Detecting adverse events in surgery: comparing events detected by the Veterans Health Administration Surgical Quality Improvement Program and the Patient Safety Indicators

Hillary J. Mull, Ph.D., M.P.P. a,b,*, Ann M. Borzecki, M.D., M.P.H. c,d, Susan Loveland, M.A.T. a, Kathleen Hickson, Ph.D., R.N. c,f, Qi Chen, Ph.D., M.D. a, Sally MacDonald, R.N., M.N. c, Marlena H. Shin, J.D., M.P.H. a, Marisa Cevasco, M.D. e,f, Kamal M. F. Itani, M.D. b,e,f, Amy K. Rosen, Ph.D. a,b

aCenter for Organization, Leadership, and Management Research, VA Boston Healthcare System, 150 S Huntington Avenue (152M), Boston, MA 02130, USA; bDepartment of Surgery, Boston University School of Medicine, Boston, MA, USA; cCenter for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, MA, USA; dDepartment of Health Policy and Management, Boston University School of Public Health, Boston, MA, USA; eVA Boston Healthcare System, Boston, MA, USA; fHarvard Medical School, Boston, MA, USA

KEYWORDS:
Veterans Health Administration; VA Surgical Quality Improvement Program; Patient Safety Indicators; Adverse surgical events; Administrative data

Abstract

BACKGROUND: The Patient Safety Indicators (PSIs) use administrative data to screen for select adverse events (AEs). In this study, VA Surgical Quality Improvement Program (VASQIP) chart review data were used as the gold standard to measure the criterion validity of 5 surgical PSIs. Independent chart review was also used to determine reasons for PSI errors.

METHODS: The sensitivity, specificity, and positive predictive value of PSI software version 4.1a were calculated among Veterans Health Administration hospitalizations (2003–2007) reviewed by VASQIP (n = 268,771). Nurses re-reviewed a sample of hospitalizations for which PSI and VASQIP AE detection disagreed.

RESULTS: Sensitivities ranged from 31% to 68%, specificities from 99.1% to 99.8%, and positive predictive values from 31% to 72%. Reviewers found that coding errors accounted for some PSI-VASQIP disagreement; some disagreement was also the result of differences in AE definitions.

CONCLUSIONS: These results suggest that the PSIs have moderate criterion validity; however, some surgical PSIs detect different AEs than VASQIP. Future research should explore using both methods to evaluate surgical quality.

Published by Elsevier Inc.

Adverse events (AEs) that occur during inpatient surgical care may result in significant patient harm and attract widespread negative publicity. 1,2 The landmark Institute of Medicine report To Err Is Human 3 emphasized the importance of AE surveillance, but >10 years later, there is little...
consensus on the best method for AE detection. Recent studies have shown that different methods, such as chart review and administrative data–based screening tools, may detect different types of AEs, leading some researchers to postulate that no single method is ideal for identifying all AEs that occur within a hospital. Although medical chart review is still considered the “gold standard” for AE detection, it is possible that methods based on administrative data may detect a more diverse range of inpatient AEs. The aim of this study was to compare surgical AE detection by 2 frequently used surgical AE detection methods, the VA Surgical Quality Improvement Program (VASQIP) and the Patient Safety Indicators (PSIs) developed by the Agency for Healthcare Research and Quality (AHRQ).

The VASQIP method uses nurse chart review to detect a defined list of 20 surgical AEs occurring within 30 days after a major operation. Currently, VASQIP is the most reliable and validated method for surgical AE detection, and it has been adopted in the private sector by the American College of Surgeons (National Surgical Quality Improvement Program [NSQIP]). The program’s major limitation is the highly resource intensive medical record review process; to minimize resource utilization, as many as 30% of major surgeries at Veterans Health Administration (VA) hospitals do not undergo VASQIP review. Although some private sector hospitals use NSQIP, most surgical AE surveillance nationally is performed using administrative data–based methods. Many data-based AE detection tools were developed and validated using either VASQIP or NSQIP chart review data as the gold standard.

The AHRQ PSIs are an example of a widely used administrative data–based medical and surgical AE detection method. Currently, there are 10 PSI algorithms (among 17 PSI algorithms) that are used to screen administrative data for potentially preventable inpatient surgical AEs. Since the AHRQ developed and publicly released the software in 2003, the PSIs have been increasingly used in a number of health care settings to detect AEs and target specific areas for patient safety improvement. Studies have demonstrated good predictive and construct validity for several of the surgical PSIs. Additionally, some of the surgical PSIs were recently endorsed by the National Quality Forum for quality reporting and hospital performance measurement. The Centers for Medicare and Medicaid Services publicly report private sector PSI rates on the Hospital Compare Web site.

As use of the PSIs for public reporting becomes more widespread, there is increasing demand for evidence that the PSIs accurately detect AEs. Several recent studies have examined the criterion validity of the PSIs. These focused on identifying how many of the PSI-flagged hospitalizations had true AEs through explicit nurse chart review to confirm whether AEs occurred (ie, positive predictive value [PPV]). A few studies have also focused on assessing how well the PSIs detect cases in which true AEs occurred (ie, sensitivity) and, conversely, how well the PSIs perform in not flagging cases in which AEs did not occur (ie, specificity). For example, in previous work, we used chart review data from VASQIP as the gold standard to estimate the criterion validity of 5 of the surgical PSIs using fiscal year (FY) 2000 data. We demonstrated that the PSIs had moderate sensitivity and high specificity in detecting AEs, and our findings led to revision of 3 of the 5 PSI algorithms to improve sensitivity or PPV (the updated algorithms were incorporated into PSI version 3.1a). Our results also highlighted the need to better understand the differences between the AEs detected by the PSIs and VASQIP.

This research, part of the VA Health Services Research and Development–funded PSI Validation Study, builds on our previous work by furthering our understanding of the differences between these 2 surgical AE detection methods. In our first study objective, we assessed the criterion validity of a newer version of the PSIs (version 4.1a) using a longer time frame for comparison with VASQIP chart review data. Although the newer version of the PSI algorithms is expected to improve PSI validity, particularly specificity, we hypothesized that the criterion validity estimates in our present study would be similar to our earlier findings. In our second objective, we explored reasons why not all of the PSI events overlap with those of VASQIP using our own nurse chart review. By reviewing cases with AEs identified by only 1 of the 2 methods, we have the opportunity to explore coding issues that may lead to PSI false-positives and false-negatives. We hypothesized that substantive differences in AE definitions between the 2 methods, in addition to coding errors affecting the PSIs, might explain poor overlap in AE detection.

Methods

We used a retrospective study design to compare surgical AEs detected by VASQIP with AEs detected by the PSIs from October 1, 2002, through September 30, 2007 (FY 2003 to FY 2007) in the VA. After our assessment of PSI criterion validity, we sampled cases for chart review to understand why some AEs were detected by only 1 method. Our study was approved by the VA Boston Healthcare System Institutional Review Board and the VA Surgical Quality Data Use Group.

VA Surgical Quality Improvement Program

VASQIP was designed to identify complications occurring up to 30 days after surgery and does not attempt to measure the preventability of an AE. VASQIP nurse reviewers abstract preoperative, intraoperative, and postoperative data from surgeries sampled according to the VASQIP sampling framework: the first 36 VASQIP-eligible surgeries in an 8-day period are assessed, with restrictions on the number of high-volume, low-risk procedures.
<table>
<thead>
<tr>
<th>PSI #10: PMD</th>
<th>VASQIP ARF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Various diabetes codes in secondary diagnosis field or 2. Dialysis as a secondary procedure and various necrosis/renal failure codes in any secondary diagnosis field</td>
<td>In a patient who did not require dialysis preoperatively, worsening of renal dysfunction postoperatively requiring</td>
</tr>
<tr>
<td></td>
<td>• hemodialysis,</td>
</tr>
<tr>
<td></td>
<td>• peritoneal dialysis,</td>
</tr>
<tr>
<td></td>
<td>• hemofiltration,</td>
</tr>
<tr>
<td></td>
<td>• hemodiafiltration, or</td>
</tr>
<tr>
<td></td>
<td>• ultrafiltration.</td>
</tr>
<tr>
<td>Note: Part 1 of the PSI algorithm does not align with the VASQIP AE</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PSI #11: postoperative respiratory failure</th>
<th>VASQIP FW or on ventilator &gt;48 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Acute respiratory failure code in any secondary diagnosis field or 2. Reintubation (96.04) or mechanical ventilation (96.70, 96.71, 96.72), codes in any secondary procedure field in the following time frame:</td>
<td></td>
</tr>
<tr>
<td>• (96.04) ≥ 1 d after the major operating room procedure code</td>
<td></td>
</tr>
<tr>
<td>• (96.70 or 96.71) ≥ 2 d after the major operating room procedure code</td>
<td></td>
</tr>
<tr>
<td>• (96.72) ≥ 0 d after the major operating room procedure code</td>
<td></td>
</tr>
<tr>
<td>Note: Part of the PSI #11 algorithm overlaps with part of the VASQIP AE definition for “cardiac arrest requiring CPR”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PSI #12: PE/DVT</th>
<th>VASQIP PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Code for PE in any secondary diagnosis field or 2. Codes for DVT in any secondary diagnosis field</td>
<td>Definition: 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. Enter “yes” if the patient has a V/Q scan interpreted as high probability of PE or a positive pulmonary angiogram or positive CT angiogram or positive spiral CT exam. Treatment usually consists of</td>
</tr>
<tr>
<td></td>
<td>• initiation of anticoagulation therapy and</td>
</tr>
<tr>
<td></td>
<td>• placement of mechanical interruption (e.g., Greenfield filter), for patients in whom anticoagulation is contraindicated or already instituted.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VASQIP DVT/thrombophlebitis</th>
</tr>
</thead>
</table>

| Definition: 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. The patient must be treated with anticoagulation therapy, and/or placement of a vena cava filter or clipping of the vena cava. |
diagnosis field (ie, strep septicemia, staph, pneumococcal septicemia, septicemia due to anaerobes, septic shock, postoperative shock, septicemia due to gram-negative organism, hemophilus influenza, *Escherichia coli*, or SIRS due to infectious process with or without organ dysfunction)

**VASQIP SS**

1. **Sepsis:** Sepsis is the systemic response to infection. Report this variable if the patient has clinical signs and symptoms of SIRS. SIRS is a widespread inflammatory response to a variety of severe clinical insults. This syndrome is clinically recognized by the presence of ≥2 of the following:
   - Temperature >38°C or <36°C
   - HR >90 beats/min
   - RR >20 breaths/min or PaCO₂ <32 mm Hg (<4.3 kPa)
   - WBC count >12,000/mm³, <4,000/mm³, or >10% immature (band)forms
   - Anion gap acidosis, defined by either
     - \[\text{Na} + \text{K} - (\text{Cl} + \text{HCO}_3 \text{ (or serum CO}_2\text{)})\] (if >16, anion gap acidosis is present)
     - \[\text{Na} - (\text{Cl} + \text{HCO}_3 \text{ (or serum CO}_2\text{)})\] (if >12, anion gap acidosis is present)
   And by 1 of the following:
   - Positive blood culture
   - Clinical documentation of purulence or positive culture from any site thought to be causeative

2. **Severe sepsis/septic shock:** Sepsis is considered severe when it is associated with organ and/or circulatory dysfunction. Report this variable if the patient has the clinical signs and symptoms of SIRS or sepsis and documented organ and/or circulatory dysfunction. Examples of organ dysfunction include: oliguria, acute alteration in mental status, acute respiratory distress. Examples of circulatory dysfunction include hypotension, requirement of inotropic or vasopressor agents.

*From 2001 to June 2004, the VASQIP definition of sepsis was as follows:* If the primary physician or the chart states that the patient had SS within the 30 d postoperatively, choose from the following choices for sepsis. If neither is present, follow these definitions and choose the most applicable:

1. **Sepsis:** Definitive evidence of infection, plus evidence of a systemic response to infection. This systemic response is manifested by ≥2 of the following conditions:
   - Temperature >38°C or <36°C
   - HR >90 beats/min
   - RR >20 breaths/min or PaCO₂ <32 mm Hg (<4.3 kPa)
   - WBC count >12,000/mm³, <4,000/mm³, or >10% immature (band)forms

2. **Septic shock:** Sepsis with hypotension despite adequate fluid resuscitation combined with perfusion abnormalities that may include, but are not limited to, lactic acidosis, oliguria, or an acute alteration in mental status. Patients who are on inotropic or vasopressor agents may not be hypotensive at the time that perfusion abnormalities are measured.

**PSI #14: postoperative WD**

Discharges with code for reclosure of postoperative disruption of abdominal wall (54.61) in any procedure field

The VASQIP data dictionary definitions were valid from 2002 to 2007, with the exception of sepsis, as described in the table. PSI exclusion criteria are not presented. PSI AEs are detected before discharge, whereas VASQIP AEs can occur up to 30 days postoperatively, regardless of inpatient status.

AE = adverse event; AHRQ = Agency for Healthcare Research and Quality; ARF = acute renal failure; CPR = cardiopulmonary resuscitation; CT = computed tomographic; DVT = deep vein thrombosis; FW = failure to wean; HR = heart rate; PaCO₂ = partial pressure of carbon dioxide; PE = pulmonary embolism; PMD = postoperative physiologic and metabolic derangement; PSI = Patient Safety Indicator; RR = respiratory rate; R/UI = reintubation or unplanned intubation; SIRS = systemic inflammatory response syndrome; SS = systemic sepsis; VASQIP = VA Surgical Quality Improvement Program; V/Q = ventilation/perfusion; WBC = white blood cell; WD = wound dehiscence.
VASQIP nurses document each postoperative AE and the date the AE occurred. Definitions of postoperative events were consistent from FY 2003 to FY 2007; we used the FY 2007 VASQIP data dictionary.

**Agency for Healthcare Research and Quality Patient Safety Indicators**

The PSIs use International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis and procedure codes from patient administrative data to detect potentially preventable inpatient AEs. Each PSI algorithm uses eligibility criteria to exclude cases in which the patient’s diagnoses or procedures indicate that the condition was present on admission or that the patient as at high risk for the event. Applying the PSI software to administrative data generates a flag for cases that meet the eligibility criteria (ie, in the PSI denominator) and a flag if the patient has the PSI event (ie, the PSI numerator).

**Comparison of Patient Safety Indicator-detected and VA Surgical Quality Improvement Program-detected events**

Five of the AHRQ surgical PSIs identify AEs similar to 7 of the VASQIP postoperative complications (see Table 1). On the basis of our prior work, we recognized a priori that there were obvious differences in how the PSI and VASQIP AEs were defined; therefore, we eliminated as many of these differences as possible before identifying cases for chart review. For example, in comparing PSI postoperative physiologic and metabolic derangement (PMD) and VASQIP acute renal failure (ARF), we dropped the diabetes component of the PSI algorithm because VASQIP did not include diabetes in its definition of ARF. Additionally, because the PSIs focus on AEs that occur only during hospitalizations, we excluded VASQIP AEs if the VASQIP-recorded dates of the events occurred after the discharge dates in the Patient Treatment File (PTF).

**Data sources**

We ran version 4.1a of the AHRQ PSIs on VA administrative data from the PTF for FY 2003 to FY 20007 (n = 2,813,169 discharges). A credentialed VA programmer (S.L.) extracted PTF data from the Austin Information Technology Center. The VA PTF includes ICD-9-CM diagnosis and procedure codes, as well as admission and discharge dates, for each inpatient episode in the VA health care system. The process of applying the AHRQ PSIs to VA data has been described in previous studies.

We obtained FY 2003 to FY 2007 VASQIP data on all inpatient surgeries reviewed by submitting a request to VA Patient Care Services, National Office of Surgery (n = 290,542 surgeries). Our requested sample of VASQIP data included whether any of 7 relevant postoperative complications occurred and data on all patient risk factors for these AEs. We also requested the dates and Current Procedural Terminology codes characterizing surgical procedures.

Chart review data in our study came from the PSI Validation Study, as well as from the VA electronic medical record system, VistA. Because there are some limitations to national VistA access for researchers (eg, radiologic images are not always available) compared with the VistaA access available to local VA staff members, it is plausible that VASQIP nurses with local VistA access reviewed more complete patient records than the chart reviewers in our study.

**Merging administrative and VA Surgical Quality Improvement Program data**

VASQIP generates a record for each surgery reviewed, whereas the administrative data in the PTF are based on hospitalization. We followed an algorithm developed in our previous study to merge PTF and VASQIP data (ie, the operation date recorded in VASQIP had to fall between the admission and discharge dates recorded in the PTF for cases to match). Our final sample included 283,397 matched surgeries (97.5% of the VASQIP sample) that occurred during 268,771 hospitalizations in 117 VA hospitals (Fig. 1).

Not all cases could be matched. We had 2,544,327 unmatched PTF hospitalizations, because these hospital stays involved procedures performed outside an operating room or surgeries that were not assessed by the VASQIP. We were unable to match 7,145 VASQIP-assessed surgeries to the PTF for several reasons, including patient discharge dates after FY 2007 and apparent differences in definition of “outpatient” surgery between VASQIP and the PTF (ie, some of the unmatched VASQIP surgeries were labeled as inpatient in the VASQIP data but were considered as outpatient in VA administrative data).

**Assessment of Patient Safety Indicator criterion validity**

We assessed criterion validity by comparing the presence of a PSI-detected AE with a VASQIP-detected AE among hospitalizations that met the PSI eligibility criteria in our matched sample. We estimated sensitivity, specificity, and PPV for each PSI using the VASQIP findings as the gold standard. Sensitivity evaluates the extent to which the PSIs flagged cases with true AEs on the basis of the total number of cases with AEs identified by VASQIP, specificity examines the number of cases that were not flagged by PSIs and did not have AEs detected by VASQIP, and PPV is a calculation of how many cases with true AEs, as determined by the VASQIP, were identified among all the PSI-flagged hospitalizations. We calculated 95% confidence intervals for each estimate and

Assessment of disagreement: VA Surgical Quality Improvement Program-only cases

For each PSI-VASQIP comparison group, we began by examining cases for which only VASQIP detected AEs (VASQIP-only cases). Our analysis of disagreement included several steps. First, we analyzed the administrative data to determine if there were coding issues that may have explained disagreement. For example, part of the PSI criteria for postoperative respiratory failure specifies a secondary ICD-9-CM procedure code of mechanical ventilation within a specific time frame of \( R \) days after the surgical procedure (depending on the code; see Table 1). We searched the administrative data for VASQIP-only cases coded with mechanical ventilation outside the respiratory failure time frame as an explanation for disagreement in AE definition between the PSIs and VASQIP.

In a second step, we performed chart review on a sample of 20 VASQIP-only cases for each comparison group. Because our goal was to explore possible reasons for disagreement, and we expected that most of the disagreement would be explained by differences in definition or coding errors, we used nonprobability, or convenience, sampling to identify cases for chart review. One of 2 trained research nurses (S.M., K.H.) abstracted patients’ medical records in VistAWeb using the PSI-specific chart abstraction tools developed for the PSI Validation Study. Nurse chart reviewers were provided with the VASQIP AE definitions but did not undergo formal VASQIP training. The nurses were familiar with each PSI abstraction tool and had achieved \( \geq 90\% \) agreement for each PSI reviewed; as a result, we did not repeat interrater reliability tests for this project. The chart review process involved searching the patients’ medical records (ie, VistAWeb) for the VASQIP-detected AE and then examining the characteristics of the event to determine why the PSI failed to detect it. If the nurse chart reviewer could find no evidence of either a difference in AE definition or a coding error, the case was labeled “unable to determine.” A study clinician (A.M.B.) reviewed all the findings.

Assessment of disagreement: Patient Safety Indicator-only cases

To assess PSI-only cases, we compared our findings with chart review results from the larger PSI Validation Study. The PSI Validation Study used independent chart review of VistAWeb to detect true AEs among PSI-flagged cases. For each PSI-VASQIP comparison group, we matched our PSI-only cases to PSI Validation Study results and identified PSI true-positives and false-positives. We then selected a random sample of up to 20 PSI true-positive cases and performed chart review using VistAWeb to determine why these cases were not identified by VASQIP (eg, because of differences in AE definitions). As with VASQIP-only chart reviewed cases, if there was no evidence of a difference in AE definition, the case was labeled “unable to determine.” Findings were reviewed by a study clinician.

Results

Eighty-nine percent of the 268,771 hospitalizations in our PTF-VASQIP matched sample were eligible for \( \geq 1 \) of the 5 PSIs we examined, and 6,100 of these were flagged. In comparison, VASQIP found \( \geq 1 \) of the AEs of interest in

### Figure 1: Results of merge between VA administrative and VASQIP data sets.

<table>
<thead>
<tr>
<th>PTF Only</th>
<th>Matched (by SCRSSN and dates of care)</th>
<th>VASQIP Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,544,327 hospitalizations</td>
<td>283,397 surgeries</td>
<td>7,145 surgeries</td>
</tr>
<tr>
<td>290,542 surgeries</td>
<td>208,771 hospitalizations</td>
<td>117 VA hospitals</td>
</tr>
</tbody>
</table>

NOTES:
VASQIP = VA Surgical Quality Improvement Program
PTF = Patient Treatment File
DRG = Diagnosis-Related Group
SCRSSN = Scrambled Social Security number.
approximately 15,000 hospitalizations (6%); however, most of these occurred after discharge or in cases that did not meet PSI eligibility criteria. Table 2 shows the number of hospitalizations in our matched sample that were eligible for the respective PSI, as well as the number of VASQIP-detected and PSI-flagged events.

### Assessment of Patient Safety Indicator criterion validity

We calculated sensitivity, specificity, and PPV for each of the 5 PSIs using the VASQIP-detected AEs as the gold standard (Table 2). The PSIs and VASQIP detected similar rates of renal failure (PSI PMD and VASQIP ARF), respiratory failure (PSI respiratory failure and VASQIP failure to wean or reintubation or unplanned intubation), and sepsis (PSI sepsis or VASQIP systemic sepsis) AEs; however, the PSIs flagged more than twice as many hospitalizations with either pulmonary embolism (PE) or deep vein thrombosis (DVT) AEs and half as many cases with wound dehiscence (WD) events. We hypothesized that our present study findings would be similar to the criterion validity results of our previous research using FY 2001 data and PSI version 3.0.23 Our findings were comparable, other than an improvement in PPV of PE or DVT.

### Assessment of disagreement

We found evidence to support our hypothesis that differences in definition explained disagreement in AE detection between PSI and VASQIP; however, we also found evidence of coding errors in the VA administrative data. Table 3 presents the results of our chart review of VASQIP-only and PSI-only cases. We found that differences in AE definitions explained disagreement in AE detection for all PSI-VASQIP comparisons except WD. We also found that each of the 5 PSIs failed to flag some AEs detected by VASQIP because of coding errors, particularly for PE or DVT and PSI WD and VASQIP WD (75% of the chart review sample for each PSI). Finally, we found many VASQIP-only (between 5% and 45% of the chart review sample) and PSI-only (between 69% and 100% of the chart review sample) cases that did not have discernible differences in AE definitions or evidence of coding errors. We present detailed results for each PSI-VASQIP comparison, highlighting specific areas in which improvements in definitions or methods may be warranted, in the Appendix.

### Comments

Using a large sample of VA acute hospitalizations and 5 years of data, we compared AEs identified by PSI version 4.1a and VASQIP and estimated PSI criterion validity. We found high specificity estimates for all 5 PSIs; low sensitivity for the PSIs PMD, sepsis, and WD (31%–48%), and moderate sensitivity estimates for the remaining
The PSIs PE or DVT and sepsis had low PPV estimates (31% and 44%, respectively), and the PSIs PMD, respiratory failure, and WD had moderate to good PPV estimates of 66%, 72%, and 72%, respectively.

We confirmed our hypothesis that differences in PSI and VASQIP AE definitions contribute to poor overlap in AE detection for some PSIs. When examining reasons for discrepancies in VASQIP-identified and PSI-identified events, we found that relatively subtle differences in AE definitions led VASQIP to detect a broader range of events that are beyond the scope of AE detection on the basis of administrative data. For example, we found several PMD cases with advanced directives; these resulted in termination or avoidance of procedures that were necessary components of the PSI algorithm. In the case of sepsis, the VASQIP definition included patient symptoms and signs, such as fever, as well as laboratory values that are generally not part of administrative data sets. At present, the PSI definitions are limited to the constraints of ICD-9-CM coding, which does not include signs and symptoms. In the future, laboratory and vital sign data may be used to modify the definitions of PMD and sepsis to identify more cases without sacrificing specificity.

We also found that 2 of the PSIs had more narrow definitions of AEs than VASQIP. Expanding the PSI numerator criteria to include patients reintubated on the same day as the operation in the case of respiratory failure or to include some cases of bacteremia, urinary tract infection, or pneumonia in the case of sepsis would increase the overlap with VASQIP. The differences in definition for these 2 PSIs reflect the different purpose of the PSIs compared with VASQIP: the PSIs were designed to maximize specificity and to detect only those AEs that may be potentially preventable. Comparing respiratory failure and sepsis AEs detected by the PSIs and VASQIP may not be appropriate given these differences in definition.

Our chart review results also demonstrated that coding issues affected each of the PSIs we evaluated. Previous studies have cited coding issues as a significant factor limiting the use of the PSIs for public reporting and performance measurement. We expected that the increased attention to coding postoperative complications might improve the PPV of some of the PSIs compared with our previous results. The improvement in PPV for PE or DVT from 22% to 31% may reflect improvements in the documentation or coding of PE and DVT events, a focus of many health care systems, including the VA, since the introduction of the Surgical Care Improvement Program in 2005. Although we found instances of coding errors in each PSI-VASQIP comparison group, there were other factors that explained disagreement.

First, we were unable to determine why the VASQIP and PSIs did not agree on the detection of several of the AEs in our chart review sample. Studies have shown that

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Reasons for disagreement in surgical AE detection between PSI version 4.1a and VASQIP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparison group</strong></td>
<td><strong>Chart review results</strong> for PSI-only cases</td>
</tr>
<tr>
<td><strong>VASQIP AE</strong></td>
<td><strong>Total VASQIP-only cases reviewed</strong></td>
</tr>
<tr>
<td>#10: PMD ARF</td>
<td>20/2,288</td>
</tr>
<tr>
<td>#11: RF FW</td>
<td>20/683</td>
</tr>
<tr>
<td>#12: PE/DVT</td>
<td>20/554</td>
</tr>
<tr>
<td>#13: sepsis</td>
<td>20/554</td>
</tr>
<tr>
<td>#14: WD Dehiscence</td>
<td>20/862</td>
</tr>
</tbody>
</table>

AE, adverse event; ARF, acute renal failure; DVT, deep vein thrombosis; FW, failure to wean; Na, not applicable; PE, postoperative pulmonary embolism; PMD, postoperative physiologic and metabolic derangement; PSI, Patient Safety Indicator; RF, postoperative respiratory failure; R/UI, reintubation/unplanned intubation; SS, systemic sepsis; VASQIP, VA Surgical Quality Improvement Program.
conducting chart review to identify rare AEs, particularly in complicated patients, can be challenging. Our nurse chart reviewers did not undergo formal VASQIP chart review training, and it is possible that many of the “unable to determine” cases represented differences in AE definition from the PSIs that were described in the VASQIP chart review guidelines but were not present in the VASQIP AE definitions. It is also possible that characteristics of the AE that justified the VASQIP determination were not available through national VistAWeb review. Finally, these cases may also represent possible VASQIP errors; however, we are unable to confirm this. Furthermore, VASQIP data have been shown to be highly reliable.

Given that the PSIs and VASQIP are tools to assess quality of care, it may be useful for health care systems to use these methods in conjunction to identify a broader range of AEs with greater accuracy. Currently, the VA and many private sector hospitals use VASQIP or NSQIP to measure surgical outcomes and target quality improvement efforts. In addition, the Centers for Medicare and Medicaid Services Hospital Compare Web site uses the AHRQ PSIs respiratory failure, PE or DVT, and WD to report hospital performance. Previous research has found poor overlap between surgical AEs detected by the PSIs and NSQIP; in our assessment of the overlap, we found that many of the AEs detected only by the PSIs or the VASQIP were nonetheless true events that point to potential problems. Polk et al proposed that successful quality improvement initiatives depend on a combination of administrative and chart review data. Using the PSIs, in addition to VASQIP, to screen for AEs may be a cost-effective approach to detecting a wide range of AEs for surgical quality assessment. With the advent of improved coding specificity as a result of International Classification of Disease, Tenth Revision, Clinical Modification codes, the PSIs may become a more accurate and reliable source of hospital surgical performance data. Until that time, hospitals may detect a greater yield of true AEs by performing VASQIP nurse review of PSI-flagged cases.

One of our study’s strengths is that by performing chart review, we were able to explore the extent to which differences in definition and administrative data-based coding may have affected the lack of overlap in AE detection. However, we also discovered that our chart reviewers could not always determine through medical record review why the PSIs and VASQIP disagreed, possibly as a result of using VistAWeb or because VASQIP nurses can contact patients and clinicians for data beyond what is recorded in the electronic medical record. Our results may have been limited by the chart review process in that we could not obtain VASQIP guidelines, and we used a small, convenience sample of charts for review. Despite these limitations, our study expands on our previous research with a larger data set and a more recent version of the PSIs. To our knowledge, it is the first study to use independent chart review, including re-review by a physician, to compare PSI and VASQIP AE detection.

Our findings generate a number of recommendations. First, it is critically important that VA hospitals, as well as private sector institutions, improve coding processes. In addition, although VASQIP is a highly valuable method of assessing surgical quality, our chart review results suggest that it is not always possible for non-VASQIP nurses to understand VASQIP AE detection. On the basis of our findings, we recommend that future studies comparing PSI-detected and VASQIP-detected events use VASQIP-trained nurses for chart review. Finally, because VASQIP AE definitions are not perfectly aligned with the PSI algorithms, we recommend using independent chart review as the gold standard to measure PSI criterion validity.

References

7. Olsen S, Neale G, Schwab K, et al. Hospital staff should use more than one method to detect adverse events and potential adverse events: incident reporting, pharmacist surveillance and local real-time record review may all have a place. Qual Saf Health Care 2007;16:40–4.
14. Davenport DL, Holtsapple CW, Conigliaro J. Assessing surgical quality using administrative and clinical data sets: a direct comparison of the University HealthSystem Consortium Clinical Database and the


Appendix: detailed assessment of disagreement

PSI PMD (PSI #10) and VASQIP ARF

Overlap. There were 214 hospitalizations that matched both the PSI definition of PMD (a dialysis procedure and an ARF diagnosis), and the VASQIP definition of ARF (postoperative renal dysfunction requiring dialysis in a patient who did not require dialysis preoperatively).

VASQIP only. We examined the administrative codes of the 228 VASQIP-only cases to determine why they were not flagged by the PSIs and found that 38 cases (17%) met the PSI dialysis criterion only, and 131 other cases (57%) had ARF diagnoses only. There were 59 VASQIP-only hospitalizations (26%) with neither the procedure nor the diagnosis codes required in the PSI algorithm.

In the first phase of chart review, we assessed a random sample of 20 of the 228 VASQIP-only hospitalizations and found that some (6 cases [30%]) would not have been detected by the PSI because of AE definition differences (see Table 3). These cases should have received dialysis but did not for several reasons, including death and advanced directives; thus, they met the VASQIP AE definition, which specifies only the requirement for dialysis, but did not meet the PSI definition because they could not be coded for dialysis. We found 5 VASQIP-only cases (25%) that fit the PSI PMD clinical definition but were lacking appropriate codes (ie, PSI false-negatives) and 9 “unable to determine” cases (45%). Seven of these cases did not have evidence of dialysis (or of refusing dialysis) in the medical record, and 2 had unclear histories of renal failure.
The remaining 11 cases (69%) could not determine why VASQIP did not identify AEs in the patient, we conclude that these 4 cases are evidence of preoperative dialysis that occurs outside of the VA and can be detected only through patient communication. We found 5 cases (31%) that were not flagged by VASQIP because of differences in definition (see Table 3). In 4 of these cases, the VASQIP data showed that the patient had preoperative dialysis, a VASQIP exclusion criterion. Although preoperative dialysis is also a PSI exclusion criterion, these cases were flagged by the PSI, and no dialysis was detected preoperatively by the nurse reviewers in the patient medical record. Because VASQIP chart review data includes contact with the patient, we conclude that these 4 cases are evidence of preoperative dialysis that occurs outside of the VA and can be detected only through patient communication. We could not determine why VASQIP did not identify AEs in the remaining 11 cases (69%).

**PSI postoperative respiratory failure (PSI #11) and VASQIP failure to wean (FW) or VASQIP reintubation/unplanned intubation**

**Overlap.** Among the PTF-VASQIP matched sample, we identified 1,468 hospitalizations with respiratory failure and either the VASQIP FW or reintubation/unplanned intubation AE. The PSI algorithm and the VASQIP definitions differed on the basis of the timing of the reintubation or ventilation procedures, which may explain disagreement in findings.

**VASQIP only.** When we examined the coding for a reintubation procedure among the 683 VASQIP-only cases, 177 had the PSI algorithm–specified reintubation or ventilation code but were not flagged by the PSIs because of the procedures’ timing.

Our first round of chart review focused on 20 of the 683 VASQIP-only cases and found that differences in AE definition explained much of the disagreement in findings (13 of the cases [65%]; see Table 3). Of these, 5 of the VASQIP-only cases were coded by VASQIP as FW and did not have any of the PSI codes for respiratory failure, and 8 were reintubated on the same day as the initial operation; these do not meet the PSI numerator definition (see Table 1). We also found 6 PSI false-negatives (30%); 5 of these had events that fit the timing of the PSI algorithm but were missing either ventilation or intubation codes, and another case was not captured by the PSI despite having an appropriately timed intubation code, because there was an earlier intubation and ventilation code on the same day as the operation (the PSI algorithm is set up such that only the first procedure code is examined; downstream codes are ignored, even if appropriately timed). The medical chart of the 1 “unable to determine” VASQIP-only case (5%) indicated that the reintubation procedure occurred intraoperatively (the patient inadvertently coughed up his endotracheal tube).

**PSI only.** We also examined the 579 PSI-only cases, and matched PSI Validation Study chart review results to identify 7 PSI false-positives and 15 PSI true-positives. Given the differences in AE definition, we expected that some of the 15 PSI-only true-positives would not have met the FW or reintubation definitions; however, our chart review findings did not confirm this expectation. Only 1 of the PSI-only cases reviewed was the result of differences in AE definition because of the type of ventilation used. We were unable to determine why VASQIP did not identify AEs in the remaining 14 cases (93%) (see Table 3).

**PSI PE or DVT (PSI #12) and VASQIP PE or VASQIP DVT**

**Overlap.** We found 1,034 hospitalizations that were PSI flagged for PE or DVT (defined as a secondary diagnosis of PE or DVT) and that had VASQIP-detected PE or DVT (both the VASQIP PE and DVT definitions require confirmation of the AE through diagnostic tests or treatment procedures).

**VASQIP only.** We did not find any evidence of differences in the PSI and VASQIP AE definitions among the 20 cases randomly selected for chart review out of the 554 VASQIP-only PE or DVT cases (see Table 3). Fifteen of the VASQIP-only cases (75%) were PSI false-negatives because the administrative data did not code the PE or DVT. We could not determine why the PSIs did not detect the VASQIP AEs in the 5 remaining cases (25%).

**PSI only.** Of the 2,289 PSI-only cases, we found 479 coded with secondary diagnoses of PE and 1,881 with DVT. We compared the PSI-only PE and DVT cases with PSI Validation Study chart reviewed cases and found 127 matches. The PSI Validation Study nurses identified 86 PSI false-positives and 41 PSI true-positives. In our second phase of chart review, nurses reviewed 20 of the PE or DVT true-positives and determined that 6 cases (30%) were not detected by VASQIP because of differences in definition; these cases did not have sufficient diagnostic testing evident in the medical record to meet the VASQIP definition. Nurses’ review of the remaining 14 PSI-only PE or DVT cases (70%) could not determine why VASQIP did not identify PE or DVT (see Table 3).

**PSI postoperative sepsis (PSI #13) and VASQIP systemic sepsis**

**Overlap.** The definition of sepsis varies between the PSIs and the VASQIP, which may account for why only 175 cases, among 397 hospitalizations flagged by the PSIs and 570 hospitalizations with AEs identified by VASQIP, were detected by both methods. The PSI algorithm includes a limited list of diagnosis codes, whereas the VASQIP definition requires a combination of factors, many of which
are not recorded in administrative data (ie, fever). In June 2004, VASQIP revised its sepsis definition to include a diagnosis of systemic inflammatory response syndrome; the PSI algorithm includes systemic inflammatory response syndrome but is more narrowly defined.

**VASQIP only.** In our first phase of chart review, nurses reviewed 20 of the 395 VASQIP-only cases and found 11 cases (55%) for which the VASQIP and PSI definitions did not align (see Table 3). Although the patients had systemic inflammatory response syndrome, they were coded for bacteremia, urinary tract infections, or pneumonia; none were coded with ICD-9-CM codes that were part of the PSI definition of sepsis. Eight patients (40%) had clinical documentation of sepsis in the medical record but were not flagged by the PSIs because of coding errors. Nurse chart reviewers were unable to determine why the PSIs did not detect 1 of the VASQIP-only cases (5%).

**PSI only.** Of the 215 PSI-only cases, we matched 30 cases with PSI Validation Study chart review results for sepsis. Of these, 15 were PSI false-positives. The remaining 15 cases were reviewed again, and nurses found that 4 (27%) met the PSI criteria but did not meet the VASQIP sepsis definition (see Table 3). The remaining 11 cases (73%) could not be determined (see Table 3).

---

**PSI postoperative WD (PSI #14) and VASQIP dehiscence**

**Overlap.** Of the 1,254 hospitalizations with VASQIP-detected AEs, and the 544 hospitalizations flagged by the PSIs, only 392 were detected by both methods. Although the overlap in detection was poor in this comparison group, the PSI and VASQIP definitions of dehiscence appeared to match well (separation of the surgical wound with disruption of the fascia).

**VASQIP only.** In our first round of chart review, nurses reviewed 20 of the 862 VASQIP-only cases and detected 15 PSI false-negative cases (75%) that had undergone wound reclosure procedures but were missing the appropriate code. Nurse reviewers did not find any cases of differences in AE definition. The remaining 5 cases (25%) could not be determined (see Table 3).

**PSI only.** In our second phase of chart review, we matched 30 of the 152 PSI-only cases to the PSI Validation Study chart review results. Of these, 9 were PSI false-positives. Our study nurses reviewed 20 of the 21 PSI true-positives and could not determine why VASQIP did not identify the AEs (see Table 3).