Patient Safety

Validity of Selected AHRQ Patient Safety Indicators Based on VA National Surgical Quality Improvement Program Data

Patrick S. Romano, Hillary J. Mull, Peter E. Rivard, Shibe Zhao, William G. Henderson, Susan Loveland, Dennis Tsilimingras, Cindy L. Christiansen, and Amy K. Rosen

Objectives. To examine the criterion validity of the Agency for Health Care Research and Quality (AHRQ) Patient Safety Indicators (PSIs) using clinical data from the Veterans Health Administration (VA) National Surgical Quality Improvement Program (NSQIP).

Data Sources. Fifty five thousand seven hundred and fifty two matched hospitalizations from 2001 VA inpatient surgical discharge data and NSQIP chart-abstracted data.

Study Design. We examined the sensitivities, specificities, positive predictive values (PPVs), and positive likelihood ratios of five surgical PSIs that corresponded to NSQIP adverse events. We created and tested alternative definitions of each PSI.

Data Collection. FY01 inpatient discharge data were merged with 2001 NSQIP data abstracted from medical records for major noncardiac surgeries.

Principal Findings. Sensitivities were 19–56 percent for original PSI definitions; and 37–63 percent using alternative PSI definitions. PPVs were 22–74 percent and did not improve with modifications. Positive likelihood ratios were 65–524 using original definitions, and 64–744 using alternative definitions. “Postoperative respiratory failure” and “postoperative wound dehiscence” exhibited significant increases in sensitivity after modifications.

Conclusions. PSI sensitivities and PPVs were moderate. For three of the five PSIs, AHRQ has incorporated our alternative, higher sensitivity definitions into current PSI algorithms. Further validation should be considered before most of the PSIs evaluated herein are used to publicly compare or reward hospital performance.

Key Words. Patient safety indicators, criterion validity, administrative data, medical errors

Patient safety persists as a national concern since the Institute of Medicine's landmark report on medical errors (Kohn, Corrigan, and Donaldson 2000). The Agency for Health Care Research and Quality (AHRQ) recently released a methodology, the Patient Safety Indicators (PSIs), to screen for potential patient safety events using administrative data from acute care hospitals. The PSIs are an attractive tool because they use readily available data and standardized algorithms; they are risk adjusted and therefore potentially useful for benchmarking; and they are easy to implement using free, downloadable software (AHRQ 2007a, 2008).

The evidence published to date suggests that the PSIs generally have high specificity (i.e., low false-positive rates) and modest sensitivity (i.e., moderate false-negative rates) (Gallagher, Cen, and Hannan 2005b; Zhan et al. 2007; Houchens, Elixhauser, and Romano 2008). Although several recent studies have used the PSIs to identify significant gaps and variations in safety (Romano et al. 2003; Rosen et al. 2005, 2006), the PSIs are still regarded by both AHRQ and the user community principally as screening tools to flag potential safety-related events rather than as definitive measures (AHRQ 2007a).

Increasing use of the PSIs for public reporting and pay-for-performance (HealthGrades 2008; Premier Inc. 2008) makes it imperative that the PSIs undergo more rigorous evaluation. Although previous studies have demonstrated the face, content, and predictive validity of the PSIs, there is insufficient evidence of their criterion validity to support some of these new applications. The few published studies examining the criterion validity of the PSIs are limited by small sample sizes or lack of a true gold standard (Weller et al. 2004; Gallagher, Cen, and Hannan 2005a, b; Shufelt, Hannan, and Gallagher 2005; Polancich, Restrepo, and Prosser 2006; Zhan et al. 2007).

As a national leader in patient safety (Leape 2005), the Veterans Health Administration (VA) is well positioned to evaluate the criterion validity of the PSIs. The VA has several data sources that can serve as valuable resources for this endeavor. VA administrative data, necessary for estimating risk-adjusted

Address correspondence to Patrick S. Romano, M.D., M.P.H. – UC Davis Division of General Medicine and Center for Healthcare Policy and Research, 4150 V Street, PSSB Suite 2400, Sacramento, CA 95817; e-mail address: psromano@ucdavis.edu. Amy K. Rosen, Ph.D., Hillary J. Mull, M.P.P., Shibe Zhao, M.P.H., Susan Loveland, M.A.T. are with the Center for Health Quality, Outcomes and Economic Research, Bedford VAMC (152), Bedford, MA. Peter E. Rivard, Ph.D., is with the Center for Organization, Leadership, and Management Research, Boston VA Medical Center, Boston, MA., William G. Henderson, Ph.D., M.P.H., is with the Colorado Health Outcomes Program, University of Colorado Health Sciences Center Aurora, CO., Dennis Tsilimingras, M.D., M.P.H., is with the 1108 Blairmoor CT, Grosse Pointe Woods, MI. Cindy L. Christiansen, Ph.D., is with the Health Policy and Management Department, Boston University School of Public Health, Boston, MA.

PSI rates, contain detailed diagnostic and utilization information on inpatient episodes of care. The VA also collects rich chart-abstracted data on major noncardiac surgeries through the National Surgical Quality Improvement Program (NSQIP) (Khuri et al. 1998). NSQIP was designed to promote continuous quality monitoring and improvement by providing reliable, valid, comparative information regarding surgical outcomes to all facilities performing major noncardiac surgery (Daley et al. 1997; Khuri et al. 1998). NSQIP data were used as a “gold standard” for identifying postoperative complications in one previous study (Best et al. 2002), although the mapping of clinically defined events to ICD-9-CM complication codes was somewhat inexact (Romano 2003).

The purpose of this paper is to evaluate the criterion validity of surgical PSIs that match NSQIP adverse events. Our specific objectives were to (1) estimate the sensitivity, specificity, positive predictive value (PPV), and likelihood ratio of the PSIs using NSQIP data as the gold standard; and (2) improve the sensitivity and PPV of the PSIs, if possible, through revisions to PSI algorithms. If the PSIs demonstrate high criterion validity, then public reporting and pay-for-performance activities using these indicators will likely multiply.

METHODS

Data Sources

Our primary data source was the VA Patient Treatment File (PTF), an administrative database that contains records on all patients discharged from or residing in VA acute and nonacute inpatient care facilities at the end of each fiscal year (Rosen et al. 2005). The PTF is comprised of four subfiles. The main file contains demographic, diagnostic (one principal and up to nine secondary ICD-9-CM diagnosis codes, plus the diagnosis accounting for the greatest portion of the patient’s stay, which we did not use in this study), and summary information on each episode of care (e.g., dates of admission/discharge and discharge status.) The Bedsection file contains one primary and up to four secondary diagnoses, and length of stay information, for each stay under a particular service. The procedure file includes ICD-9-CM procedure codes (procedures not performed in an operating room or under anesthesia) and their respective dates and times; the surgery file contains similar data on all surgeries (procedures performed in a surgical suite or operating room).

We used NSQIP’s clinical database for validation purposes. To ensure the reliability, validity, and comparability of information across hospitals, trained nurse reviewers collect detailed clinical information prospectively

from all VA facilities performing major surgery. The first eligible operation (excluding cardiac surgeries) that requires general, spinal, or epidural anesthesia is entered into a standard database available at each facility (Best et al. 2002). Abstracted data include preoperative patient characteristics, intraoperative process information, mortality within 30 days of surgery, and 21 postoperative adverse events that can occur within 30 days of surgery (Khuri et al. 1995). For an event to count as a complication, the nurse reviewer must establish a causal link with the prior operation. Substantial to excellent interrater reliability ($\kappa = 0.40\text{--}0.89$) has been reported for postoperative outcomes (Davis et al. 2007).

In addition to inclusion criteria, NSQIP employs certain exclusion criteria so that not all surgical cases are reviewed. Surgical procedures with very low observed mortality are excluded, while those at high-volume hospitals (>36 cases per 8-day cycle) are randomly sampled to reduce abstraction burden (see supporting Appendix S1).

Sample

We selected all discharges from the PTF during Fiscal Year 2001 (FY01) (October 1, 2000 to September 30, 2001). We excluded 4,822 hospitalizations involving nonveterans (of which over 90 percent were nonsurgical), yielding a sample of 354,470 veterans and 561,436 hospitalizations, representing 130 VA hospitals. We retained the *nonacute* portion of care because NSQIP includes all patients regardless of care setting. We linked hospitalizations by patient identifiers across all four subfiles.

Merging PTF and NSQIP Data

Because of differences between NSQIP and PTF data, several steps were necessary to match cases. NSQIP data include only surgical cases, whereas PTF data include both medical and surgical. Therefore, we selected surgical hospitalizations from the PTF (i.e., those assigned surgical DRGs using the *PSI* software, version 2.1, revision 2, applied to the principal diagnosis and all reported procedure codes), which substantially reduced the sample of hospitalizations eligible for matching from 561,436 to 101,548. We then sent NSQIP a data file containing patient identifiers, admission, and discharge dates, and facility numbers of all surgical hospitalizations from the PTF. NSQIP returned a file containing all surgical patients who matched PTF data as well as information on unmatched patients, so that we could explore reasons for mismatches.

We could not perform a simple data merge because PTF data were organized at the hospitalization level, while NSQIP data were at the surgical

procedure level. Consequently, we developed algorithms to merge only those records in which NSQIP surgery dates fell between PTF admission and discharge dates. In 2 percent of cases, multiple NSQIP surgeries occurred during a single hospitalization; these were retained to maximize power and generalizability, and each surgery was considered independently for risk of PSI events.

The matched PTF/NSQIP file contained 56,419 hospitalizations (Figure 1). Forty-four percent of the PTF hospitalizations ($n = 45,129$) could not be matched with NSQIP surgery records. Of these, 47.1 percent ($n = 21,256$) did not match because: (1) some hospitalizations with surgical DRGs did not have a “valid operating room surgery requiring anesthesia,” as defined in NSQIP; (2) VA facilities without “major surgery” capabilities do not participate in NSQIP; and (3) NSQIP groups cases by year of surgery, while the PTF groups hospitalizations by year of discharge. The remaining 53 percent of PTF hospitalizations ($n = 23,873$) were not in NSQIP due to

Figure 1: Matching 2001 VA National Surgical Quality Improvement Program (NSQIP) Records to FY01 Veteran’s Inpatient Data

NSQIP exclusion criteria (supporting Appendix S1). In addition, there were 40,476 surgery records from NSQIP that did not match PTF data; these were primarily outpatient surgeries that are not collected in the PTF. Finally, there were additional mismatches because some NSQIP cases were discharged in FY02, whereas the PTF was limited to FY01 discharges.

As a final step, we deleted 588 hospitalizations from Puerto Rico from the merged file to conform to *PSI* software requirements, as well as hospitalizations without a valid operating room procedure in the PTF (because such hospitalizations were not at risk for the PSIs that we evaluated). Our final data file consisted of 55,752 hospitalizations, representing 59,838 surgeries and 51,832 patients in 110 hospitals.

Overview of the PSIs

The AHRQ PSIs, as described in previous studies (Miller et al. 2001; Romano et al. 2003), were an outgrowth of the Complications Screening Program (CSP), which was a pioneering effort to use computerized algorithms to screen hospital discharge abstracts for adverse events suggesting lapses in quality (Iezzoni et al. 1994a, b). CSP indicators with PPVs > 75 percent according to any of three validation studies involving coders, nurse abstractors, and physician reviewers (Lawthers et al. 2000; McCarthy et al. 2000; Weingart et al. 2000) were selected as potential PSIs, along with other indicators identified from the literature and ICD-9-CM. The PSIs were designed to capture potentially preventable events related to inpatient safety; hence, patients for whom a complication seemed less likely to be preventable were excluded.

Each PSI is defined as a proportion or rate, with both a numerator (hospitalizations with the complication of interest) and a denominator (hospitalizations at risk). The final set of 20 hospital-level PSIs resulted from a four-step process that included literature review, evaluation of candidate PSIs by multidisciplinary clinical panels using a modified Delphi technique based on the RAND/UCLA Appropriateness Method (Fitch et al. 2001), consultation with coding experts, and empirical analyses of reliability, confounding bias, and construct validity (McDonald et al. 2002; Zhan and Miller 2003). Sixteen additional indicators were placed on a separate “experimental” list because panelists scored them as less useful or disagreed about their usefulness.

Comparing Adverse Events Between the Two Sources of Data

From the eight surgical PSIs, we selected five (Table 1) whose definitions, based on ICD-9-CM codes, corresponded to the clinical definitions of NSQIP

Table 1: Patient Safety Indicator (PSI) Definitions (AHRQ version 2.1, revision 2) and NSQIP Adverse Event Definitions

<i>PSI</i>	<i>PSI Definition</i>	<i>NSQIP Adverse Event</i>	<i>NSQIP Definition</i>
Postoperative physiologic/metabolic derangement	Cases of specified physiological or metabolic derangement per 1,000 elective surgical discharges with OR procedure	Acute renal failure (postop)	In a patient who did not require dialysis preoperatively, worsening of renal dysfunction postoperatively requiring hemodialysis, ultrafiltration, or peritoneal dialysis
Postoperative respiratory failure	Cases of acute respiratory failure per 1,000 elective surgical discharges with OR procedure	Failure to wean > 48 hours	On ventilator > 48 hours postoperative
		Reintubation for respiratory/cardiac failure	Patient required placement of an endo-tracheal tube and mechanical or assisted ventilation because of the onset of respiratory or cardiac failure manifested by severe respiratory distress, hypoxia, hypercarbia, or respiratory acidosis
Postoperative PE/DVT	Cases of deep vein thrombosis (DVT) or pulmonary embolism (PE) per 1,000 surgical discharges with OR procedure	Pulmonary embolism	Lodging of a blood clot in a pulmonary artery with subsequent obstruction of blood supply to the lung parenchyma. The blood clots usually originate from the deep leg veins or the pelvic venous system
		Deep vein thrombosis	The formation, development, or existence of a blood clot or thrombus within the vascular system, which may be coupled with inflammation . . . The patient must be treated with heparin and/or coumadin or warfarin, and/or placement of a vena cava filter or clipping of the vena cava

Postoperative sepsis	Cases of sepsis per 1,000 elective surgery patients with OR procedure and a length of stay of 4 days or more	Systemic sepsis	The primary physician or the chart states that the patient had systemic sepsis within the 30 days postoperatively: definitive evidence of infection, plus evidence of a systemic response . . . manifested by TWO or more of the following conditions Temp >38°C or <36°C Septic shock . . . with hypotension . . . HR >90bpm RR >20 breaths/min or PaCO ₂ <32 mmHg WBC >12,000 cells/mm ³ , <4,000 cells/mm ³ , or >10% immature forms Separation of the layers of a surgical wound, which may be partial or complete, with disruption of the fascia
Postoperative wound dehiscence	Cases of reclosure of postoperative disruption of abdominal wall per 1,000 cases of abdominopelvic surgery	Dehiscence	
Postoperative acute myocardial infarction	Cases of acute myocardial infarction per 1,000 noncardiac surgical discharges	Myocardial infarction	A new transmural acute myocardial infarction occurring during surgery or within 30 days following surgery, as manifested by new Q waves on ECG
Postoperative iatrogenic complications: cardiac	Cases of postoperative cardiac complications per 1,000 surgical discharges	Cardiac arrest requiring CPR	The absence of cardiac rhythm or presence of chaotic rhythm that results in loss of consciousness requiring the initiation of any component of BLS or ACLS

NSQIP, National Surgical Quality Improvement Program; AHRQ, Agency for Health Care Research and Quality.

events: “postoperative physiologic/metabolic derangements,” “postoperative respiratory failure,” “postoperative pulmonary embolism/deep vein thrombosis” (PE/DVT), “postoperative sepsis,” and “postoperative wound dehiscence.” We also identified two “experimental PSIs” that matched adverse events in NSQIP: “postoperative acute myocardial infarction” and “postoperative iatrogenic complications—cardiac” (McDonald et al. 2002). Despite our ability to create crosswalks between these seven indicators and NSQIP adverse events, definitions did not always correspond exactly. PSIs are defined using ICD-9-CM codes applied by professional coders who review physician documentation, whereas NSQIP complications are defined using clinical definitions applied by nurse abstractors who review laboratory and radiologic data as well as physician documentation.

To ensure fair comparisons between PSI and NSQIP events, we limited our analyses to hospitalizations that met the denominator definition of each PSI. For instance, only patients who underwent major abdominopelvic surgery were included in the denominator of “postoperative wound dehiscence,” because other types of surgery are not in the risk pool for that PSI. PSIs capture only in-hospital events while NSQIP captures adverse events within 30-days postsurgery; therefore, we deleted NSQIP events that occurred after the matched PTF hospitalization’s discharge date. Finally, to improve the match between PSI-identified and NSQIP-identified adverse events (i.e., to improve sensitivity and PPV), we explored several alternative definitions of each PSI using different combinations of ICD-9-CM diagnosis and procedure codes. Clinical and coding input was used to modify AHRQ’s PSI definitions. Our “original” (AHRQ PSI software, version 2.1, revision 2) and the best of these “alternative” PSI definitions (based on the balance between sensitivity and PPV) are shown in Table 2.

Analyses

Analyses were performed using SAS (version 8.0). We determined occurrence rates of PSI events by applying the PSI software (version 2.1, revision 2) to our VA hospital discharge summary file. Minor modifications to the PTF structure and to several PTF data elements were necessary, as described previously (Rivard et al. 2005). The occurrence of PSI events and NSQIP-defined adverse events were designated by separate dichotomous variables.

We estimated the sensitivity, specificity, PPV, and positive likelihood ratios of the five original PSIs using NSQIP as the gold standard. These parameters were reestimated using alternative definitions of the AHRQ PSIs.

Table 2: “Original” AHRQ Patient Safety Indicator (PSI) Definitions (version 2.1, revision 2) and Current/Alternative Definitions

<i>PSI</i>	<i>Original Definitions</i>	<i>Current/Alternative Definitions (Changes in Italics)</i>
Postoperative physiologic/metabolic derangement	Numerator: Discharges with acute renal failure (subgroup of physiologic and metabolic derangements, 584.x) must be accompanied by a procedure code for dialysis (39.95, 54.98)	Alternative numerator: Discharges with acute renal failure (subgroup of physiologic and metabolic derangements, including codes 584.x or 586 or 997.5 or 788.5) must be accompanied by a procedure code for dialysis (39.95, 54.98) <i>after the date of the index surgical procedure</i>
Postoperative respiratory failure	Numerator: Discharges with ICD-9-CM codes for acute respiratory failure (518.81) in any secondary diagnosis field. (After 1999, include 518.84)	Alternative numerator: Discharges with ICD-9-CM codes for acute respiratory failure (518.81, 518.84) in any secondary diagnosis field, <i>OR</i> <i>ICD-9-CM codes for reintubation/prolonged ventilation procedure as follows:</i> <ul style="list-style-type: none"> • (96.04) 1 or more days after the major operating room procedure code • (96.70 or 97.71) 2 or more days after the major operating room procedure code • (96.72) zero or more days after the major operating room procedure code
Postoperative PE/DVT	Numerator: Discharges with ICD-9-CM codes for deep vein thrombosis (45x) and/or pulmonary embolism in (415.1x) any secondary diagnostic field	Alternative numerator: Discharges with ICD-9-CM codes for deep vein thrombosis (45x) and/or pulmonary embolism in (415.1x) in any secondary diagnostic field, <i>including any hospitalization with a secondary procedure code for interruption of vena cava (38.7) on any day AFTER the day of the principal procedure</i>
Postoperative sepsis	Numerator: Discharges with ICD-9-CM code for sepsis (038.xx) in any secondary diagnostic field	Alternative numerator: Discharges with ICD-9-CM codes for sepsis (038.xx, 998.0, 998.1, 785.59, 785.50, 785.5, 785.52) in any secondary diagnostic field
Postoperative wound dehiscence	Numerator: Discharges with ICD-9-CM code for reclosure of postoperative disruption of abdominal wall (54.61) in any procedure field	Alternative numerator: Discharges with ICD-9-CM code for reclosure of postoperative disruption of abdominal wall (54.61, 998.3x) in any procedure field

continued

Table 2. *Continued*

<i>PSI</i>	<i>Original Definitions</i>	<i>Current/Alternative Definitions (Changes in Italics)</i>
Postoperative acute myocardial infarction	Numerator: Discharges with ICD-9-CM code for acute myocardial infarction (410.x0, 410.x1) in any secondary diagnosis field	Alternative numerator: Discharges with ICD-9-CM code for acute myocardial infarction, <i>initial episode of care</i> (410.x1), in any secondary diagnosis field
Postoperative iatrogenic complications cardiac	Numerator: Discharges with ICD-9-CM code for cardiac complications (997.1) in any secondary diagnosis field	Alternative numerator: Discharges with ICD-9-CM code for <i>ventricular fibrillation and flutter (472.4x)</i> or <i>cardiac arrest (427.5)</i> in any secondary diagnosis field

The alternative definitions shown for “initial episode of care,” “ventricular fibrillation and flutter,” and “cardiac arrest” were adopted in AHRQ *PSI* version 3.0 (and subsequent versions).

AHRQ, Agency for Health Care Research and Quality.

Sensitivity represents the proportion of cases with an NSQIP adverse event that were correctly flagged for the corresponding PSI. Specificity represents the proportion of cases without a NSQIP adverse event that were correctly not flagged for the corresponding PSI. PPV represents the proportion of cases flagged for a PSI that were also identified in NSQIP (confirmed) as having an adverse event. The positive likelihood ratio (sensitivity/[1-specificity]) measures how many times more likely a flagged PSI was to occur in a hospitalization that had a “true” event (based on NSQIP) than in a hospitalization that did not have the true event. This ratio can be multiplied by the prior odds of an event (which approximates prevalence for rare events) to yield the posterior odds given a flagged PSI. We calculated 95 percent confidence intervals for sensitivity, specificity, and PPV using the Wilson score method (Newcombe 1998); intervals for the likelihood ratio used the method developed by Simel, Samsa, and Matchar (1991).

RESULTS

Our sample was 95.4 percent male, with an average age of 63 years; 47 percent of the persons in our sample were over 65 years of age. Our sample was similar to the overall VA surgical population based on the entire PTF, although mean length of stay was shorter ($p < .05$) (12.6 versus 14.6 days, respectively) and cardiac, ophthalmologic, oral, plastic, and miscellaneous surgery were underrepresented in our sample, as expected (see supporting Appendix S2).

Overall Validation Results

In general, we found moderate sensitivities (29–56 percent, except for “postoperative respiratory failure” at 19 percent) and PPVs (44–74 percent, except for “postoperative PE/DVT” at 22 percent) for the original PSIs (Table 3). All PSIs had high specificities, from 99.1 percent (“postoperative PE/DVT”) to 99.9 percent (“postoperative derangements”). Positive likelihood ratios for the original PSIs ranged from a low of 65 (“postoperative PE/DVT”) to a high of 524 (“postoperative derangements”).

All of the alternative PSI definitions had higher estimates of sensitivity than the original indicators, although the only statistically significant increases were for “postoperative respiratory failure” (from 19 to 67 percent) and “postoperative wound dehiscence” (from 29 to 61 percent) ($p < 0.05$). With respect to PPVs and positive likelihood ratios, the only statistically significant changes were for “postoperative wound dehiscence”: PPV decreased from 72 to 57 percent ($p < 0.05$) and positive likelihood ratio decreased from 160 to 79 ($p < 0.05$).

Individual PSI Validation Results

“Postoperative physiologic and metabolic derangements” (PSI) versus “acute renal failure” (NSQIP). The original PSI definition was broader than NSQIP’s definition because the NSQIP definition was limited to acute renal failure requiring postoperative dialysis whereas the AHRQ definition also included diabetic complications. To facilitate comparison, we focused on the PSI-flagged renal failure cases. The original PSI definition omitted two relevant but vague diagnosis codes (997.5, “urinary complications”; 586, “renal failure, unspecified”), even though the former code includes “renal failure (acute), specified as due to procedure.” To improve the match between the PSI and NSQIP, we added 586 and 997.5 to the original PSI definition (if accompanied by a dialysis procedure code dated after the first operating room procedure). The sensitivity, PPV, and likelihood ratio of this indicator all increased slightly but not significantly. More substantial improvement in sensitivity (to over 74 percent) was achieved by dropping the dialysis requirement from the PSI definition if the patient had acute renal failure (584), but at the price of much worse PPV (23 percent).

“Postoperative respiratory failure” (PSI) versus “unplanned intubation for respiratory failure” and/or “failure to wean from ventilator > 48 hours” (NSQIP). The original PSI definition was broader than NSQIP’s definition because the AHRQ definition included all patients with acute respiratory failure after

Table 3: Criterion Validity of the Original and Alternative Patient Safety Indicators (PSIs) versus NSQIP Adverse Events

PSI	Hospital- izations at-Risk for PSI	NSQIP Adverse Events:Gold Standard	Sensitivity (%) [†]		PPV (%) [‡]		Likelihood Ratio [§]	
			Original PSI Def*	Current/ Alternative Def	Original PSI Def*	Current/ Alternative Def	Original PSI Def*	Current/ Alternative Def
Postoperative physiologic/ metabolic derangement	27,722	62	44 (32–56)	48 (36–61)	54 (40–67)	63 (48–75)	524 (319– 861)	744 (438–1261)
Postoperative respiratory failure	24,273	344	19 (15–23)	63 (57–67)	74 (63–82)	68 (62–73)	194 (122– 308)	147 (119–181)
Postoperative PE/DVT	55,682	241	56 (50–63)	58 (51–64)	22 (19–25)	22 (19–25)	65 (56–75)	64 (56–73)
Postoperative sepsis	12,011	75	32 (23–43)	37 (27–49)	44 (31–57)	45 (33–57)	123 (76– 200)	131 (84–205)
Postoperative wound dehiscence	16,904	274	29 (24–34)	61 (55–67)	72 (63–80)	57 (51–62)	160 (107– 239)	79 (65–96)
Postoperative acute myocardial infarction	26,925	111	81 (73–87)	44 (35–53)	49 (42–56)	56 (46–68)	231 (185– 288)	311 (213–456)
Postoperative iatrogenic complications: cardiac	56,305	609	17 (14–20)	27 (23–30)	8 (6–9)	49 (43–54)	8 (6–9)	86 (71–105)

Parentheses contain 95% confidence intervals. All specificity values are > 99.1%. Paired sensitivity estimates that were statistically significantly different, based on the continuity-corrected McNemar statistic ($p < .05$) for matched pairs, are shown in boldface. The “alternative definitions” shown for “postoperative physiologic/metabolic derangement,” “postoperative respiratory failure,” and “postoperative sepsis” were adopted in AHRQ PSI version 3.0 (and subsequent versions).

*Original definition from AHRQ PSI version 2.1, revision 2.

[†]Sensitivity represents the proportion of the NSQIP postoperative adverse events that were found using the AHRQ PSI algorithms.

[‡]Positive predictive value (PPV) represents the proportion of the AHRQ-defined PSIs that were confirmed as true events using NSQIP.

[§]Positive likelihood ratio measures how many times more likely a flagged PSI was to occur in a hospitalization that had a “true” event (based on NSQIP) versus a hospitalization that did not have the true event.

AHRQ, Agency for Health Care Research and Quality; PSI, Patient Safety Indicator; NSQIP, National Surgical Quality Improvement Program; PE/DVT, pulmonary embolism/deep vein thrombosis.

surgery whereas the NSQIP definition was limited to patients who were “on ventilator >48 hours postoperative” or required reintubation because of respiratory or cardiac failure. To improve the match between the PTF and NSQIP, we added postoperative reintubation/prolonged ventilation procedure codes (96.04, 96.70 or 97.71, 96.72) to the PSI numerator, with date restrictions. These changes led to a substantial, statistically significant improvement in the sensitivity of the indicator, at the cost of slight decreases in the PPV and the likelihood ratio. Adding 518.5 (“pulmonary insufficiency following trauma and surgery”) to the AHRQ definition improved sensitivity further (67 percent) but also worsened PPV (66 percent). An alternative definition relying only on procedure codes was less sensitive than the preferred definition.

“Postoperative PE/DVT” (PSI) versus “PE” and/or “DVT” (NSQIP). The NSQIP definition was more restrictive than the original PSI definition. To establish a diagnosis of PE, NSQIP required either a high probability V-Q scan or a positive angiogram or CT scan, whereas the PSI only required a physician diagnosis. For DVT, NSQIP required either anticoagulation or vena caval interruption, whereas the PSI was triggered by secondary diagnoses alone. To improve the match between the PTF and NSQIP, we added a secondary procedure code for placement of an inferior vena cava filter (38.7) occurring any day after the principal procedure. This alternative definition had minimally higher sensitivity (from 56 to 58 percent), but the PPV and positive likelihood ratio remained essentially constant. Restricting the PSI denominator to elective surgery modestly improved both sensitivity (from 56 to 67 percent) and PPV (from 22 to 30 percent).

“Postoperative sepsis” (PSI) versus “systemic sepsis” (NSQIP). The NSQIP definition was slightly narrower than the original PSI definition. The AHRQ definition included all types of “septicemia” (038.xx), plus “systemic inflammatory response syndrome due to infectious process without/with organ dysfunction” (995.91 and 995.92), whereas the NSQIP definition required “definitive evidence of infection” plus two or more findings listed in Table 1. To improve the match, we added diagnosis codes for postoperative or septic shock (998.0, 785.59, 785.52) to the PSI numerator. However, this change had a modest effect, increasing both sensitivity and positive likelihood ratio slightly (from 32 to 37 percent and 123 to 131, respectively) but not significantly.

“Postoperative wound dehiscence” (PSI) versus “dehiscence” (NSQIP). The NSQIP definition was far broader than the PSI definition, in that AHRQ required a surgical procedure to close the wound (code 54.61), whereas NSQIP relied on the wound’s appearance. To improve the match between the PTF and NSQIP, we added a diagnosis code (998.3x, “disruption of operation wound”) to the PSI numerator. This change resulted in a statistically significant increase in sensitivity (from 21 to 69 percent), but also decreases in both the PPV and the positive likelihood ratio (from 72 to 57 percent and 160 to 79, respectively). An alternative definition using this diagnosis code alone also had poor PPV. Restricting the PSI denominator to elective surgery improved both sensitivity (from 29 to 39 percent) and PPV (from 72 to 90 percent).

Experimental PSIs

The NSQIP definition of “postoperative myocardial infarction” was narrower than the experimental PSI definition, in that NSQIP only captured *Q*-wave infarcts. As a result, the PSI appeared to have high sensitivity (81 percent) but moderate PPV (49 percent). The NSQIP definition of “postoperative cardiac arrest” was also narrower than the experimental PSI definition, in that NSQIP only captured events requiring cardiopulmonary resuscitation. An alternative definition based on diagnosis codes for ventricular fibrillation/flutter and cardiac arrest had better PPV (49 versus 8 percent) and positive likelihood ratio (86 versus 8), but still poor sensitivity (27 versus 17 percent).

DISCUSSION

The purpose of this study was to evaluate the criterion validity of selected surgical PSIs in the VA using chart-abstracted data collected on surgical adverse events by NSQIP. Despite differences between the PTF and NSQIP, we were able to create a matched PTF/NSQIP file to validate five of the surgical PSIs (and two experimental PSIs) using “gold standard” clinical data. In general, we found moderate sensitivities and PPVs for the original PSIs. The proportion of adverse events identified by NSQIP that were also flagged by ICD-9-CM codes varied across the PSIs, from 19 percent for “postoperative respiratory failure” to 56 percent for “postoperative PE/DVT.” The proportion of events identified by ICD-9-CM codes that were confirmed by NSQIP had a similar range, from 22 percent for “postoperative PE/DVT” to 74 percent for

“postoperative respiratory failure.” All PSIs had high specificities and positive likelihood ratios, indicating that flagged events were from 65 to 524 times more likely to occur in a hospitalization that had a true adverse outcome (based on NSQIP) than in a hospitalization that did not have a true adverse outcome.

NSQIP events were generally defined more narrowly or precisely than PSI events, except for “postoperative wound dehiscence.” Our alternative PSI definitions improved the sensitivities of all five PSIs, although the only statistically significant increases were for “postoperative respiratory failure” and “postoperative wound dehiscence.” For these two PSIs, we witnessed a tradeoff between sensitivity and PPV, although the decrease in PPV (and positive likelihood ratio) was statistically significant only for “postoperative wound dehiscence.” In version 3.0 of the *PSI* software, AHRQ adopted our alternative definitions for “postoperative physiologic and metabolic derangements,” “postoperative respiratory failure,” and “postoperative sepsis” (AHRQ 2008). For the other two indicators, the modest improvements in sensitivity with our alternative definitions were felt to be outweighed by decreased PPV.

Our research builds on other studies that have attempted to validate or improve the PSIs by linking administrative and chart-abstracted data on adverse events. Zhan et al. (2007) used 2002–2004 Medicare discharge data to compare “postoperative PE/DVT” events identified by ICD-9-CM codes with medical record information on 20,868 beneficiaries. Their sensitivity, specificity, and PPV estimates were 68, 90, and 29 percent, respectively. Our sensitivity and PPV estimates for “postoperative PE/DVT” (56 and 22 percent, respectively) were slightly lower than those reported by Zhan and colleagues, perhaps due to superior coding in non-VA hospitals, variation in the epidemiology of thromboembolic disease, or NSQIP’s more restrictive definition of PE. NSQIP required a “high probability” nuclear scan, but PE may be diagnosed after an “intermediate-probability” scan in a high-risk patient (PIOPED 1990). Because of the indicator’s poor predictive ability, we conducted sensitivity analyses separately for PE and DVT. Using original PSI definitions, we found sensitivity and PPV of 53 and 42 percent, respectively, for *PE alone*, compared with sensitivity and PPV of 29 and 15 percent, respectively, for *DVT alone*. A recent study using administrative data from New York and California suggested that the poor PPV of “postoperative PE/DVT” is largely attributable to preexisting or chronic thromboembolic disease, as 54–57 percent of these diagnoses were reported by hospitals as present on admission (POA) (Houchens, Elixhauser, and Romano 2008). By contrast, the “POA” rates for the other PSIs evaluated herein ranged from 6–7 percent for

“postoperative respiratory failure” to 23–36 percent for “postoperative derangements.”

Gallagher, Cen, and Hannan (2005b) examined the validity of the PSI “accidental puncture or laceration.” Of 67 cases found in New York State administrative data in 2000, 75 percent (50) appeared to be true cases based on medical record abstraction. Three recent studies showed that the sensitivity of “postoperative PE/DVT” (Weller et al. 2004), “postoperative hemorrhage and hematoma” (Shufelt, Hannan, and Gallagher 2005), and “selected infections due to medical care” (Gallagher, Cen, and Hannan 2005a), could be improved by expanding the PSI definitions to capture readmissions within 30 days of a previous surgical hospitalization. AHRQ has recently revised the specifications of “postoperative hemorrhage and hematoma” to enhance sensitivity, based on the findings of Shufelt and colleagues (AHRQ 2008).

Finally, Best et al. (2002) used 1994–1995 VA administrative data to compare ICD-9-CM codes from discharge abstracts to NSQIP chart-abstracted adverse events. Eighty-six percent of the NSQIP indicators had potentially matching ICD-9-CM codes. Of these, only 23 percent had sensitivities > 50 percent and only 31 percent had PPVs > 50 percent. However, the coding of VA inpatient data has substantially improved since this study was conducted (Kashner 1998), so its applicability to present circumstances is limited.

Evaluation of the criterion validity of the PSIs remains a challenge because of the limited data available for analysis. Sensitivity and PPV estimates depend on the accuracy and completeness of chart-abstracted data. Despite our use of NSQIP as the gold standard, it assesses only major noncardiac surgeries and does not capture complications that may result from high-volume minor surgeries. Because of NSQIP’s exclusion criteria, we were only able to match about 50 percent of our flagged PSI hospitalizations with NSQIP adverse events. Further, we examined relatively infrequent events, limiting the power of our analyses.

The generalizability of our findings to non-VA administrative data sets is uncertain. VA inpatient data have a high level of completeness and are not affected by financial incentives for providers to “upcode” diagnoses (Kashner 1998). Some administrative data sets, but not the PTF, permit users to distinguish between conditions that develop during hospitalization and those that are “POA.” Incorporating this information into the PSI logic, as AHRQ now encourages, would be expected to enhance PPV with little effect on sensitivity. Administrative data sets also differ on the number of allowable diagnoses and procedures. The VA PTF Main File contains a maximum of 10 diagnosis fields, and the Bedsection Files (also used in this study) contain up to five codes each, yielding a maximum of 31 unique diagnoses per hospital stay.

By contrast, many state databases contain only 10–15 diagnosis fields. However, this difference may have little practical significance, as we recently found that the VA datasets and the HCUP Nationwide Inpatient Sample had the same average number of diagnosis codes per discharge (6.5).

Ten PSIs were recently submitted to the National Quality Forum (NQF) for consideration as hospital performance measures (AHRQ 2007b). Several of these indicators, including “postoperative physiologic and metabolic derangement,” “postoperative respiratory failure,” and “postoperative sepsis,” were withdrawn because of insufficient evidence of validity, action/ability, or both (although “postoperative respiratory failure” appears to have relatively high sensitivity and PPV). The poor PPV for “postoperative PE/DVT” is correctable, with future implementation of POA coding and proposed new codes for subacute and upper extremity thromboses (Centers for Disease Control and Prevention, 2008). Of the PSIs examined in this study, only two (“postoperative respiratory failure” and “postoperative wound dehiscence”) appear ready for use in efforts beyond quality improvement and screening (based on sensitivity and PPV exceeding 60 percent). Only the latter indicator, among those evaluated here, is now endorsed by the NQF (National Quality Forum 2008). One experimental PSI (“postoperative myocardial infarction”) also appears promising. However, the high positive likelihood ratios of all five PSIs suggest that they are valuable case-finding tools for providers and quality improvement organizations. We should continue to explore more creative algorithms based on diagnosis and procedure codes to improve the sensitivity and PPV of the PSIs. The addition of POA reporting should also help to improve PSI validity (Naessens et al. 2007; Bahl et al. 2008). Ongoing and future research, such as the AHRQ Validation Pilot Project, will build on our results by evaluating both surgical and nonsurgical PSIs, and by reviewing random samples of eligible records, without irrelevant exclusion criteria.

Efforts to improve safety will be facilitated by the availability of valid measures that can be used to evaluate hospital performance. The AHRQ PSIs represent a useful step in this direction, but our results demonstrate that health data agencies, purchaser coalitions, and other sponsors should still proceed cautiously in using administrative data to identify postoperative complications for the purpose of public reporting on hospital safety performance.

ACKNOWLEDGMENTS

Joint Acknowledgement/Disclosure Statement: The authors would like to acknowledge the contribution of clinical expertise by Dr. Ann Borzecki and admin-

istrative support by Dr. Daniel Berlowitz. This research was funded through grant number IIR 02-144 awarded to Dr. Amy Rosen by the Department of Veterans Affairs Health Services Research & Development (HSR&D) Service. The authors would also like to acknowledge the Chiefs of Surgery and the NSQIP Surgical Clinical Nurse Reviewers for their dedication and hard work in assuring the integrity of the NSQIP data.

Disclosures: The first author of this manuscript is a subcontracted member of the Support for Quality Indicators team, based at the Battelle Memorial Institute, which provides ongoing support for public use of the AHRQ Patient Safety Indicators. However, this work was not supported by the AHRQ. Data were provided by the VA's NSQIP, subject to these restrictions: NSQIP has strict data use guidelines to ensure the accuracy and integrity of all studies based on NSQIP data. The present study was approved under the 10/04 version of these guidelines, which included the following text (relevant section excerpted): "All analyses, abstracts, and papers based on your proposal using the NSQIP database must be reviewed by the Executive Committee and approved for publication and/or presentation prior to any submission for publication or presentation at local or national meetings. Executive Committee review and approval is required for all abstracts, manuscripts, and presentations. "Drs. Khuri and Henderson or their designees will be co-authors on all presentations and publications based on the VA National Surgical Quality Improvement Program data." We followed NSQIP's stipulated procedure to ensure that we used their data correctly. Neither the sponsoring organizations nor any of the authors' employers received advance copies of the manuscript. There are no other disclosures.

REFERENCES

- AHRQ 2007a. *Guide to Patient Safety Indicators Version 3.1 (Revised March 2007)*. Rockville, MD: Agency for Healthcare Research and Quality.
- . 2007b. "The AHRQ Quality Indicators in 2007". AHRQ Quality Indicators eNewsletter [accessed on May 8, 2008]. Available at <http://qualityindicators.ahrq.gov/newsletter/2007-February-AHRQ-QI-Newsletter.htm>
- . 2008. *Patient Safety Indicators Technical Specifications Version 3.2 (Revised March 2008)*. Rockville, MD: Agency for Healthcare Research and Quality.
- Bahl, V., M. A. Thompson, T. Y. Kau, H. M. Hu, and D. A. Campbell. 2008. "Do the AHRQ Patient Safety Indicators Flag Conditions that are Present at the Time of Hospital Admission?" *Medical Care* 46 (5): 516–22.
- Best, W. R., S. F. Khuri, M. Phelan, K. Hur, W. G. Henderson, J. G. Demakis, and J. Daley. 2002. "Identifying Patient Preoperative Risk Factors and Postoperative

- Adverse Events in Administrative Databases: Results from the Department of Veterans Affairs National Surgical Quality Improvement Program." *Journal of the American College of Surgeons* 194 (3): 257–66.
- Centers for Disease Control and Prevention. 2008. "ICD-9-CM Coordination and Maintenance Committee 2008 Summary" [accessed on May 13, 2008]. Available at http://www.cdc.gov/nchs/classifications_of_diseases_and_f.htm
- Daley, J., M. G. Forbes, G. J. Young, M. P. Charns, J. O. Gibbs, K. Hur, W. Henderson, and S. F. Khuri. 1997. "Validating Risk-adjusted Surgical Outcomes: Site Visit Assessment of Process and Structure." *Journal of the American College of Surgeons* 185 (4): 341–51.
- Davis, C. L., J. R. Pierce, W. Henderson, C. D. Spencer, C. Tyler, R. Langberg, J. Swafford, G. S. Felan, M. A. Kearns, and B. Booker. 2007. "Assessment of the Reliability of Data Collected for the Department of Veterans Affairs National Surgical Quality Improvement Program." *Journal of the American College of Surgeons* 204 (4): 550–60.
- Fitch, K., S. J. Bernstein, M. S. Aguilar, B. Burnand, J. R. LaCalle, P. Lazaro, M. V. H. Loo, J. McDonnell, J. P. Vader, and J. P. Kahan. 2001. *The RAND/UCLA Appropriateness Method User's Manual*. Los Angeles: RAND Health and RAND Europe.
- Gallagher, B., L. Cen, and E. L. Hannan. 2005a. "Readmission for Selected Infections Due to Medical Care: Expanding the Definition of a Patient Safety Indicator." In *Advances in Patient Safety: From Research to Implementation*, Vol. 2, edited by K. Henriksen, J. B. Battles, E. Marks, and D. I. Lewin, pp 39–50. Rockville, MD: Agency for Healthcare Research and Quality and Department of Defense.
- . 2005b. "Validation of AHRQ's Patient Safety Indicator for Accidental Puncture or Laceration." In *Advances in Patient Safety: From Research to Implementation*, Vol. 2, edited by K. Henriksen, J. B. Battles, E. Marks, and D. I. Lewin, pp. 27–38. Rockville, MD: Agency for Healthcare Research and Quality and Department of Defense.
- HealthGrades. 2008. "The Fifth Annual HealthGrades Patient Safety in American Hospitals Study" [accessed on May 13, 2008]. Available at <http://www.healthgrades.com/media/dms/pdf/patientsafetyinamericanhospitalsstudy2008.pdf>
- Houchens, R., A. Elixhauser, and P. Romano. 2008. "How Often Are Potential 'Patient Safety Events' Present on Admission?" *Joint Commission Journal on Quality and Patient Safety* 34 (3): 154–63.
- Iezzoni, L. I., J. Daley, T. Heeren, S. M. Foley, E. S. Fisher, C. Duncan, J. S. Hughes, and G. A. Coffman. 1994a. "Identifying Complications of Care Using Administrative Data." *Medical Care* 32 (7): 700–15.
- Iezzoni, L. I., J. Daley, T. Heeren, S. M. Foley, J. S. Hughes, E. S. Fisher, C. C. Duncan, and G. A. Coffman. 1994b. "Using Administrative Data to Screen Hospitals for High Complication Rates." *Inquiry* 31 (1): 40–55.
- Kashner, T. M. 1998. "Agreement between Administrative Files and Written Medical Records: A Case of the Department of Veterans Affairs." *Medical Care* 36 (9): 1324–36.

- Khuri, S. F., J. Daley, W. Henderson, G. Barbour, P. Lowry, G. Irvin, J. Gibbs, F. Grover, K. Hammermeister, and J. F. Strempel. 1995. "The National Veterans Administration Surgical Risk Study: Risk Adjustment for the Comparative Assessment of the Quality of Surgical Care." *Journal of the American College of Surgeons* 180 (5): 519–31.
- Khuri, S. F., J. Daley, W. Henderson, K. Hur, J. Demakis, J. B. Aust, V. Chong, P. J. Fabri, J. O. Gibbs, F. Grover, K. Hammermeister, G. III. Irvin, G. McDonald, E. Jr. Passaro, L. Phillips, F. Scamman, J. Spencer, and J. F. Strempel. 1998. "The Department of Veterans Affairs' NSQIP: The First National, Validated, Outcome-based, Risk-adjusted, and Peer-controlled Program for the Measurement and Enhancement of the Quality of Surgical Care. National VA Surgical Quality Improvement Program." *Annals of Surgery* 228 (4): 491–507.
- Kohn, L. T., J. M. Corrigan, and M. S. Donaldson. 2000. *To Err is Human: Building A Safer Health System*. Washington, DC: Institute of Medicine, National Academy Press.
- Lawthers, A. G., E. P. McCarthy, R. B. Davis, L. E. Peterson, R. H. Palmer, and L. I. Iezzoni. 2000. "Identification of In-hospital Complications from Claims Data." *Medical Care* 38 (8): 785–95.
- Leape, L. L. 2005. "Where the Rubber Meets the Road." In *Advances in Patient Safety: From Research to Implementation*, Vol. 3, edited by K. Henriksen, J. B. Battles, E. Marks, and D. I. Lewin, pp 1–3. Rockville, MD: Agency for Healthcare Research and Quality and Department of Defense.
- McCarthy, E. P., L. I. Iezzoni, R. B. Davis, R. H. Palmer, M. Cahalane, M. B. Hamel, K. Mukamal, R. S. Phillips, and D. T. Jr. Davies. 2000. "Does Clinical Evidence Support ICD-9-CM Diagnosis Coding of Complications?" *Medical Care* 38 (8): 868–76.
- McDonald, K. M., P. S. Romano, J. J. Geppert, S. M. Davies, B. W. Duncan, K. G. Shojania, and A. Hansen. 2002. *Measures of Patient Safety Based on Hospital Administrative Data: The Patient Safety Indicators*. Rockville, MD: Agency for Healthcare Research and Quality [accessed on May 8, 2008]. Available at http://qualityindicators.ahrq.gov/downloads/technical/psi_technical_review.zip
- Miller, M. R., A. Elixhauser, C. Zhan, and G. S. Meyer. 2001. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." *Health Services Research* 36 (6, part 2): 110–32.
- Naessens, J. M., C. R. Campbell, B. Berg, A. R. Williams, and R. Culbertson. 2007. "Impact of Diagnosis-timing Indicators on Measures of Safety, Comorbidity, and Case Mix Groupings from Administrative Data Sources." *Medical Care* 45 (8): 781–8.
- National Quality Forum. 2008. "National Quality Forum Endorses Consensus Standards for Quality of Hospital Care" [accessed on May 15, 2008]. Available at <http://www.qualityforum.org/news/releases/051508-endorsed-measures.asp>
- Newcombe, R. G. 1998. "Two-sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods." *Statistics in Medicine* 17 (8): 857–72.
- PIOPED Investigators. 1990. "Value of the Ventilation/Perfusion Scan in Acute Pulmonary Embolism. Results of the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED)." *Journal of the American Medical Association* 263 (20): 2753–9.

- Polancich, S., E. Restrepo, and J. Prosser. 2006. "Cautious Use of Administrative Data for Decubitus Ulcer Outcome Reporting." *American Journal of Medical Quality* 21 (4): 262–8.
- Premier, Inc. 2008. "CMS/Premier Hospital Quality Incentive Demonstration" [accessed on May 8, 2008]. Available at <http://www.premierinc.com/all/quality/hqi/>
- Rivard, P., A. R. Elwy, S. Loveland, S. Zhao, D. Tsilimingras, A. Elixhauser, P. S. Romano, and A. K. Rosen. 2005. "Applying Patient Safety Indicators (PSIs) across Healthcare Systems: Achieving Data Comparability." In *Advances in Patient Safety: From Research to Implementation*, Vol. 2, edited by K. Henriksen, J. B. Battles, E. Marks, and D. I. Lewin, pp 7–25. Rockville, MD: Agency for Healthcare Research and Quality and Department of Defense.
- Romano, P. S. 2003. "Asking Too Much of Administrative Data?" *Journal of the American College of Surgeons* 196 (2): 337–8; author reply 38–9.
- Romano, P. S., J. J. Geppert, S. Davies, M. R. Miller, A. Elixhauser, and K. M. McDonald. 2003. "A National Profile of Patient Safety in U.S. Hospitals." *Health Affairs* 22 (2): 154–66.
- Rosen, A. K., P. Rivard, S. Zhao, S. Loveland, D. Tsilimingras, C. L. Christiansen, A. Elixhauser, and P. S. Romano. 2005. "Evaluating the Patient Safety Indicators: How Well Do They Perform on Veterans Health Administration Data?" *Medical Care* 43 (9): 873–84.
- Rosen, A. K., S. Zhao, P. Rivard, S. Loveland, M. E. Montez-Rath, A. Elixhauser, and P. S. Romano. 2006. "Tracking Rates of Patient Safety Indicators over Time: Lessons from the Veterans Administration." *Medical Care* 44 (9): 850–61.
- Shufelt, J. L., E. L. Hannan, and B. K. Gallagher. 2005. "The Postoperative Hemorrhage and Hematoma Patient Safety Indicator and its Risk Factors." *American Journal of Medical Quality* 20 (4): 210–8.
- Simel, D. L., G. P. Samsa, and D. B. Matchar. 1991. "Likelihood Ratios with Confidence: Sample Size Estimation for Diagnostic Test Studies." *Journal of Clinical Epidemiology* 44 (8): 763–70.
- Weingart, S. N., L. I. Iezzoni, R. B. Davis, R. H. Palmer, M. Cahalane, M. B. Hamel, K. Mukamal, R. S. Phillips, D. T. Jr. Davies, and N. J. Banks. 2000. "Use of Administrative Data to Find Substandard Care: Validation of the Complications Screening Program." *Medical Care* 38 (8): 796–806.
- Weller, W. E., B. K. Gallagher, L. Cen, and E. L. Hannan. 2004. "Readmissions for Venous Thromboembolism: Expanding the Definition of Patient Safety Indicators." *Joint Commission Journal on Quality and Patient Safety* 30 (9): 497–504.
- Zhan, C., J. Battles, Y. Chiang, and D. Hunt. 2007. "The Validity of ICD-9-CM Codes in Identifying Postoperative Deep Vein Thrombosis and Pulmonary Embolism." *Joint Commission Journal on Quality and Patient Safety* 33 (6): 326–31.
- Zhan, C., and M. R. Miller. 2003. "Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries during Hospitalization." *Journal of the American Medical Association* 290 (14): 1868–74.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix S1: NSQIP Case Selection Methodology.

Appendix S2: Sample Characteristics as Compared to Overall VA.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.