# 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, PhD, MPP**[a,b], **Ann M. Borzecki, MD, MPH**[c,d], **Susan Loveland, MAT**[a], **Kathleen Hickson, PhD, RN**[c,f], **Qi Chen, PhD, MD**[a], **Sally MacDonald, RN, MN**[c], **Marlena H. Shin, JD, MPH**[a], **Marisa Cevasco, MD**[e,f], **Kamal M. F. Itani, MD**[b,e,f], and **Amy K. Rosen, PhD**[a,b]

[a]Center for Organization, Leadership, and Management Research, VA Boston Healthcare System, Boston, MA

[b]Department of Surgery, Boston University School of Medicine, Boston, MA

[c]Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford, MA

[d]Department of Health Policy and Management, Boston University School of Public Health, Boston, MA

[e]VA Boston Healthcare System, Boston, MA

[f]Harvard Medical School, Boston, MA

## Keywords

Veterans Health Administration; VA Surgical Quality Improvement Program; Patient Safety Indicators; Adverse Surgical Events; Administrative Data

## Background

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*, emphasized the importance of AE surveillance;[3] yet more than ten years later, there is little consensus on the best method for AE detection.[4] Recent studies show that different methods, such as chart review and administrative data-based screening tools, may detect different types of AEs,[5,6] leading some researchers to postulate that no single method is ideal for identifying all AEs that occur within a hospital.[7] While medical chart review is still considered the "gold standard" for AE detection,[8] it is possible that methods based on administrative data may detect a more diverse range of inpatient AEs. This study compares surgical AE detection by two frequently used surgical

Author for correspondence and reprint requests: Hillary J. Mull, Ph.D., M.P.P., Center for Organization, Leadership, and Management Research (COLMR), VA Boston Healthcare System, 150 S. Huntington Ave. (152M), Boston, MA 02130, Telephone number (857) 364-2766; Fax number (857) 364-4438, hjmull@bu.edu or hillary.mull@va.gov.

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 following a major operation.[9] Currently, the VASQIP is the most reliable and validated method to date for surgical AE detection,[10,11] and has been adopted in the private sector by the American College of Surgeons (ACS-NSQIP).[12] The program's major limitation is the highly resource-intensive medical record review process; to minimize resource utilization, as many as 30% of major VA surgeries do not undergo VASQIP review.[13] Although some private sector hospitals use ACS-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 ACS-NSQIP chart review data as the gold standard.[14,15]

The AHRQ PSIs are an example of a widely used administrative data-based medical and surgical AE detection method.[16] Currently, there are ten PSI algorithms (out of 17 PSI algorithms) that are used to screen administrative data for potentially preventable inpatient surgical AEs.[17] Since AHRQ developed and publicly released the software in 2003, the PSIs have been increasingly used in a number of healthcare settings to detect AEs and target specific areas for patient safety improvement.[18–20] Studies have demonstrated good predictive and construct validity for several of the surgical PSIs.[21–23] Additionally, some of the surgical PSIs were recently endorsed by the National Quality Forum (NQF) for quality reporting and hospital performance measurement.[24] The Centers for Medicare and Medicaid Services (CMS) publicly report private sector PSI rates on the Hospital Compare website.[25]

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 a true AE through explicit nurse chart review to confirm whether an AE occurred (i.e., positive predictive value (PPV)).[26–28] A few studies have also focused on assessing how well the PSIs detect cases in which a true AE occurred (i.e., sensitivity), and conversely, how well the PSIs perform in not flagging cases in which an AE did not occur (i.e., specificity).[29] For example, in previous work, we used chart review data from VASQIP as the gold standard to estimate the criterion validity of five of the surgical PSIs using fiscal year (FY) 2000 data.[23] We demonstrated that the PSIs had moderate sensitivity and high specificity in detecting AEs, and our findings led to revision of three of the five PSI algorithms to improve sensitivity or PPV (the updated algorithms were incorporated into PSI version 3.1a).[30] Our results also highlighted the need to better understand the differences between the AEs detected by the PSIs and VASQIP.

This research, part of a Veterans Health Administration (VA) Health Services Research and Development (HSR&D)-funded PSI Validation Study,[28] builds on our previous work by furthering our understanding of the differences between these two surgical AE detection methods. In our first study objective, we assess 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 hypothesize that the criterion validity estimates in our current study will be similar to our earlier findings. Our second objective explores reasons why not all of the PSI events overlap with those of VASQIP using our own nurse chart review. By reviewing cases with an AE identified by only one of the two methods, we have the opportunity to explore coding issues that may lead to PSI false positives and negatives. We hypothesized that substantive differences in AE definitions between the two methods, in addition to coding errors affecting the PSIs, may explain poor overlap in AE detection.

## Methods

We used a retrospective study design to compare surgical AEs detected by the VASQIP with AEs detected by the PSIs from October 1, 2002 through September 30, 2007 (FY03-FY07) in the VA. Following our assessment of PSI criterion validity, we sampled cases for chart review to understand why some AEs were detected by only one 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

The VASQIP was designed to identify complications occurring up to 30 days after surgery and does not attempt to measure the preventability of the AE. VASQIP nurse reviewers abstract pre-, intra-, and postoperative data from surgeries sampled according to the VASQIP sampling framework: the first 36 VASQIP-eligible surgeries in an eight day period are assessed, with restrictions on the number of high-volume, low-risk procedures.[9,11,13] VASQIP nurses document each postoperative AE and the date the AE occurred.[9] Definitions of postoperative events were consistent from FY03-07; we used the FY07 VASQIP data dictionary.

### AHRQ Patient Safety Indicators

The PSIs use International Classification of Disease 9th Revision Clinical Modification (ICD-9-CM) diagnosis and procedure codes from patient administrative data to detect potentially preventable inpatient AEs.[17,31] Each PSI algorithm uses eligibility criteria to exclude cases in which the patient's diagnoses or procedures indicate the condition was present-on-admission (POA) or they were at high risk of the event.[17,31] Applying the PSI software to administrative data generates a flag for cases that meet the eligibility criteria (i.e., in the PSI denominator) and a flag if the patient has the PSI event (i.e., PSI numerator).

### Comparison of PSI- versus VASQIP-detected Events

Five of the AHRQ surgical PSIs identify AEs similar to seven of the VASQIP postoperative complications (see Table 1). Based on our prior work, we recognized a priori that there were obvious differences in how the PSI and VASQIP AEs were defined;[23] 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 since VASQIP did not include diabetes in their definition of ARF.

Additionally, as the PSIs focus on AEs that occur only during hospitalization, we excluded VASQIP AEs if the VASQIP recorded date of the event occurred after the discharge date in the PTF.

### Data Sources

We ran version 4.1a of the AHRQ PSIs on VA administrative data from the Patient Treatment File (PTF) for FY03-07 (n=2,813,169 discharges).[32] A credentialed VA programmer (co-author SL) extracted PTF data from the Austin Information Technology Center (AITC). 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 healthcare system. The process of applying the AHRQ PSIs to VA data has been described in previous studies.[18,23]

We obtained FY03-07 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 or not any of seven relevant postoperative complications occurred, and data on all patient risk factors for these AEs. We also requested the dates and Current Procedural Terminology (CPT) codes characterizing surgical procedures.

Chart review data in our study came from the PSI Validation Study,[28] as well as from the VA electronic medical record (EMR) system, VistA.[33] Because there are some limitations to national VistAWeb access for researchers (e.g., radiology images are not always available) versus the VistA access available to local VA staff [34], it is plausible that VASQIP nurses with local VistA access reviewed a more complete patient record than the chart reviewers on our study.

### Merging Administrative and VASQIP Data

The VASQIP generates a record for each surgery reviewed, whereas the administrative data in the PTF is based on hospitalization. We followed an algorithm developed in our previous study to merge PTF and VASQIP data (i.e., the operation date recorded in VASQIP had to fall between the admission and discharge dates recorded in the PTF for cases to match).[23] Our final sample included 283,397 matched surgeries (97.5% of the VASQIP sample) that occurred during 268,771 hospitalizations in 117 VA hospitals (Figure 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 FY07 and apparent differences in definition of "outpatient" surgery between VASQIP and the PTF files (i.e., some of the unmatched VASQIP surgeries were labeled as inpatient in the VASQIP data but were considered as outpatient in VA administrative data).[35]

### Assessment of PSI Criterion Validity

We assessed criterion validity by comparing the presence of a PSI-detected AE to a VASQIP-detected AE among hospitalizations that met the PSI eligibility criteria in our

matched sample. We estimated sensitivity, specificity and positive predictive value (PPV) for each PSI using the VASQIP findings as the gold standard. Sensitivity evaluates the extent to which the PSI flagged cases with a true AE based on the total number of cases with AEs identified by the VASQIP; specificity examines the number of cases that were not flagged by a PSI and did not have an AE detected by the VASQIP; and PPV is a calculation of how many cases with a true AE, as determined by the VASQIP, were identified among all the PSI-flagged hospitalizations. We calculated 95% confidence intervals for each estimate and performed all analyses using SAS software version 9.1.

### Assessment of Disagreement – VASQIP-only Cases

For each PSI-VASQIP comparison group, we began by examining cases where only the VASQIP detected an AE (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 at least one or more 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 timeframe, as an explanation of disagreement in AE definition between the PSI 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 non-probability, or convenience, sampling to identify cases for chart review.[36,37] One of two trained research nurses (SM, KH) abstracted patient medical records in VistAWeb using the PSI-specific chart abstraction tools developed for the PSI Validation Study.[38] 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 at least 90% agreement for each PSI reviewed;[28] as a result, we did not repeat inter-rater reliability tests for this project. The chart review process involved searching the patients' medical record (i.e., VistAWeb) for the VASQIP-detected AE, and then examining the characteristics of the event to determine why the PSI failed to detect the AE. 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 (AB) reviewed all the findings.

### Assessment of Disagreement – PSI-only Cases

To assess PSI-only cases, we compared our findings to chart review results from the larger PSI Validation Study.[28] 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 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 the VASQIP (e.g., because of differences in AE definition). 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 at least one of the five PSIs we examined, and 6,100 of these were flagged. In comparison, the VASQIP found at least one of the AEs of interest in 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 PSI Criterion Validity

We calculated sensitivity, specificity and PPV for each of the five PSIs using the VASQIP-detected AEs as the gold standard (Table 2). The PSIs and the VASQIP detected similar rates of renal failure (PSI PMD and VASQIP ARF), respiratory failure (PSI Respiratory Failure and VASQIP Failure to Wean (FW) or Reintubation OR Unplanned Intubation (R/UI)), and sepsis (PSI Sepsis or VASQIP Systemic Sepsis (SS)) AEs; however, the PSIs flagged more than twice as many hospitalizations with either a pulmonary embolism or deep vein thrombosis (PSI Pulmonary Embolism or Deep Vein Thrombosis (PE/DVT) and VASQIP PE and DVT) AE, and half as many cases with wound dehiscence (PSI Wound Dehiscence (WD) and VASQIP WD) events. We hypothesized that our current study findings would be similar to the criterion validity results of our previous research using FY01 data and PSI version 3.0.[23] Our findings were comparable, other than an improvement in PPV of PE/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 five PSIs failed to flag some AEs detected by VASQIP because of coding errors, particularly for PE/DVT and PSI Wound Dehiscence (WD) and VASQIP WD (75% of the chart review sample for each PSI). Finally, we found many VASQIP-only (between 5–45% of the chart review sample) and PSI-only (between 69–100% of the chart review sample) cases that did not have discernible differences in AE definition or evidence of a coding error. We present detailed results for each PSI-VASQIP comparison, highlighting specific areas where improvements in definitions or methods may be warranted, in the Appendix.

## Discussion

Using a large sample of VA acute hospitalizations and five years of data, we compared AEs identified by PSI version 4.1a and the VASQIP and estimated PSI criterion validity. We found high specificity estimates for all five PSIs; low sensitivity for PSIs PMD, Sepsis, and WD (31–48%), and moderate sensitivity estimates for the remaining two PSIs, Respiratory Failure, and PE/DVT (65–68%). PSIs PE/DVT and Sepsis had low PPV estimates (31% and 44% respectively), and 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 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 based on 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 datasets. 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 two of the PSIs had more narrow definitions of AEs than VASQIP. Expanding the PSI numerator criteria to include cases reintubated on the same day of 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 two PSIs reflect the different purpose of the PSIs as compared to VASQIP: 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.[22,39] We expected that the increased attention to coding postoperative complications might improve the PPV of some of the PSIs as compared to our previous results. The improvement in PPV for PE/DVT from 22% to 31% may reflect improvements in the documentation or coding of PE and DVT events, a focus of many healthcare systems, including the VA, following the introduction of the Surgical Care Improvement Program (SCIP) in 2005.[40] 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 PSI did not agree on the detection of several of the AEs in our chart review sample. Studies have shown that conducting chart

review to identify rare AEs, particularly in complicated patients, can be challenging.[41] 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. Further, VASQIP data have been shown to be highly reliable.[11]

Given that the PSIs and the VASQIP are tools to assess quality of care, it may be useful for healthcare 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 ACS-NSQIP to measure surgical outcomes and target quality improvement efforts.[10,12] In addition, the CMS Hospital Compare website uses the AHRQ PSIs Respiratory Failure, PE/ DVT, and WD to report hospital performance.[25] Previous research has found poor overlap between surgical AEs detected by the PSIs and ACS-NSQIP;[29] 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 proposed that successful QI initiatives depend on a combination of administrative and chart review data.[42] Utilizing 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 ICD-10-CM 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 impacted 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 the 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 EMR. 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 upon our previous research with a larger dataset 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.[23]

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, while 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. Based on our findings, we recommend that future studies comparing PSI and VASQIP-detected events employ a VASQIP-trained nurse for chart review. Finally, as 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

1. Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. Health Technol Assess. 2001; 5(22):1–194. [PubMed: 11532239]

2. Morris JA Jr, Carrillo Y, Jenkins JM, et al. Surgical adverse events, risk management, and malpractice outcome: morbidity and mortality review is not enough. Ann Surg. Jun; 2003 237(6): 844–851. discussion 851–842. [PubMed: 12796581]

3. Institute of Medicine. To Err is Human: Building A Safer Health System. Washington, DC: National Academy Press; 1999.

4. Pronovost PJ, Goeschel CA, Marsteller JA, Sexton JB, Pham JC, Berenholtz SM. Framework for patient safety research and improvement. Circulation. Jan 20; 2009 119(2):330–337. [PubMed: 19153284]

5. Jha AK, Kuperman GJ, Teich JM, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. J Am Med Inform Assoc. May-Jun;1998 5(3):305–314. [PubMed: 9609500]

6. Naessens JM, Campbell CR, Huddleston JM, et al. A comparison of hospital adverse events identified by three widely used detection methods. Int J Qual Health Care. Aug; 2009 21(4):301–307. [PubMed: 19617381]

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. Feb; 2007 16(1):40–44. [PubMed: 17301203]

8. Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med. Feb 7; 1991 324(6):377–384. [PubMed: 1824793]

9. Khuri SF, Daley J, Henderson W, et al. 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. Ann Surg. Oct; 1998 228(4):491–507. [PubMed: 9790339]

10. Itani KM. Fifteen years of the National Surgical Quality Improvement Program in review. Am J Surg. Nov; 2009 198(5 Suppl):S9–S18. [PubMed: 19874939]

11. Davis CL, Pierce JR, Henderson W, et al. Assessment of the reliability of data collected for the Department of Veterans Affairs national surgical quality improvement program. J Am Coll Surg. Apr; 2007 204(4):550–560. [PubMed: 17382213]

12. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. Ann Surg. Sep; 2009 250(3):363–376. [PubMed: 19644350]

13. Henderson WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? Am J Surg. Nov; 2009 198(5 Suppl):S19–27. [PubMed: 19874930]

14. Davenport DL, Holsapple CW, Conigliaro J. Assessing surgical quality using administrative and clinical data sets: a direct comparison of the University HealthSystem Consortium Clinical Database and the National Surgical Quality Improvement Program data set. Am J Med Qual. Sep-Oct;2009 24(5):395–402. [PubMed: 19584374]

15. Hanauer DA, Englesbe MJ, Cowan JA Jr, Campbell DA. Informatics and the American College of Surgeons National Surgical Quality Improvement Program: automated processes could replace manual record review. J Am Coll Surg. Jan; 2009 208(1):37–41. [PubMed: 19228500]

16. Miller MR, Elixhauser A, Zhan C, Meyer GS. Patient Safety Indicators: using administrative data to identify potential patient safety concerns. Health Serv Res. Dec; 2001 36(6 Pt 2):110–132. [PubMed: 16148964]

17. Agency for Healthcare Research and Quality. [Accessed May 3, 2011] AHRQ Patient Safety Indicators Overview. AHRQ Homepage. 2011. http://www.qualityindicators.ahrq.gov/Modules/psi_overview.aspx

18. Rosen AK, Rivard P, Zhao S, et al. Evaluating the patient safety indicators: how well do they perform on Veterans Health Administration data? Med Care. Sep; 2005 43(9):873–884. [PubMed: 16116352]

19. Weiner BJ, Alexander JA, Baker LC, Shortell SM, Becker M. Quality improvement implementation and hospital performance on patient safety indicators. Med Care Res Rev. Feb; 2006 63(1):29–57. [PubMed: 16686072]

20. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA. Oct 8; 2003 290(14):1868–1874. [PubMed: 14532315]

21. Rivard P, Elixhauser A, Christiansen C, Shibei Z, Rosen A. Testing the association between patient safety indicators and hospital structural characteristics in VA and nonfederal hospitals. Med Care Res Rev. Jun; 2010 67(3):321–341. [PubMed: 19880671]

22. West AN, Weeks WB, Bagian JP. Rare adverse medical events in VA inpatient care: reliability limits to using patient safety indicators as performance measures. Health Serv Res. Feb; 2008 43(1 Pt 1):249–266. [PubMed: 18211528]

23. Romano PS, Mull HJ, Rivard PE, et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. Health Serv Res. Feb; 2009 44(1): 182–204. [PubMed: 18823449]

24. Agency for Healthcare Research and Quality. AHRQ Quality Indicator: National Quality Forum (NQF) Endorsed Composite and Individual Measures (June 2010). Rockville, MD: AHRQ; 2010.

25. Centers for Medicare and Medicaid Services (CMS). CMS Hospital Compare Website. 2011. http://www.cms.gov/HospitalQualityInits/11_HospitalCompare.asp2010

26. Henderson K, Recktenwald A, Reichley R, Bailey T, Waterman B, Diekemper R, Storey P, Ireland B, Dunagan W. Clinical validation of the AHRQ postoperative venous thromboembolism patient safety indicator. Jt Comm J Qual Patient Saf. Jul; 2009 35(7):370–376. [PubMed: 19634805]

27. Utter G, Cuny J, Sama P, Silver M, Zrelak P, Baron R, Drosler S, Romano P. Detection of postoperative respiratory failure: how predictive is the Agency for Healthcare Research and Quality's Patient Safety Indicator? J Am Coll Surg. Sep; 2010 211(3):347–354. e341–329. [PubMed: 20800191]

28. Rosen AK, Itani KM, Cevasco M, et al. Validating the patient safety indicators in the Veterans Health Administration: do they accurately identify true safety events? Med Care. Jan; 2012 50(1): 74–85. [PubMed: 21993057]

29. Cima RR, Lackore KA, Nehring SA, et al. How best to measure surgical quality? comparison of the Agency for Healthcare Research and Quality Patient Safety Indicators (AHRQ-PSI) and the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) postoperative adverse events at a single institution. Surgery. Aug 27.2011

30. Agency for Healthcare Research and Quality. AHRQ Guide to Patient Safety Indicators Version 3.1a. Rockville, MD: AHRQ; 2007.

31. McDonald, KM.; Romano, PS.; Geppert, JJ., et al. Measures of Patient Safety Based on Hospital Administrative Data: The Patient Safety Indicators. Stanford Evidence-based Practice Center, Agency for Healthcare Research and Quality; 2002.

32. Agency for Healthcare Research and Quality. Patient Safety Indicators Technical Specifications Version 4.1 (October 2010). Rockville, MD: AHRQ; 2010.

33. VA Office of Information and Technology. [Accessed May 3, 2011] VistAWeb Overview. 2011. http://vista.med.va.gov/vistaweb/

34. VA Information Resource Center (VIReC). [Accessed Aug 13, 2012] Veterans Health Information Systems and Technology Architecture (VistA), VistA Data for Research. 2012. http://www.virec.research.va.gov/VistA/Overview.htm

35. Neily J, Mills PD, Eldridge N, et al. Incorrect surgical procedures within and outside of the operating room. Arch Surg. Nov; 2009 144(11):1028–1034. [PubMed: 19917939]

36. Lavrakas, PJ., editor. Encyclopedia of Survey Research Methods. 1. Thousand Oaks, CA: SAGE Publications; 2008.

37. Gearing RE, Mian IA, Barber J, Ickowicz A. A methodology for conducting retrospective chart review research in child and adolescent psychiatry. J Can Acad Child Adolesc Psychiatry. Aug; 2006 15(3):126–134. [PubMed: 18392182]

38. Utter GH, Borzecki AM, Rosen AK, et al. Designing an abstraction instrument: lessons from efforts to validate the AHRQ patient safety indicators. Jt Comm J Qual Patient Saf. Jan; 2011 37(1):20–28. [PubMed: 21306062]

39. Hussey, PS.; Mattke, S.; Morse, L.; Ridgely, MS. Evaluation of the Use of AHRQ and Other Quality Indicators. Program Evaluation: Final Contract Report. Agency for Healthcare Research and Quality; Rockville, MD: Dec. 2007 AHRQ Publication No. 08-M012-EFhttp://www.ahrq.gov/about/evaluations/qualityindicators/

40. Bratzler DW. The Surgical Infection Prevention and Surgical Care Improvement Projects: promises and pitfalls. Am Surg. Nov; 2006 72(11):1010–1016. discussion 1021–1030, 1133–1048. [PubMed: 17120942]

41. Thomas EJ, Lipsitz SR, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. Ann Intern Med. Jun 4; 2002 136(11):812–816. [PubMed: 12044129]

42. Polk HC Jr, Lewis JN, Ali MK, Jones T, Robbins R. The complementary value of trained abstractors and surgeons in the more accurate assessment of surgical quality. Am J Med Qual. Nov-Dec;2010 25(6):444–448. [PubMed: 21115982]

## APPENDIX: Detailed Assessment of Disagreement

## Postoperative Physiologic and Metabolic Derangement (PMD) PSI #10 & VASQIP Acute Renal Failure (ARF)

### Overlap

There were 214 hospitalizations that matched both the PSI definition of PMD (a dialysis procedure and an acute renal failure diagnosis), and the VASQIP definition of acute renal failure (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 PSI, and found that 38 cases (17%) met the PSI dialysis criterion only and 131 other cases (57%) had an acute renal failure diagnosis 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, or 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 only specifies the *requirement* for dialysis, but did not meet the PSI definition because they could not be coded for dialysis. We found five VASQIP-only cases (25%) that fit the PSI PMD clinical definition but were lacking appropriate codes (i.e., PSI false negatives), and nine 'unable to determine' cases (45%). Seven of these cases did not have evidence of dialysis (or of refusing dialysis) in the medical record, and two had an unclear history of renal failure.

**PSI-Only**

In the second phase of chart review, our comparison with the PSI Validation Study chart review data revealed 4 PSI false positives, and 16 PSI-flagged true positive cases for re-review. We found 5 cases (31%) that were not flagged by VASQIP because of differences in definition (see Table 3). In four of these cases the VASQIP data showed that the patient had pre-operative dialysis, a VASQIP exclusion criterion. Although pre-operative dialysis is also a PSI exclusion criterion, these cases were flagged by the PSI and no dialysis was detected pre-operatively by the nurse reviewers in the patient medical record. As VASQIP chart review data includes contact with the patient, we conclude that these four cases are evidence of pre-operative dialysis that occurs outside of the VA and can only be detected through patient communication. We could not determine why the VASQIP did not identify an AE in the remaining 11 cases (69%).

## Postoperative Respiratory Failure PSI #11 & VASQIP Failure to Wean (FW) or VASQIP Reintubation/Unplanned Intubation (R/UI)

**Overlap**

Among the PTF-VASQIP matched sample, we identified 1,468 hospitalizations with Respiratory Failure and either the VASQIP FW or R/UI AEs. The PSI algorithm and the VASQIP definitions differed based on 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 codes, but were not flagged by the PSI because of the procedure's <u>timing</u>.

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, or 65%) (see Table 3). Of these, five of the VASQIP-only cases were coded by VASQIP as FW and did not have any of the PSI codes for respiratory failure, and eight were reintubated on the same day as the initial operation – these do not meet the PSI numerator definition (see Table 1). We also found six PSI false negatives (30%); five 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 one 'unable to determine' VASQIP-only case (5%) indicated the reintubation procedure occurred intra-operatively (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 seven PSI false positives and 15 PSI true positives. Given the differences in AE definition, we expected that some of the 15 PSI-only true positive cases would not have met the FW or Reintubation definitions; however, our chart review findings did not confirm this expectation. Only one 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 an AE in the remaining 14 cases (93%) (see Table 3).

# Pulmonary Embolism or Deep Vein Thrombosis (PE/DVT) PSI #12 & VASQIP PE or VASQIP DVT

**Overlap**

We found 1,034 hospitalizations that were PSI-flagged for PE/DVT, (defined as a secondary diagnosis of PE or DVT) and that had a 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, or 75%, of the VASQIP-only cases were PSI false negatives because the administrative data did not code the PE or DVT. We could not determine why the PSI did not detect the VASQIP AE in the five remaining cases (25%).

**PSI-Only**

Of the 2,289 PSI-only cases, we found 479 coded with a secondary diagnosis of PE and 1,881 with a 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/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/DVT cases (70%) could not determine why the VASQIP did not identify the PE or DVT (see Table 3).

# Postoperative Sepsis PSI#13 & VASQIP Systemic Sepsis

**Overlap**

The definition of sepsis varies between the PSI and the VASQIP which may account for why only 175 cases, out of 397 hospitalizations flagged by the PSI and **570** hospitalizations with AEs identified by the 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 (i.e., fever). In June of 2004, VASQIP revised their sepsis definition to include a diagnosis of Systemic Inflammatory Response Syndrome (SIRS); the PSI algorithm includes SIRS, 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 eleven cases (55%) where the VASQIP and PSI definitions did not align (see Table 3). While the cases had SIRS, 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 cases (40%) had clinical documentation of sepsis in the medical record but were not flagged by the PSI because of coding errors. Nurse chart reviewers were unable to determine why the PSI did not detect one 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 four (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).

## Postoperative Wound Dehiscence (WD) PSI#14 & VASQIP Dehiscence

### Overlap

Of the 1,254 hospitalizations with a VASQIP-detected AE, and the 544 hospitalizations flagged by the PSI, 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 a wound reclosure procedure but were missing the appropriate code. Nurse reviewers did not find any cases of differences in AE definition. The remaining five 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 the VASQIP did not identify the AE (see Table 3).
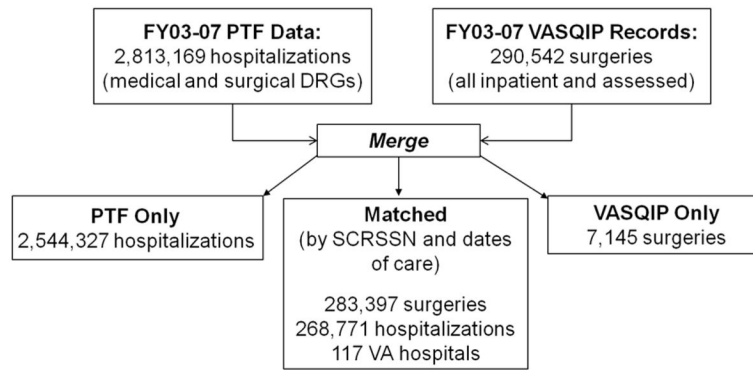
**Figure 1. Results of Merge between VA Administrative and VASQIP Datasets**
NOTES:

VASQIP=VA Surgical Quality Improvement Program

PTF=Patient Treatment File

DRG=Diagnosis-Related Group

SCRSSN=Scrambled social security number

**Table I**

Comparison of AHRQ PSI version 4.1a and VASQIP Data Dictionary AE Definitions

| AHRQ PSI v4.1a Numerator Criteria | VASQIP 2007 Data Dictionary Definition of Postoperative Complication |
|---|---|
| **PSI#10 Postoperative Physiologic and Metabolic Derangement (PMD):** | **VASQIP Acute Renal Failure (ARF):** In a patient who did not require dialysis preoperatively, worsening of renal dysfunction postoperatively requiring either |
| **1** Various diabetes codes in secondary diagnosis field | • hemodialysis, |
| **OR** | • peritoneal dialysis, |
| **2** Dialysis as a secondary procedure AND various necrosis/renal failure codes in any secondary diagnosis field | • hemofiltration, |
| | • hemodiafiltration or |
| | • ultrafiltration |
| | If the patient refuses dialysis the answer is Yes to this variable, because he/she did require dialysis. |
| *NOTE: Part 1 of the PSI algorithm does not align with the VASQIP AE* | |
| **PSI#11 Postoperative Respiratory Failure:** | • **VASQIP Failure to Wean (FW) OR On Ventilator > 48 hours:** |
| **1** Acute respiratory failure code in any secondary diagnosis field | Total duration of ventilator-assisted respirations during postoperative hospitalization was >48 hours. This can occur at any time during the 30-day period postoperatively. |
| **OR** | This time assessment is CUMULATIVE, not necessarily consecutive. Ventilator-assisted respirations can be via endotracheal tube, nasotracheal tube, or tracheostomy tube. |
| **2** Reintubation (96.04) or mechanical ventilation (96.70, 96.71, 96.72), codes in any secondary procedure field in the following time frame: | • **VASQIP Reintubation OR Unplanned Intubation (R/UI) for Respiratory/Cardiac Failure:** |
| • (96.04) one or more days after the major operating room procedure code | Patient required placement of an endotracheal 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. In patients who were intubated for their surgery, unplanned intubation occurs after they have been extubated after surgery. In patients who were not intubated during surgery, intubation at any time after their surgery is considered unplanned. |
| • (96.70 or 96.71) two or more days after the major operating room procedure code | |
| • (96.72) zero or more days after the major operating room procedure code | |
| *NOTE: Part of the PSI#11 algorithm overlaps with part of the VASQIP AE definition for "Cardiac Arrest Requiring CPR"* | |
| **PSI#12 Postoperative Pulmonary Embolism (PE)/Deep Vein Thrombosis (DVT):** | • **VASQIP Pulmonary Embolism (PE):** |
| **1** Code for pulmonary embolism in any secondary diagnosis field | 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 pulmonary embolism or a positive pulmonary arteriogram or positive CT angiogram or positive Spiral CT exam. Treatment usually consists of: |
| **OR** | – Initiation of anticoagulation therapy |
| **2** Codes for deep vein thrombosis in any secondary diagnosis field | – Placement of mechanical interruption (e.g. Greenfield Filter), for patients in whom anticoagulation is contraindicated or already instituted. |
| | • **VASQIP Deep Vein Thrombosis (DVT)/ Thrombophlebitis:** |

| AHRQ PSI v4.1a<br>Numerator Criteria | VASQIP 2007 Data Dictionary<br>Definition of Postoperative Complication |
|---|---|
| | Definition: the identification of a new blood clot or thrombus within the venous system, which may be coupled with inflammation. This d...<br>Definition: the identification of a new blood clot or thrombus within the venous system, which may be coupled with inflammation. This d...<br>Definition: the identification of a new blood clot or thrombus within the venous system, which may be coupled with inflammation. This d... |
| **PSI#13 Postoperative Sepsis:**<br>Discharges with various sepsis codes in any secondary diagnosis field (i.e., strep septicemia, staph, pneumococcal septicemia, septicemia due to anaerobes, septic shock, postoperative shock, septicemia due to gram negative organism, hemophilus influenza, E. coli, or SIRs due to infectious process with or without organ dysfunction) | **VASQIP Systemic Sepsis (SS):**<br><br>**1**    **Sepsis:** Sepsis is the systemic response to infection. Report this variable if the patient has clinical signs and symptoms of Systemic Inflammatory Response Syndrome (SIRS). SIRS is a widespread inflammatory response to a variety of severe clinical insults. This syndrome is clinically recognized by the presence of **two** or more of the following:<br><br>   • Temp >38 degrees C or <36 degrees C<br>   • HR >90 bpm<br>   • RR >20 breaths/min or PaCO2 <32 mmHg(<4.3 kPa)<br>   • WBC >12,000 cell/mm3, <4000 cells/mm3, or >10% immature (band)forms<br>   • Anion gap acidosis: this is defined by either:<br><br>      –   $[Na + K] - [Cl + HCO3$ (or serum CO2)]. If this number is greater than 16, then an anion gap acidosis is present.<br>      –   $Na - [Cl + HCO3$ (or serum CO2)]. If this number is greater than 12, then an anion gap acidosis is present.<br><br>     And **one** of the following:<br><br>   • positive blood culture<br>   • clinical documentation of purulence or positive culture from any site thought to be causative<br><br>**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.<br><br>***From 2001- June, 2004 the VASQIP definition of sepsis was as follows:***<br>  If the primary physician or the chart states that the patient had systemic sepsis within the 30 days postoperatively, choose from the following choices for sepsis. If neither is present, follow these definitions and choose the most applicable:<br><br>**1**    **Sepsis:** Definitive evidence of infection, plus evidence of a systemic response to infection. This systemic response is manifested by **two or more** of the following conditions:<br><br>   **a.**   Temp > 38 degrees C or < 36 degrees C<br>   **b.**   HR > 90 bpm<br>   **c.**   RR > 20 breaths/min or PaCO2 < 32 mmHg (<4.3 kPa)<br>   **d.**   WBC > 12,000 cell/mm3, < 4000 cells/mm3, or >10% immature (band) forms<br><br>**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 |

| AHRQ PSI v4.1a Numerator Criteria | VASQIP 2007 Data Dictionary Definition of Postoperative Complication |
|---|---|
| | are on inotropic or vasopressor agents may not be hypotensive at the time that perfusion abnormalities are measured. are on inotropic or vasopressor agents may not be hypotensive at the time that perfusion abnormalities are measured. |
| **PSI#14 Postoperative Wound Dehiscence (WD):** Discharges with code for reclosure of postoperative disruption of abdominal wall (54.61) in any procedure field | • **VASQIP Dehiscence:** Separation of the layers of a surgical wound, which may be partial or complete, with disruption of the fascia. |

NOTES:

The VASQIP data dictionary definitions were valid from 2002–2007 with the exception of sepsis, as described in the table.

PSI exclusion criteria are not presented;

PSI AEs are detected prior to discharge while VASQIP AEs can occur up to 30 days postoperatively, regardless of inpatient status.

**Table II**

Criterion Validity Estimates of PSI version 4.1a and VASQIP AEs (FY03-07) and PSI version 3.0 and VASQIP AEs (FY01)

| Comparison Group | | Current Results for PSI version 4.1a, FY03-07 Data | | | | | Previous Results for PSI version 3.0, FY01 Data | |
|---|---|---|---|---|---|---|---|---|
| PSI | VASQIP AE | Cases eligible for PSI (%)* | VASQIP-Detected AEs (%)** | PSI- Detected AEs (%)** | PSI Sensitivity§ (95% CI) | PSI PPV† (95% CI) | PSI Sensitivity§ (95% CI) | PSI PPV† (95% CI) |
| #10 PMD | ARF | 131,711 (49%) | 442 (0.3%) | 322 (0.2%) | 48% (44–53%) | 66% (61–72%) | 48% (36–61%) | 63% (48–75%) |
| #11 RF | FW / R/UI | 109,916 (41%) | 2,151 (2.0%) | 2,047 (1.9%) | 68% (66–70%) | 72% (70–74%) | 63% (57–67%) | 68% (62–73%) |
| #12 PE/DVT | PE / DVT | 253,090 (95%) | 1,588 (0.6%) | 3,323 (1.3%) | 65% (63–67%) | 31% (30–33%) | 56% (50–63%) | 22% (19–25%) |
| #13 Sepsis | SS | 47,523 (18%) | 570 (1.2%) | 397 (0.8%) | 31% (27–35%) | 44% (39–49%) | 37% (27–49%) | 45% (33–57%) |
| #14 WD | Dehiscence | 74,251 (28%) | 1,254 (1.7%) | 544 (0.7%) | 31% (29–34%) | 72% (68–76%) | 29% (24–34%) | 72% (63–80%) |

Notes: Parentheses contain 95% confidence intervals; PSI specificity estimates exceeded 99.1%; Italicized values indicate a significant difference between estimates in our previous study (Romano, 2009).

*
Percent of the PTF-VASQIP matched sample of 268,771 hospitalizations.

**
Percent of the hospitalizations eligible for the corresponding PSI.

§
Sensitivity represents the proportion of the VASQIP-detected AEs flagged by the PSI.

†
PPV - positive predictive value - represents the proportion of PSI-flagged hospitalizations with a VASQIP-detected AE.

PMD= Postoperative Physiologic and Metabolic Derangement

RF= Postoperative Respiratory Failure

PE= Postoperative Pulmonary Embolism

DVT= Postoperative Deep Vein Thrombosis

WD= Postoperative Wound Dehiscence

ARF= Acute Renal Failure

FW=Failure to Wean

R/UI= Reintubation/Unplanned Intubation W

SS=Systemic Sepsis

**Table III**

Reasons for Disagreement in Surgical AE Detection between PSI version 4.1a and VASQIP

| Comparison Group | | Total VASQIP-only Cases Reviewed* | Chart Review Results for VASQIP - Only Cases | | | Total PSI- Only True Positives Re-Reviewed** | Chart Review Results for PSI-Only Cases | | |
|---|---|---|---|---|---|---|---|---|---|
| PSI | VASQIP AE | | Differences in Definition (%) | PSI False Negative (%) | Unable to Determine (%) | | Differences in Definition (%) | PSI False Positive (%)† | Unable to Determine (%) |
| #10 PMD | ARF | 20/228 | 6 (30%) | 5 (25%) | 9 (45%) | 16/108 | 5 (31%) | NA | 11 (69%) |
| #11 RF | FW | 20/683 | 13 (65%) | 6 (30%) | 1 (5%) | 15/579 | 1 (7%) | NA | 14 (93%) |
| | R/UI | | | | | | | | |
| #12 PE/DVT | PE | 20/554 | - | 15 (75%) | 5 (25%) | 20/2,289 | 6 (30%) | NA | 14 (70%) |
| | DVT | | | | | | | | |
| #13 Sepsis | SS | 20/395 | 11 (55%) | 8 (40%) | 1 (5%) | 15/222 | 4 (27%) | NA | 11 (73%) |
| #14 WD | Dehiscence | 20/862 | - | 15 (75%) | 5 (25%) | 20/152 | - | NA | 20 (100%) |

*
This column shows the number of hospitalizations reviewed out of the number of PSI-eligible hospitalizations with an AE detected by VASQIP only.

**
This column shows the number of PSI Validation Study true positive hospitalizations re-reviewed out of the number of PSI-eligible hospitalizations in our sample with an AE detected by the PSI only.

†
We matched PSI-only cases with PSI Validation Study true positive chart review results, and limited our chart review to cases that were deemed true AEs in the Validation Study. As a result, we do not have any PSI false positive cases.

PMD= Postoperative Physiologic and Metabolic Derangement

RF= Postoperative Respiratory Failure

PE= Postoperative Pulmonary Embolism

DVT= Postoperative Deep Vein Thrombosis

WD= Postoperative Wound Dehiscence

ARF= Acute Renal Failure

FW=Failure to Wean

R/UI= Reintubation/Unplanned Intubation

SS=Systemic Sepsis