</td>
</tr>
</tbody>
</table>

*Denotes a procedure that could be performed by multiple specialties.

Table 3. Comparison of a Natural Language Processing–Based Approach to the Agency for Healthcare Research and Quality Patient Safety Indicators in Identifying Postoperative Complications

<table>
<thead>
<tr>
<th>Occurrence</th>
<th>Event Rate</th>
<th>Test Characteristic</th>
<th>Natural Language Processing</th>
<th>Patient Safety Indicator</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute renal failure</td>
<td>39/1924</td>
<td>Sensitivity</td>
<td>0.82 (0.67-0.91)</td>
<td>0.38 (0.25-0.54)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specificity</td>
<td>0.94 (0.93-0.95)</td>
<td>1.00 (1.00-1.00)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pulmonary embolism/ deep vein thrombosis</td>
<td>46/2327</td>
<td>Sensitivity</td>
<td>0.59 (0.44-0.72)</td>
<td>0.46 (0.32-0.60)</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specificity</td>
<td>0.91 (0.90-0.92)</td>
<td>0.98 (0.98-0.99)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sepsis</td>
<td>61/866</td>
<td>Sensitivity</td>
<td>0.89 (0.78-0.94)</td>
<td>0.34 (0.24-0.47)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specificity</td>
<td>0.94 (0.93-0.96)</td>
<td>0.99 (0.98-0.99)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>222/1405</td>
<td>Sensitivity</td>
<td>0.64 (0.58-0.70)</td>
<td>0.05 (0.03-0.09)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specificity</td>
<td>0.95 (0.94-0.96)</td>
<td>0.99 (0.98-0.99)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>35/1822</td>
<td>Sensitivity</td>
<td>0.91 (0.78-0.97)</td>
<td>0.89 (0.74-0.96)</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specificity</td>
<td>0.95 (0.94-0.96)</td>
<td>0.99 (0.98-0.99)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
algorithms had a sensitivity of 0.33 (95% CI, 0.21-0.49) and a positive predictive value of 0.93 (95% CI, 0.69-1.00).

The patient safety indicator algorithm had a sensitivity of 0.46 (95% CI, 0.32-0.60) and specificity of 0.98 (95% CI, 0.98-0.99) for venous thromboembolism. The natural language processing approach for venous thromboembolism had a sensitivity of 0.99 (95% CI, 0.98-0.72) and a specificity of 0.91 (95% CI, 0.90-0.92; Figure 2 and eTable 2).

The patient safety indicator approach for pneumonia had a sensitivity of 0.05 (95% CI, 0.03-0.09) and a specificity of 0.99 (95% CI, 0.99-1.00; Figure 2 and eTable 3). A search strategy that identified postoperative occurrences of lung consolidations recorded within progress notes or discharge summaries had a lower sensitivity of 0.64 (95% CI, 0.58-0.70), and a specificity of 0.94 (95% CI, 0.94-0.96).

Occurrences of postoperative sepsis were identified using the patient safety indicator method with a sensitivity of 0.34 (95% CI, 0.24-0.47) and a specificity of 0.99 (95% CI, 0.98-0.99; Figure 2 and eTable 4). An identification strategy combining query searches for multiorgan failure, septic shock, systemic infection, or bacterial or fungal organisms on blood culture reports resulted in a sensitivity of 0.89 (95% CI, 0.78-0.94) with a specificity of 0.95 (95% CI, 0.93-0.96).

The patient safety indicator algorithms identified postoperative myocardial infarctions with a sensitivity of 0.89 (95% CI, 0.74-0.96) and a specificity of 0.99 (95% CI, 0.98-0.99; Figure 2 and eTable 5) Combining cardiac biomarker results obtained from structured data with text searches of postoperative progress notes for SNOMED terms related to “electrocardiographic ST segment changes” resulted in a sensitivity of 0.74 (95% CI, 0.58-0.86) and a specificity of 0.98 (95% CI, 0.98-0.99).

**COMMENT**

We found that automated searches of an EMR using a natural language processing–based approach was able to identify occurrences of acute renal failure requiring dialysis, deep vein thrombosis, pulmonary embolism, pneumonia, sepsis, and acute myocardial infarctions in patients following surgery. Varying the search strategies and source documents resulted in differing levels of case finding and false positive alerts; however, for many outcomes, rules could be developed with both high sensitivities and high positive predictive values. For some outcomes, the choice of search strategy required substantial tradeoffs between case finding and false-positive alerts.

Although the patient safety indicator algorithms offered consistently high specificities, the natural language processing approach in general had significantly greater sensitivities with only a small reduction in specificities. In addition, depending on one’s chosen search strategy, positive predictive values could be moderate to high. In contrast to the patient safety indicator approach, for which test characteristics are fixed, the natural language processing offered a wide array of search strategies with varying test characteristics. Nevertheless in some cases, specifically postoperative myocardial infarction, the patient safety indicator algorithm had excellent test characteristics that were not improved through the natural language processing approach.

A natural language processing–based approach offers several potential advantages over administrative-code based strategies to identify health care quality concerns. First is the flexibility of the approach to meet the individual institutional needs. Once documents have been processed, different approaches and query strategies to identify a specific outcome can be implemented at a relatively low programming effort using standard database query applications. Second, as opposed to administrative codes, search strategies using daily progress notes, microbiology reports, or imaging reports could be monitored on a prospective basis. Thus, this approach could potentially identify complications while a patient is still in the hospital, which could greatly facilitate real-time quality assurance processes. A natural language processing–based search strategy is far more scalable than manual abstraction, potentially allowing surveillance on an entire health care system population rather than a subsample. Finally, in systems with highly integrated EMRs, prospective surveillance could be extended to the outpatient setting for individuals remaining with the health care system.

Only a few studies have used text-based approaches to identify medical complications. In a study by Melton and Hripcsak,4 the natural language processing system MedLEE was used to identify 45 adverse events tracked as part of the New York Patient Occurrence Reporting and Tracking System. The overall sensitivity of the system was 0.28 (95% CI, 0.17-0.42) with a specificity of 0.99 (95% CI, 0.98-0.99). This system was limited in that the only electronically available text source was discharge summaries. Penz et al15 compared 2 automated techniques to identify adverse events related to the use of central venous catheters. An approach using a phrase-matching algorithm had a higher sensitivity but lower specificity than an approach using natural language processing. Improvements in using automated approaches to extract information from the clinical narrative are still ongoing, but this approach is well-regarded as a current strategy for the detection of adverse events associated with medical care.15,23

Our patient safety indicator results are similar to previously published studies. Romano et al8 and Rivard et al18 found the sensitivity of the patient safety indicators in a VA population was 44% (95% CI, 32%-56%) to detect acute renal failure, 56% (95% CI, 50-63) to detect pulmonary embolism or deep vein thrombosis; and 32% (95% CI, 23%-43%) to detect sepsis. Differences in patient safety indicator event rates between VA and non-VA populations have been generally small and inconsistent.24 Rosen et al25,26 have suggested that these differences are likely a result of inadequate case-mix adjustment.

©2011 American Medical Association. All rights reserved.
A strength of our study is its large sample size. In addition, we only applied our natural language processing queries to cases that would have been included within the patient safety indicator denominator. Although this reduced the number of total events that we would have detected, this approach helped ensure the best compatibility between the patient safety indicator and the natural language processing approaches. Another strength of our study was the use of VASQIP nurse-reviewed events as our referent standard. The VASQIP program has been in operation for more than 15 years, and nurse reviewers undergo a rigorous training protocol that has been determined to be reliable.²⁷ In addition, this study applied natural language processing methods for extraction of clinical information across multiple types of medical documentation occurring over a longitudinal period of hospitalizations.

Our study has several limitations. One is that the patient safety indicators were not originally designed for VA data and some of the methodological issues in modifying the patient safety indicators for VA data have been previously described.⁹,²¹ Nevertheless, patient safety indicator rates appear similar between the VA and non-VA populations.²⁴ Perhaps the greatest limitation is that although the adoption of EMRs by health systems is improving, only a small minority of institutions currently use them, so some of the query strategies would not be feasible at all institutions.²⁷ Nevertheless, our results should contribute to the growing literature supporting the utility of EMRs and help to encourage future adoption of such technology and the integration of such systems across health care systems.

In conclusion, using natural language processing with an electronic medical record greatly improves postoperative complication identification compared with the patient safety indicators, an administrative-code based algorithm. Different query strategies pro-

---

**Figure 2. Receiver Operating Characteristic Plots for Different Search Queries of a Natural Language Processing–Approach or Patient Safety Indicator Algorithm**

1. **Postoperative acute renal failure requiring dialysis**
   - Sensitivity
   - Specificity
   - Search queries:
     - Single rule
     - Combined
     - Patient safety indicator algorithm

2. **Postoperative deep vein thrombosis or pulmonary embolism**
   - Sensitivity
   - Specificity
   - Search queries:
     - Deep vein thrombosis
     - Pulmonary embolism
     - Single rule
     - Combined
     - Both (combined)
     - Patient safety indicator algorithm

3. **Postoperative pneumonia**
   - Sensitivity
   - Specificity
   - Search queries:
     - Single rule
     - Combined
     - Patient safety indicator algorithm

4. **Postoperative myocardial infarction**
   - Sensitivity
   - Specificity
   - Search queries:
     - Single rule
     - Combined
     - Patient safety indicator algorithm
duced varying sensitivity and specificity, which in many cases could be improved through combining individual queries to optimize test characteristics. A natural language processing–based approach designed to detect postoperative complications within an EMR identifies several surgical complications with moderate to good sensitivities and specificities. Developing natural language processing–based algorithms was an iterative process and in many cases query combinations resulted in improvement of poorly functioning rules. As additional institutions develop fully integrated EMR, electronic chart reviews for quality purposes should be further developed and evaluated.

**REFERENCES**


**Funding/Support:** This study was supported by grant SA-F:03-223 from the Department of Veterans Affairs.

**Role of the Sponsor:** The sponsors were not involved in the design and conduct of the study; collection, management, analysis, or interpretation of the data; and preparation, review, or approval of the manuscript.

**Disclaimer:** The view expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

**Online-Only Material:** eTables 1 through 5 are available at http://www.jama.com.

**Additional Contributions:** We thank the VA Surgical Quality Data Use Group (SQDUC) for its role as scientific advisors and for the critical review of data use and analysis presented in the manuscript. We also thank the collaborative investigators for their work in this study: Debra Jo Barrett, MSN, Lexington VA Medical Center, Lexington, Kentucky; William G. Cheadle, MD, Louisville VA Medical Center, Louisville, Kentucky; Brad L. Roper, PhD, Memphis VA Medical Center, Memphis, Tennessee; Teresa England, RN, James H. Quil- len VA Medical Center, Mountain Home, Tennessee; and Sandra E. Shaw, RN, Huntington VA Medical Center, Huntington, West Virginia, none of whom received compensation.

©2011 American Medical Association. All rights reserved.