NATIONAL QUALITY FORUM

Measure Evaluation 4.1
January 2010

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The sub-criteria and most of the footnotes from the evaluation criteria are provided in Word comments and will appear if your cursor is over the highlighted area (or in the margin if your Word program is set to show revisions in balloons). Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized): Complete all yellow highlighted areas of the form. Evaluate the extent to which each sub-criterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: If there is no TAP or workgroup, the SC also evaluates the sub-criteria (yellow highlighted areas).

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the sub-criterion, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few sub-criteria as indicated)

(for NQF staff use) NQF Review #: PS-006-10 NQF Project: Patient Outcomes Measures: Phases I and II

MEASURE DESCRIPTIVE INFORMATION

De.1 Measure Title: Risk Adjusted Surgical Site Infection Outcome Measure

De.2 Brief description of measure: This is a hospital based, risk adjusted, case mix adjusted surgical site infection measure of adults 18 years of age and over.

De.1-2 Type of Measure: outcome

De.3 If included in a composite or paired with another measure, please identify composite or paired measure n/a

De.4 National Priority Partners Priority Area: population health, safety
De.5 IOM Quality Domain: effectiveness, efficiency, equity, safety
De.6 Consumer Care Need: Getting Better

CONDITIONS FOR CONSIDERATION BY NQF

Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:

A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.

A.1 Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)? Yes

A.2 Indicate if Proprietary Measure (as defined in measure steward agreement):

A.3 Measure Steward Agreement: agreement signed and submitted

A.4 Measure Steward Agreement attached:

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
</table>

C. The intended use of the measure includes both public reporting and quality improvement.

- Purpose: public reporting, quality improvement 0,0,0,

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
</table>

D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement.

- Testing: No, testing will be completed within 12 months
- Have NQF-endorsed measures been reviewed to identify if there are similar or related measures? Yes

(for NQF staff use) Have all conditions for consideration been met?

Staff Notes to Steward (if submission returned):

Staff Notes to Reviewers (Issues or questions regarding any criteria):

Staff Reviewer Name(s):

---

**TAP/Workgroup Reviewer Name:**

- Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. **Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)**

**1. IMPORTANCE TO MEASURE AND REPORT**

**1a. High Impact**

(for NQF staff use) **Specific NPP goal:**

1a.1 Demonstrated High Impact Aspect of Healthcare: affects large numbers, severity of illness, frequently performed procedure, a leading cause of morbidity/mortality, patient/societal consequences of poor quality, high resource use

1a.2

1a.3 Summary of Evidence of High Impact: Surgical site infections (SSIs) are the second leading cause of nosocomial infections. Approximately 290,000 SSIs were diagnosed in the United States in 2002, resulting in 8,207 associated deaths. [1] The mortality rate of patients with SSIs is approximately 2-12 times that of patients who do not have a SSI. [2, 3] Surgical site infections result in an additional 7-10 days of hospitalization for each postoperative infection per patient. [4] Furthermore, SSIs represent a significant financial burden to the healthcare system. The attributable direct cost per infection ranges from $6,000 to $29,000 depending on the operative procedure and the type of infecting pathogen. [3-5] Estimates indicate that SSIs accounted for $3.45-10.07 billion in direct costs in 2007. [6]

1a.4 Citations for Evidence of High Impact:

1b. Opportunity for improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure:

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
SSI rates are highly variable by institution. ACS NSQIP uses clinical, audited, third-party collection, and risk adjusted data. An analysis of ACS NSQIP data shows that O/E ratios for SSI range from 0 to 3.01 for all participating hospitals. The interquartile range for O/E ratios is 0.70-1.24, and the 10th and 90th percentile O/E ratios were 0.45 and 1.52, respectively. These statistics demonstrate the significance of the performance gap in SSI outcomes across hospital providers.

1b.3 Citations for data on performance gap:
The data cited above is unpublished, obtained from an internal analysis of ACS NSQIP data. However, these gaps have been repeatedly demonstrated since the inception of the program.

1b.4 Summary of Data on disparities by population group:
Certain patient-related factors have been associated with an increased risk of SSI, including: advanced age, obesity, [4, 7-9] and gender [10], as well as characteristics associated with certain population groups such as hyperglycemia/diabetes, dyspnea, hypoxia, ASA classification>2, smoking, [4, 7, 8, 13] and alcoholism.

1b.5 Citations for data on Disparities:

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population): Measurement of SSI as a risk-adjusted outcome.
adjusted outcome is highly relevant to the general surgical population. Despite a wealth of modern infection-control practices (sterilization, antimicrobial prophylaxis, antisepsis and barriers, etc.), SSIs persist in causing a significant number of morbidity and mortality events among hospitalized patients. Since risk of SSI increases with certain patient preoperative factors, a risk-adjusted measure is necessary to ensure that hospitals are receiving an accurate benchmark of their performance based on their patient case-mix. Finally our analyses demonstrate that evidence based NQF endorsed process measures have little to no correlation with clinical risk adjusted outcomes. Evidence that demonstrates the significance and relevance of SSI to the population may be found in RCTs, observational trials, cohort studies, etc. (See below)

1c.2-3. **Type of Evidence:** evidence based guideline, observational study, cohort study, randomized controlled trial, systematic synthesis of research, meta-analysis, expert opinion

1c.4 **Summary of Evidence** (as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome):
A number of existing guidelines and recommendations exist that detail measures that can help prevent SSIs. However, there is little evidence to show that these measures correlate with SSI outcomes. The most highly recommended processes for prevention of SSI are outlined below (SCIP). There are six evidence-based preventive measures for SSI. These processes were selected based on literature detailing the effect of these measures on surgical site infection outcomes.

1. Administer prophylactic antibiotics within one hour prior to surgical incision (vancomycin and fluoroquinolones should be administered 2 hours prior to surgery. (A-I)
2. Select the appropriate antimicrobial prophylaxis based upon published guidelines (A-I)
3. Discontinue use of the prophylactic antibiotic within 24 hours after surgery (48 hours for cardiothoracic procedures in adult patients) (A-I)
4. Remove hair only if it interferes with the operation. If hair removal is necessary, use clippers instead of razors (A-II)
5. Monitor and maintain glucose levels (<200mg/DL) in cardiothoracic surgery patients (including non-diabetic patients) on postoperative days one and two. (A-I)
6. Maintain normothermia perioperatively for patients undergoing colorectal surgery. (B-I)

Occurrence of surgical site infections is likely multi-factorial and there are a number of additional processes that are also highly recommended for implementation based on their potential effect on improving outcomes. The problem being faced currently is the degree to which already-identified SSI processes affect outcomes in the real world setting. Analyses within ACS NSQIP to date show little to no correlation between performance on the SCIP process measures and risk-adjusted outcomes. Thus, an alternative metric for evaluating surgical patient care is to use an SSI outcome-based performance measure, as opposed to measures based on processes.

1c.5 **Rating of strength/quality of evidence** (also provide narrative description of the rating and by whom):
There are no ratings for an SSI outcome measure simply because it is the outcome of interest. The SCIP process measures cited above are generally level I -II evidence. Ratings for SSI related processes are not applicable to this application, however, they are available upon request.

1c.6 **Method for rating evidence:** Rating method adapted from the Canadian Task Force on Health Examination. These ratings apply to the process measures under "Summary of Evidence"

<table>
<thead>
<tr>
<th>Category/grade Definition</th>
<th>Strength of recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Good evidence to support a recommendation for use</td>
</tr>
<tr>
<td>B</td>
<td>Moderate evidence to support a recommendation for use</td>
</tr>
<tr>
<td>C</td>
<td>Poor evidence to support a recommendation</td>
</tr>
</tbody>
</table>

Quality of evidence
I  Evidence from one or more properly randomized, controlled trial
II Evidence from one or more well-designed clinical trial, without randomization; from cohort or case-
control analytic studies (preferably from >1 center); from multiple time series; or from dramatic results from uncontrolled experiments

III Evidence from opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees

1c.7 Summary of Controversy/Contradictory Evidence: Contradictory evidence exists on the effect of process measures on outcomes. In a highly controlled setting (controlled clinical study) high performance on SCIP measures is related to high performance on outcomes, but in an observational setting, there is little correlation between process and outcomes. As mentioned above, ACS NSQIP data were used to conduct a cross-sectional study (unpublished data) to determine whether adherence with Surgical Care Improvement Project (SCIP) process measures correlates with risk-adjusted ACS NSQIP outcomes. Thirty-day risk-adjusted outcomes after colorectal surgery, including mortality, serious morbidity, morbidity, surgical site infections, venous thromboembolism (VTE), and cardiac events, at ACS NSQIP hospitals that submitted performance on seven process measures to The Joint Commission between July 1, 2007, and June 30, 2008, were correlated with process measure compliance. Multivariable forward step-wise logistic regression models were constructed to assess 30-day morbidity and mortality adjusted for patient comorbidities, operative risk factors, and process measure compliance. The results of the regression models showed that SCIP process measure compliance was not an important predictor of ACS NSQIP risk-adjusted outcomes. The above study illustrates that occurrence of SSI is probably multifactorial and it is quite likely that the process measures identified by SCIP for prevention of SSI do not accurately reflect ALL of the processes that account for risk-adjusted SSI outcomes.

1c.8 Citations for Evidence (other than guidelines):


S.1 Do you have a web page where current detailed measure specifications can be obtained?
S.2 If yes, provide web page URL:

2a. Precisely Specified

2a.1 Numerator Statement (Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome):
The outcome of interest is a hospital-specific risk-adjusted Deep Incisional Surgical Site Infection (SSI) or Organ/Space SSI as defined by American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) within 30 days of any ACS NSQIP listed (CPT) surgical procedure.

2a.2 Numerator Time Window (The time period in which cases are eligible for inclusion in the numerator):
Targeted events within 30 days of the operation are included.

2a.3 Numerator Details (All information required to collect/calculate the numerator, including all codes, logic, and definitions):
Deep Incisional SSI: Deep Incision SSI is an infection that occurs within 30 days after the operation and the infection appears to be related to the operation and infection involves deep soft tissues (for example, fascial and muscle layers) of the incision and at least one of the following: Purulent drainage from the deep incision but not from the organ/space component of the surgical site; A deep incision spontaneously dehisces or is deliberately opened by a surgeon when the patient has at least one of the following signs or symptoms: fever (> 38 C), localized pain, or tenderness, unless site is culture-negative; An abscess or other evidence of infection involving the deep incision is found on direct examination, during reoperation, or by histopathologic or radiologic examination; Diagnosis of a deep incision SSI by a surgeon or attending physician.

Organ/Space SSI: Organ/Space SSI is an infection that occurs within 30 days after the operation and the infection involves any part of the anatomy (for example, organs or spaces), other than the incision, which was opened or manipulated during an operation and at least one of the following: Purulent drainage from a drain that is placed through a stab wound into the organ/space; Organisms isolated from an aseptically obtained culture of fluid or tissue in the organ/space; An abscess or other evidence of infection involving the organ/space that is found on direct examination, during reoperation, or by histopathologic or radiologic examination; Diagnosis of an organ/space SSI by a surgeon or attending physician.

2a.4 Denominator Statement (Brief, text description of the denominator - target population being measured):
Patients undergoing any ACS NSQIP listed (CPT) surgical procedure. (see separate list of ACS NSQIP CPT codes

2a.5 Target population gender: Female, Male
2a.6 Target population age range: Any patient greater than or equal to 18 years of age

2a.7 Denominator Time Window (The time period in which cases are eligible for inclusion in the denominator):
Data are derived from a systematic sample collected over a one year period constructed to as to meet sample size requirements specified for the measure.

2a.8 Denominator Details (All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions):
Cases are collected so as to match ACS NSQIP inclusion and exclusion criteria, thereby permitting valid application of ACS NSQIP model-based risk adjustment.

2a.9 Denominator Exclusions (Brief text description of exclusions from the target population): Trauma and transplant surgeries are excluded as are surgeries not on the ACS NSQIP CPT list as eligible for selection. Patients who are ASA 6 (brain-death organ donor) are not eligible surgical cases.

2a.10 Denominator Exclusion Details (All information required to collect exclusions to the denominator, including all codes, logic, and definitions):
A patient who is admitted to the hospital with acute trauma and has surgery for that trauma is excluded

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
though any operation performed after the patient has been discharged from the trauma stay can be included. A patient who is admitted to the hospital for a transplant and has a transplant procedure and any additional surgical procedures during the transplant hospitalization will be excluded, tough any operation performed after the patient has been discharged from the transplant stay is eligible for selection. A patient classified as ASA Class 6 is not eligible for inclusion.

2a.11 Stratification Details/Variables (All information required to stratify the measure including the stratification variables, all codes, logic, and definitions):
There is no stratification of the measure

2a.12-13 Risk Adjustment Type: case-mix adjustment

2a.14 Risk Adjustment Methodology/Variables (List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method):
From 271,368 patient records in the 2008 ACS NSQIP Data file; 254,200 acceptable records from 211 hospitals (mean/hospital=1,205) were analyzed. Records were excluded either because of missing values for critical variables or because the primary CPT code could not be categorized into 1 of the 136 pre-established CPT “Groups”. These categorizations have been defined and implemented for risk adjustment in previously published research.*

Surgical site infection (SSI) was defined as either a deep SSI or an organ space SSI, according to ACS NSQIP definitions (concordant with CDC). Of the 254,200 patients, 4,532 (1.8%) experienced an SSI event as defined. CPT Group was originally considered a categorical variable but, because of frequent empty cells, which precluded logistic model convergence (quasi-complete separation), CPT Group was converted to continuous risk variable. This was accomplished by making the categorical Group variable a single predictor for SSI and invoking the Firth penalized likelihood method in the logistic modeling software (SAS PROC LOGISTIC). The patient-based predicted log odds from this model was then used as a continuous predictor in subsequent logistic models which also included the standard predictors.

Step-wise logistic regression (P<0.05 for inclusion), which selected from a total of 26 NSQIP predictors, identified 12 predictors for inclusion in the model. In order of inclusion these variables were: Log Odds CPT Group, ASA Class, Wound Class, Age Group, Steroid Use, BMI Class, Smoking, Disseminated Cancer, Emergent, Pneumonia, Weight Loss, and Alcohol Use. The c-statistic was 0.810 and the Hosmer-Lemeshow was 0.043. Because of the very large sample sizes studied here, a statistically significant Hosmer-Lemeshow statistic is not considered informative with respect to calibration.

Using only the first three selected variables (Log Odds CPT Group, ASA Class, and Wound Class), the c-statistic was 0.806 and the Hosmer-Lemeshow was 0.002. The use of these three predictors for modeling was further evaluated. Using a 95% confidence interval for the ratio of observed to expected events (O/E), this three-variable logistic model identified 50 statistical outliers (26 low outliers and 24 high outliers). When the same three variables were used in a random intercept, fixed slope, hierarchical model (SAS PROC GLIMMIX) using only the fixed portion of the prediction equation (NOBLUP option), 49 outliers were detected (22 low outliers and 27 high outliers). Thus, using a 95% confidence interval, logistic and hierarchical models identified between 11% and 13% of hospitals as high outliers. When the logistic model parameters were applied to an independent validation data set (the 2007 Data file composed of 201,837 patients) after coding CPT Groups with log odds derived from the original 1-variable model on 2008 data, the c-statistic was essentially unchanged (c-statistic=0.801).

A GEE (generalized estimating equations) approach (SAS PROC GENMOD) with compound symmetry was used to estimate the intraclass correlation (ICC) which is reported in GENMOD as the exchangeable working correlation. The ICC was 0.00156. The relationship between sample size, the ICC, and reliability is defined as: N-R / [ICC(1 - R)] - R / (1 - R), where N is the required number patients per hospital and R is reliability. Based on the estimated ICC, patients per hospital to achieve reliability levels of 0.3, 0.4, 0.5, 0.6, and 0.7 are 275, 428, 641, 961, and 1,495, respectively.

For the table detailing risk factors, odds ratios, and parameters for the logistic model, please see attachment (Parsimonious Model for SSI.doc)

For initial year(s) of measure use, ACS NSQIP data-derived model parameters will be used to construct risk-adjusted O/E ratios for participating hospitals. Once data from measure-participating hospitals is substantial, models will derived from that data.
References utilizing CPT groups


2a.15-17 Detailed risk model available Web page URL or attachment: Attachment Parsimonious Model for SSI.doc

Type of Score: ratio

Interpretation of Score: better quality = lower score

Calculation Algorithm: For data collected during the one year time interval at each hospital: (a) O = the number of observed adverse events at the hospital; (b) using parameters from the applicable model derived logistic equation, compute predicted event probabilities for each patient in the hospital’s data set; (c) the sum of these predicted probabilities defines E; (d) compute the hospital’s O/E ratio and applicable confidence intervals.

Describe the method for discriminating performance (e.g., significance testing): The default methodology for discrimination performance will be based on the computed 95% CI for the O/E ratio. If the interval is above, and does not overlap 1.0, the hospital is identified as having performance significantly worse than expected. If the interval is below, and does not overlap 1.0, the hospital is identified as having performance significantly better than expected. Depending on programmatic objectives, the implementing organization could also opt for outlier status being defined by percentile rank, for example, in upper or lower distributional deciles of O/E ratios.

Sampling (Survey) Methodology: For each data collection year, hospitals would need to estimate their number of qualifying surgeries. Based on that denominator and the required sample size to achieve reliability of 0.4 (see Risk-adjustment Methodology section), hospitals would take a systematic sample (e.g., every 3rd qualifying case), to achieve the minimum sample size. In the event that the required sample size can not be achieved, hospitals would collect data on all eligible patients.

Data Source: Documentation of original self-assessment, Management data, pharmacy data

Data source/data collection instrument: ACS NSQIP Data File


Level of Measurement/Analysis: Facility/Agency
2a.36-37 Care Settings (Check the setting(s) for which the measure is specified and tested)
Hospital

2a.38-41 Clinical Services (Healthcare services being measured, check all that apply)
Clinicians: Physicians (MD/DO), Clinicians: Pharmacist, Clinicians: Nurses, Clinicians: PA/NP/Advanced Practice Nurse, Clinicians: Respiratory Therapy, Clinicians: Dietician/Nutritional professional

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample (description of data/sample and size): See Risk-adjustment Methodology in Specifications. Models were constructed using a large sample derived from the ACS NSQIP database for 2008.

2b.2 Analytic Method (type of reliability & rationale, method for testing): See Risk-adjustment Methodology in Specifications. Reliability was determined using ICCs estimated by SAS PROC GENMOD.

2b.3 Testing Results (reliability statistics, assessment of adequacy in the context of norms for the test conducted):
See Risk-adjustment Methodology in Specifications.

The relative variation between hospitals defined by the intra-class correlation coefficient (ICC) for hospitals can be estimated for continuous outcomes using linear mixed models, but the within-hospital variation needed to calculate ICCs is not routinely estimated for dichotomous outcomes. Hence, the usual measure of ICC based on a latent variable formulation using the standard logistic distribution was estimated. The between-hospital variation component of the ICC was estimated from SAS PROC GENMOD regressing the composite outcome on the significant predictors for SSI. Together with procedure volumes, these ICCs were entered into the following equation to estimate reliability:

\[ R = \frac{n \cdot ICC}{1 + (n - 1) \cdot ICC} \]

where R is the reliability, n is the case load per hospital and ICC is the intra-class correlation.

There are no definitive criteria for what level of reliability is acceptable, but it is proposed to be similar to inter-rater reliability standards used for assessing survey instruments.

<table>
<thead>
<tr>
<th>RELIABILITY ESTIMATE</th>
<th>INTERPRETATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00-0.20</td>
<td>Slight</td>
</tr>
<tr>
<td>0.21-0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41-0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61-0.80</td>
<td>Substantial</td>
</tr>
<tr>
<td>0.81-1.00</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

The ICC was estimated at 0.00156. Using a minimum acceptable reliability for SSI of 0.4, the proportions of hospitals likely to have a “minimally acceptable” reliability estimate are as follows: 89.3% of all U.S. hospitals and 92.4% of ACS NSQIP hospitals meet the 0.4 reliability requirement.

Table 1. Estimates of Procedure Volume Required to Achieve Specified Measure Reliability, and Proportions of U.S. Hospitals and ACS NSQIP Hospitals Meeting the Volume Requirements.

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Required Cases %U.S. Hosp Mts Rqmnt+</th>
<th>%NSQIP Hosp Mts Rqmnt+</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>275</td>
<td>94.0</td>
</tr>
<tr>
<td>0.4</td>
<td>428</td>
<td>92.4</td>
</tr>
<tr>
<td>0.5</td>
<td>641</td>
<td>89.3</td>
</tr>
<tr>
<td>0.6</td>
<td>961</td>
<td>84.8</td>
</tr>
<tr>
<td>0.7</td>
<td>1495</td>
<td>79.7</td>
</tr>
</tbody>
</table>

*Based on volume data from the 2005 National Inpatient Survey and inflated to account for outpatient procedures.
### 2c. Validity testing

#### 2c.1 Data/sample (description of data/sample and size): See Risk-adjustment Methodology in Specifications. Models were constructed using a large sample derived from the ACS NSQIP database for 2008.

#### 2c.2 Analytic Method (type of validity & rationale, method for testing): See Risk-adjustment Methodology in Specifications. C-statistics and Hosmer-Lemeshow P-values for the developmental data set were computed; c-statistics were computed for an independent validation data set base on 2007 data.

#### 2c.3 Testing Results (statistical results, assessment of adequacy in the context of norms for the test conducted): See Risk-adjustment Methodology in Specifications. Model validity (a similar c-statistic, discrimination) was demonstrated when the 2008 model was applied to 2007 data.

### 2d. Exclusions Justified

#### 2d.1 Summary of Evidence supporting exclusion(s):
The ACS NSQIP CPT list includes all surgeries that would be appropriate for measurement of quality and it would be unreasonable to provide documentation on the thousands of inapplicable codes. In addition, we have explicitly excluded surgeries related to trauma, transplant, and ASA Class 6 (brain-death organ donors). The ASA 6 exclusion as regards prediction of postoperative mortality and morbidity does not require explanation. As this measure is intended to apply generally to all hospitals doing surgery, inclusion of trauma and transplant cases, which tend to be directed towards metropolitan or regional centers, could adversely affect the efficacy of risk-adjustment (non-overlap of these types of cases across hospitals might be profound).

#### 2d.2 Citations for Evidence:
As exclusions are based on reasoned argument rather empirical findings neither published evidence nor research findings are provided.

#### 2d.3 Data/sample (description of data/sample and size): N/A

#### 2d.4 Analytic Method (type analysis & rationale): N/A

#### 2d.5 Testing Results (e.g., frequency, variability, sensitivity analyses): N/A

### 2e. Risk Adjustment for Outcomes/ Resource Use Measures

#### 2e.1 Data/sample (description of data/sample and size): The data sample is derived from the most recent ACS NSQIP Data file. The SSI model used 254,400 patient records. Future models can be constructed using the most recent Data file. If this measure is adopted by sufficient numbers of non-NSQIP hospitals re-modeling can be based on data from the broader sample of hospitals.

#### 2e.2 Analytic Method (type of risk adjustment, analysis, & rationale): Preliminary risk-adjustment models were constructed for these developmental purposes using step-wise logistic regression. Compared to hierarchical models this methodology poses fewer convergence problems, has step-wise variable-selection methodology, and we have found that it provides nearly identical risk-adjustment as random intercept hierarchical models. Odds ratios and parameters reported here are derived from hierarchical model methodology applied to the predictor set established using step-wise logistic regression methods.

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
2e.3 Testing Results (risk model performance metrics):
See Risk-adjustment Methodology in Specifications. A parsimonious predictor set was constructed from the full step-wise set. Step-wise logistic regression (P<0.05 for inclusion), which selected from a total of 26 predictors, identified 12 predictors for inclusion in the model. In order of inclusion these variables were: Log Odds CPT Group, ASA Class, Wound Class, Age Group, Steroid Use, BMI Class, Smoking, Disseminated Cancer, Emergent, Pneumonia, Weight Loss, and Alcohol Use. The c-statistic was 0.810 and the Hosmer-Lemeshow was 0.043. Because of the very large sample sizes studied here, a statistically significant Hosmer-Lemeshow statistic is not considered informative with respect to calibration. Using only the first three selected variables (Log Odds CPT Group, ASA Class, and Wound Class), which is being advocated as the risk-adjustment model, the c-statistic was 0.806 and the Hosmer-Lemeshow was 0.002. The use of these three predictors for modeling was further evaluated. Using a 95% confidence interval for the ratio of observed to expected events (O/E), this three-variable logistic model identified 50 statistical outliers (26 low outliers and 24 high outliers). When the same three variables were used in a random intercept, fixed slope, hierarchical model (SAS PROC GLIMMIX) using only the fixed portion of the prediction equation (NOBLUP option), 49 outliers were detected (22 low outliers and 27 high outliers). Thus, using a 95% confidence interval, logistic and hierarchical models identified between 11% and 13% of hospitals as high outliers.

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use (description of data/sample and size): See Risk-adjustment Methodology in Specifications.

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (type of analysis & rationale):

The default methodology for discrimination performance will be based on the computed 95% CI for the O/E ratio. If the interval is above, and does not overlap, 1.0, the hospital is identified as having performance significantly worse than expected. If the interval is below, and does not overlap, 1.0, the hospital is identified as having performance significantly better than expected. Depending on programmatic objectives, the implementing organization could also opt for outlier status being defined by percentile rank, for example, in upper or lower distributional percentiles of O/E ratios.

2f.3 Provide Measure Scores from Testing or Current Use (description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance):

See Risk-adjustment strategy Testing Results

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample (description of data/sample and size): The only sources of data are those indicated above. This measure will require mostly clinical data (electronic or paper records), with administrative data added only as necessary. The advantage of clinical data versus administrative or claims data in identifying risk-adjusted outcomes is exemplified in the study by Steinberg et al (2008). The study compared comorbidities collected and postsurgical complications from the ACS NSQIP database and the University HealthSystem Consortium (UHC). Comorbidities per patient were identified twice as often in the UHC system, while there was a discordance of 26% in identifying complications (UHC complication rate, 2% vs. ACS NSQIP complication rate, 28%). Using administrative or claims data may result in significant differences in risk-adjusted outcomes than using clinical data.


2g.2 Analytic Method (type of analysis & rationale):
see above

2g.3 Testing Results (e.g., correlation statistics, comparison of rankings):

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [K19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 73%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.
see above

2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results (scores by stratified categories/cohorts): Measure is not stratified; measure is case mix adjusted.

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:

n/a

TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria for Scientific Acceptability of Measure Properties?

Steering Committee: Overall, to what extent was the criterion, Scientific Acceptability of Measure Properties, met?
Rationale:

2

3. Usability

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (Evaluation criteria)

3a. Meaningful, Understandable, and Useful Information

3a.1 Current Use: in use

3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large) (If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years):
Not currently in use for a public reporting initiative

3a.3 If used in other programs/initiatives (If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years):
ACS NSQIP - www.acsnsqip.org

Testing of Interpretability (Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement)

3a.4 Data/sample (description of data/sample and size): Although this specific measure has not been formally tested for interpretability, the ACS NSQIP has been using similar O/E ratios to measure outcomes in the program for over 15 years from its inception in the VA. The success of this program and the satisfaction of participants provide evidence of interpretability of this outcome measure. Hospitals are able to compare their observed complications with their number of expected complications in a ratio that provides a very straightforward measure of performance, while simultaneously being complex enough to adjust for each hospital’s case mix. Hospitals are also able to benchmark their performance against other participating hospitals, so that better and worse performers are easily identified.

This risk-adjusted and benchmarked measure provides enormous motivation for hospitals to see their outcomes improve. A recent analysis (Hall et al, 2009) has shown that 66% of ACS NSQIP hospitals improved their risk-adjusted mortality and 82% of hospitals improved their risk-adjusted complication rates. The effect on avoided complications is also significant, as the analysis demonstrates that between 250 and 500 complications per hospital were avoided in 2007.

The data for the above study was ACS NSQIP data collected over 3 years (2005-2007) from 118 hospitals. This measure will be reported annually.

### NQF #PSM-006-10

3a.5 Methods (e.g., focus group, survey, QI project):
The data for the above study was ACS NSQIP data collected over 3 years (2005-2007) from 118 hospitals.

3a.6 Results (qualitative and/or quantitative results and conclusions):
See above section on “Testing of Interpretability”: Data Sample

3b/3c. Relation to other NQF-endorse measures

3b.1 NQF # and Title of similar or related measures:
NQF #0299, Surgical Site Infection (Centers for Disease Control, Centers for Medicare and Medicaid Services).

(for NQF staff use) Notes on similar/related endorsed or submitted measures:

3b. Harmonization

If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population):

3b.2 Are the measure specifications harmonized? If not, why?
The measures were not harmonized due to the different intents for the measures. The CDC SSI measure was developed for surveillance and employs stratification of cases rather than risk adjustment. Thus, raw unadjusted rates are used in each stratified sample. The current ACS NSQIP SSI measure is directed towards accountability and employs an inclusive random sample with risk adjustment. To reiterate, because of the separate aims of the measures (surveillance versus accountability) the measures have not been harmonized. Another reason is that the specifications for the CDC measure include surgical site infections occurring within one year after the procedure if an implant is in place. We are not able to harmonize with this additional numerator due to the difficulty in collecting accurate data on SSIs up to one year after the procedure.

3c. Distinctive or Additive Value

3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorse measures:
This proposed measure is related to the endorsed NQF measure, but is different in several aspects. As briefly mentioned above in the prior item, first, the CDC measure appears geared more toward surveillance than quality improvement, since it is based on the reporting of stratified percentages of SSI’s per procedure. This reporting method is not as conducive to quality improvement as O/E ratio reporting as proposed here. The CDC measure is stratified according to several risk factors, whereas our proposed measure is risk-adjusted. The experience of the ACS NSQIP is that risk-adjustment is a more robust method for hospitals to accomplish targeted quality improvement, as it facilitates comparisons of each hospital’s own performance over time and benchmarking with other hospitals. In addition, the specifications, including follow-up horizon, are rigorously implemented in this proposed measure to optimize implementation, control burden, and provide information targeted to specific quality improvement opportunities.

5.1 Competing Measures If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), describe why it is a more valid or efficient way to measure quality:

<table>
<thead>
<tr>
<th>TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria for Usability?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steering Committee: Overall, to what extent was the criterion, Usability, met? Rationale:</td>
</tr>
<tr>
<td>Extent to which the required data are readily available, retrievable without undue burden, and can be Evaluated</td>
</tr>
</tbody>
</table>

Rationale:

### 4. FEASIBILITY

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
### 4a. Data Generated as a Byproduct of Care Processes

4a.1-2 How are the data elements that are needed to compute measure scores generated?

Data generated as byproduct of care processes during delivery, coding/abstraction performed by someone other than person obtaining original information.

### 4b. Electronic Sources

4b.1 Are all the data elements available electronically? (elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims)

No

4b.2 If not, specify the near-term path to achieve electronic capture by most providers.

A completely electronic medical record will be needed to capture the risk factors that enter into the model. In addition, a software module (currently available to ACS NSQIP subscribers) will be required to transfer information from the EMR to a measure submission database.

### 4c. Exclusions

4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?

No

4c.2 If yes, provide justification.

### 4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences

4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.

Based upon experience with ACS NSQIP data collection, there are very few problems with errors or inaccuracies. Data collectors in the ACS NSQIP receive extensive training and support for accurate data collection. In addition, data collectors are audited for inter-rater reliability and are held to a 95% or better concordance rate for all variables. Additionally, chart audits have been planned in accordance with CMS stipulations for measure participants who are not ACS NSQIP participants.

### 4e. Data Collection Strategy/Implementation

4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/implementation issues:

ACS NSQIP has been open to subscription by private sector hospitals since 2004. Ten years prior to this time the program was implemented in the U.S. Department of Veterans Affairs. Thus we have long term experience with the data collection and operational use of the O/E ratio for quality improvement and benchmarking on which this measure is based. Historically, the use of trained data collectors within ACS NSQIP and a comprehensive support system has resulted in high reliability of data and very few problems with missing data. Participants in the program are required to assign a dedicated person for data collection to ensure reliable assessment of clinical data.

Data definitions are continually evaluated and inter-rater reliability audits are regularly performed. ACS NSQIP has placed a very high value on accuracy of data collection while maintaining a sample size large enough for statistical modeling and keeping within regulations for patient confidentiality. The methodology of our program has been highly successful with increasing numbers of participants every year, and measureable improvements in surgical outcomes over time based on the O/E ratios for mortality and various post surgical complications. Due to the much smaller number of variables needed for participation in this measure than in the full program, we expect that hospitals that are not ACS NSQIP participants will also be able to achieve highly reliable results.

4e.2 Costs to implement the measure (costs of data collection, fees associated with proprietary costs of data collection, fees associated with proprietary
Using a conservative estimate, approximately .125 to .333 of a FTE will be needed to collect the data for the measure. There are no fees associated with this measure. Hospitals do not have to be ACS NSQIP hospitals in order to participate in the proposed measure.

4e.3 Evidence for costs:
Costs are based upon an estimate from ACS NSQIP data collection, in which one FTE can collect 1600 cases per year, but is required to collect a large number of variables, as well as 30-day follow up which can consume many hours. In contrast, this measure does not require many variables, and sample size is such that reliable results can be achieved after collection of 200-500 cases.

4e.4 Business case documentation: A business case has not been developed for this measure; however, literature results show that the direct costs for each surgical site infection can range from $6,000 to $29,000 and require an extra 7 days of hospitalization per infected patient.

TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria for Feasibility?

Steering Committee: Overall, to what extent was the criterion, Feasibility, met?
Rationale:

RECOMMENDATION
(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.

Steering Committee: Do you recommend for endorsement?
Comments:

CONTACT INFORMATION
Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
American College of Surgeons | 633 N. Saint Clair St. | Chicago | Illinois | 60611

Co.2 Point of Contact
Karen | Richards, Director, Division of Research and Optimal Patient Care | krichards@facs.org | 312-202-5282

Measure Developer If different from Measure Steward
Co.3 Organization
American College of Surgeons | 633 N. Saint Clair St. | Chicago | Illinois | 60611

Co.4 Point of Contact
Karen | Richards, Director, Division of Research and Optimal Patient | krichards@facs.org | 312-202-5282

Co.5 Submitter If different from Measure Steward POC
Karen | Richards, Director, Division of Research and Optimal Patient Care | krichards@facs.org | 312-202-5282

Co.6 Additional organizations that sponsored/participated in measure development

ADDITIONAL INFORMATION
Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members’ names and organizations. Describe the members’ role in measure development.
American College of Surgeons, Area of Continuous Quality Improvement

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
This group used ACS NSQIP data to develop the statistical risk-adjusted model on which this measure is based. The workgroup also reviewed and summarized the literature that supports the importance of using this measure to as a tool to improve surgical quality.

### Measure Developer/Steward Updates and Ongoing Maintenance

**Year the measure was first released:**

**Month and Year of most recent revision:**

**What is your frequency for review/update of this measure?**

**When is the next scheduled review/update for this measure?**

**Copyright statement/disclaimers:**

**Type of measure** *Outcome*

Four conditions must be met before a proposed measure may be considered and evaluated for suitability as voluntary consensus standards:

A. The measure steward is a governmental organization or a Measure Steward Agreement is signed. Public domain only applies to governmental organizations. All non-government organizations must sign a Measure Steward Agreement even if measures are made publicly and freely available.

Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)? *

Yes

Please check if either of the following apply

**Proprietary measure**

**Measure Steward Agreement** *

Agreement will be signed and submitted prior to or at the time of measure submission

**B.** The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. *

Yes, information will be provided in the contact section (in the Additional tab)

C. The intended use of the measure includes both public reporting and quality improvement.

**Purpose** *

Public reporting

Internal quality improvement

Additional purposes

None

D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 24 months of endorsement.

**Testing** *

No, testing will be completed within 24 months

Have NQF-endorsed® measures been reviewed to identify if there are similar or related measures? *

If there are similar or related measures, be sure to address those items in the Usability tab.

Yes
Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status - patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Clinically necessary measure exclusions are identified and must be:
• supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
  AND
• a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
  AND
• precisely defined and specified:
  − if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
## Parsimonious Hierarchical Model for SSI
### Data from 1/1/2008 – 12/31/2008 (2008 PUF)

<table>
<thead>
<tr>
<th>Predictor Set</th>
<th>Risk Factor</th>
<th>Odds Ratio</th>
<th>Lower 95% Boundary</th>
<th>Upper 95% Boundary</th>
<th>Parameter Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intercept</td>
<td>-0.5151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Log Odds CPT Group (continuous)</td>
<td>2.599</td>
<td>2.517</td>
<td>2.683</td>
<td>0.9550</td>
</tr>
<tr>
<td>2</td>
<td>ASA Class:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-Severe Disturbance vs. 1/2-No/Mild Disturbance</td>
<td>1.403</td>
<td>1.313</td>
<td>1.499</td>
<td>0.3386</td>
</tr>
<tr>
<td>2</td>
<td>4/5- Life Threatening/Moribund vs. 1/2-No/Mild</td>
<td>1.335</td>
<td>1.202</td>
<td>1.483</td>
<td>0.2888</td>
</tr>
<tr>
<td>2</td>
<td>Disturbance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Wound Class:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3/4 Contaminated/Dirty/Infected vs. 1/2 Clean/Clean</td>
<td>1.454</td>
<td>1.359</td>
<td>1.556</td>
<td>0.3745</td>
</tr>
<tr>
<td></td>
<td>Contaminated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observations = 254,200; Hospitals = 211; Patients with SSI event = 4,532, Rate = 1.8%, (logistic) c-statistic = 0.806, HL = 0.002.