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** #: PSM-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.
1.1-2 **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:	**NQF Staff**
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:**	**A** **Y**☐ **N**☐

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	**B** Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** public reporting, quality improvement 0,0,0,	**C** Y☐ N☐
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. D.1**Testing:** No, testing will be completed within 12 months D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
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) **1a. High Impact**	Eval Rating
(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:** 1. Anderson, D.J., et al., Strategies to prevent surgical site infections in acute care hospitals. Infect Control Hosp Epidemiol, 2008. 29 Suppl 1: p. S51-61. 2. Engemann, J.J., et al., Adverse clinical and economic outcomes attributable to methicillin resistance among patients with Staphylococcus aureus surgical site infection. Clin Infect Dis, 2003. 36(5): p. 592-8. 3. Kirkland, K.B., et al., The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs. Infect Control Hosp Epidemiol, 1999. 20(11): p. 725-30.	**1a** C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

4. Klevens, R.M., et al., Estimating health care-associated infections and deaths in U.S. hospitals, 2002. Public Health Rep, 2007. 122(2): p. 160-6.
5. Olsen, M.A., et al., Hospital-associated costs due to surgical site infection after breast surgery. Arch Surg, 2008. 143(1): p. 53-60; discussion 61.
6. Scott II, R.D., The Direct Medical Costs of Healthcare-Associated Infections in U.S. Hospitals and the Benefits of Prevention. 2009, Centers for Disease Control and Prevention: Atlanta.

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 percentile 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, [7] [8] obesity, [4, 7-9] and gender [10], as well as characteristics associated with certain population groups such as hyperglycemia/diabetes, [4, 7, 8, 11] dyspnea, [8] hypoxia, [11] ASA classification>2, [8, 12]smoking, [4, 7, 8, 13] and alcoholism.

1b.5 Citations for data on Disparities:

4. Anderson, D.J., et al., Strategies to prevent surgical site infections in acute care hospitals. Infect Control Hosp Epidemiol, 2008. 29 Suppl 1: p. S51-61.
5. Olsen, M.A., et al., Hospital-associated costs due to surgical site infection after breast surgery. Arch Surg, 2008. 143(1): p. 53-60; discussion 61.
6. Scott II, R.D., The Direct Medical Costs of Healthcare-Associated Infections in U.S. Hospitals and the Benefits of Prevention. 2009, Centers for Disease Control and Prevention: Atlanta.
7. Mangram, A.J., et al., Guideline for prevention of surgical site infection, 1999. Hospital Infection Control Practices Advisory Committee. Infect Control Hosp Epidemiol, 1999. 20(4): p. 250-78; quiz 279-80.
8. Neumayer, L., et al., Multivariable predictors of postoperative surgical site infection after general and vascular surgery: results from the patient safety in surgery study. J Am Coll Surg, 2007. 204(6): p. 1178-87.
9. Anaya, D.A. and E.P. Dellinger, The obese surgical patient: a susceptible host for infection. Surg Infect (Larchmt), 2006. 7(5): p. 473-80.
10. Rogers, M.A., et al., Increased risk of infection and mortality in women after cardiac surgery related to allogeneic blood transfusion. J Womens Health (Larchmt), 2007. 16(10): p. 1412-20.
11. Cheadle, W.G., Risk factors for surgical site infection. Surg Infect (Larchmt), 2006. 7 Suppl 1: p. S7-11.
12. Culver, D.H., et al., Surgical wound infection rates by wound class, operative procedure, and patient risk index. National Nosocomial Infections Surveillance System. Am J Med, 1991. 91(3B): p. 152S-157S.
13. Theadom, A. and M. Cropley, Effects of preoperative smoking cessation on the incidence and risk of intraoperative and postoperative complications in adult smokers: a systematic review. Tob Control, 2006. 15(5): p. 352-8.

1b C☐ P☐ M☐ N☐

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-

1c C☐ P☐ M☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
oEfficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Comment [k5]: 4 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., ... [1]

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 cardio-thoracic 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"

Category/grade Definition Strength of recommendation
Strength of recommendation
A Good evidence to support a recommendation for use
B Moderate evidence to support a recommendation for use
C Poor evidence to support a recommendation

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-

N☐

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

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***):** 1. Arrowsmith, V.A., et al., Removal of nail polish and finger rings to prevent surgical infection. Cochrane Database Syst Rev, 2001(4): p. CD003325.
2. Auerbach, A.D., Chapter 20. Prevention of Surgical Site Infections. Making Health Care Safer: A Critical Analysis of Patient Safety Practices.
3. Barie, P.S., Surgical site infections: epidemiology and prevention. Surg Infect (Larchmt), 2002. 3 Suppl 1: p. S9-21.
4. Belda, F.J., et al., Supplemental perioperative oxygen and the risk of surgical wound infection: a randomized controlled trial. JAMA, 2005. 294(16): p. 2035-42.
5. Bratzler, D.W. and P.M. Houck, Antimicrobial prophylaxis for surgery: An advisory statement from the National Surgical Infection Prevention Project. The American Journal of Surgery, 2005. 189(4): p. 395-404.
6. Bratzler, D.W. and D.R. Hunt, The surgical infection prevention and surgical care improvement projects: national initiatives to improve outcomes for patients having surgery. Clin Infect Dis, 2006. 43(3): p. 322-30.
7. Bucher, P., et al., Mechanical bowel preparation for elective colorectal surgery: a meta-analysis. Arch Surg, 2004. 139(12): p. 1359-64; discussion 1365.
8. Chow, T.T. and X.Y. Yang, Ventilation performance in operating theatres against airborne infection: review of research activities and practical guidance. J Hosp Infect, 2004. 56(2): p. 85-92.
9. Chura, J.C., A. Boyd, and P.A. Argenta, Surgical site infections and supplemental perioperative oxygen in colorectal surgery patients: a systematic review. Surg Infect (Larchmt), 2007. 8(4): p. 455-61.
10. Dellinger, E.P., Preventing surgical-site infections: the importance of timing and glucose control. Infect Control Hosp Epidemiol, 2001. 22(10): p. 604-6.
11. Dellinger, E.P., Increasing inspired oxygen to decrease surgical site infection: time to shift the quality improvement research paradigm. JAMA, 2005. 294(16): p. 2091-2.
12. Dellinger, E.P., Roles of temperature and oxygenation in prevention of surgical site infection. Surg Infect (Larchmt), 2006. 7(Suppl 3): p. s27-32.
13. Dellinger, E.P., What is the ideal time for administration of antimicrobial prophylaxis for a surgical procedure? Ann Surg, 2008. 247(6): p. 927-8.
14. Dellinger, E.P. and D.A. Anaya, Infectious and immunologic consequences of blood transfusion. Crit Care, 2004. 8 Suppl 2: p. S18-23.
15. Dharan, S. and D. Pittet, Environmental controls in operating theatres. J Hosp Infect, 2002. 51(2): p. 79-84.
16. Digison, M.B., A review of anti-septic agents for pre-operative skin preparation. Plast Surg Nurs,

2007. 27(4): p. 185-9; quiz 190-1.
17. Edmiston, C.E., Jr., et al., Comparative of a new and innovative 2% chlorhexidine gluconate-impregnated cloth with 4% chlorhexidine gluconate as topical antiseptic for preparation of the skin prior to surgery. Am J Infect Control, 2007. 35(2): p. 89-96.
18. Edwards, P.S., A. Lipp, and A. Holmes, Preoperative skin antiseptics for preventing surgical wound infections after clean surgery. Cochrane Database Syst Rev, 2004(3): p. CD003949.
19. Greif, R., et al., Supplemental perioperative oxygen to reduce the incidence of surgical-wound infection. Outcomes Research Group. N Engl J Med, 2000. 342(3): p. 161-7.
20. Kluytmans, J.A. and H.F. Wertheim, Nasal carriage of Staphylococcus aureus and prevention of nosocomial infections. Infection, 2005. 33(1): p. 3-8.
21. Latham, R., et al., The association of diabetes and glucose control with surgical-site infections among cardiothoracic surgery patients. Infect Control Hosp Epidemiol, 2001. 22(10): p. 607-12.
22. Leaper, D., Effects of local and systemic warming on postoperative infections. Surg Infect (Larchmt), 2006. 7 Suppl 2: p. S101-3.
23. Niel-Weise, B.S., J.C. Wille, and P.J. van den Broek, Hair removal policies in clean surgery: systematic review of randomized, controlled trials. Infect Control Hosp Epidemiol, 2005. 26(12): p. 923-8.
24. Perl, T.M., et al., Intranasal mupirocin to prevent postoperative Staphylococcus aureus infections. N Engl J Med, 2002. 346(24): p. 1871-7.
25. Pryor, K.O., et al., Surgical site infection and the routine use of perioperative hyperoxia in a general surgical population: a randomized controlled trial. JAMA, 2004. 291(1): p. 79-87.
26. Tanner, J., K. Moncaster, and D. Woodings, Preoperative hair removal: a systematic review. J Perioper Pract, 2007. 17(3): p. 118-21, 124-32.
27. Webster, J. and S. Osborne, Preoperative bathing or showering with skin antiseptics to prevent surgical site infection. Cochrane Database Syst Rev, 2007(2): p. CD004985.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
There are no ratings for an SSI outcome measure simply because it is the outcome of interest.

1c.10 Clinical Practice Guideline Citation: N/A
1c.11 National Guideline Clearinghouse or other URL: N/A

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
N/A

1c.13 Method for rating strength of recommendation (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:**
N/A

1c.14 Rationale for using this guideline over others:
N/A

TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria **for** *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** Y☐ N☐
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

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 involved 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 appears to be related to 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

2a-specs
C☐
P☐
M☐
N☐

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

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. If surgeries do not appear in the list of ACS NSQIP CPT codes, they are not 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 the136 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

Hall BL, Hamilton BH, Richards K, et al. Does Surgical Quality Improve in the American College of Surgeons National Surgical Quality Improvement Program: An Evaluation of All Participating Hospitals. Ann Surg, in press.

Hall BL, Hsiao EY, Majercik S, et al. The impact of surgeon specialization on patient mortality: examination of a continuous Herfindahl-Hirschman index. Ann Surg 2009; 249(5):708-16.

Cohen ME, Bilimoria KY, Ko CY, Hall BL. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. J Am Coll Surg 2009; 208(6):1009-16.

Schilling PL, Dimick JB, Birkmeyer JD. Prioritizing quality improvement in general surgery. J Am Coll Surg 2008; 207(5):698-704.

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

2a.18-19 Type of Score: ratio
2a.20 Interpretation of Score: better quality = lower score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
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.

2a.22 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.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
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.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Documentation of original self-assessment, Management data, pharmacy data

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
ACS NSQIP Data File

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL
https://acsnsqip.org/puf/PufRequestHomepage.aspx

2a.29-31 Data dictionary/code table web page URL or attachment: URL
https://acsnsqip.org/documents_section/documents_appendix_c-2.pdf

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
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 = nICC/(1 + (n -1)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.

RELIABILITY ESTIMATE	INTERPRETATION
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Excellent

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.

Reliability	RequiredCases	%U..S.HospMtgRqrmnt*	%NSQIPHospMtg Rqrmnt+
0.3	275	94.0	94.8
0.4	428	89.3	92.4
0.5	641	84.8	82.0
0.6	961	79.7	65.9
0.7	1495	73.1	32.7

*Based on volume data from the 2005 National Inpatient Survey and inflated to account for outpatient procedures.

2b
C☐
P☐
M☐
N☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

+Based on ACS NSQIP Data file 2008 and inflated to account for procedures that might be excluded for over-representation

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.

2c
C☐
P☐
M☐
N☐

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

2d
C☐
P☐
M☐
N☐
NA☐

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 254,200 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.

2e
C☐
P☐
M☐
N☐
NA☐

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. 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 ca ... [2]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR rationale/data support no risk adjustment.

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

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

2f
C☐
P☐
M☐
N☐

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.

Steinberg, S.M., et al., Comparison of risk adjustment methodologies in surgical quality improvement. Surgery, 2008. 144(4): p. 662-7; discussion 662-7.

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

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

2g
C☐
P☐
M☐
N☐
NA☐

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. 75%) 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	**2h** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria **for** *Scientific Acceptability of Measure Properties*?	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties***, met?** **Rationale:**	*2* C☐ P☐ M☐ N☐
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)	Eval Rating
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. 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.	**3a** C☐ P☐ M☐ N☐

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

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-endorsed 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.	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed 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:**	**3c** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria **for** *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be	Eval

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare).

Comment [k26]: 5. Demonstration that the measure is superior to competing measures – new submissions and/or endorsed measures (e.g., is a more valid or efficient way to measure).

implemented for performance measurement. (evaluation criteria)	Rating
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,	**4a** C☐ P☐ M☐ N☐
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.	**4b** C☐ P☐ M☐ N☐
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.**	**4c** C☐ P☐ M☐ N☐ NA☐
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.	**4d** C☐ P☐ M☐ N☐
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*	**4e** C☐ P☐ M☐ N☐

Comment [KP27]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP28]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP29]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP30]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP31]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

measures)**:** 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*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** **C☐** **P☐** **M☐** **N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐** **N☐** **A☐**

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

Clifford Ko
Karen Richards
Bruce Hall
Mark Cohen
Mehul Raval
Mira Shiloach
Angela Ingraham
Stanley Frencher

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.

Ad.2 If adapted, provide name of original measure: n/a
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released:
Ad.7 Month and Year of most recent revision:
Ad.8 What is your frequency for review/update of this measure?
Ad.9 When is the next scheduled review/update for this measure?

Ad.10 Copyright statement/disclaimers: UPDATED CONDITIONS SECTION:
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

Ad.11 -13 Additional Information web page URL or attachment: Attachment BP Guideline SSI-633888790679944712.pdf

Date of Submission (*MM/DD/YY*)**:**

Page 3: [1] Comment [k5] **Karen Pace** **10/5/2009 8:59:00 AM**

4 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.

Page 11: [2] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. 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)					
Predictor Set	Risk Factor	Odds Ratio	Lower 95% Boundary	Upper 95% Boundary	Parameter Estimate
	Intercept				-0.5151
1	Log Odds CPT Group (continuous)	2.599	2.517	2.683	0.9550
2	ASA Class: 3-Severe Disturbance vs. 1/2-No/Mild Disturbance 4/5- Life Threatening/Moribund vs. 1/2-No/Mild Disturbance	 1.403 1.335	 1.313 1.202	 1.499 1.483	 0.3386 0.2888
3	Wound Class: 3/4 Contaminated/Dirty/Infected vs. 1/2 Clean/Clean Contaminated	1.454	1.359	1.556	0.3745
Observations = 254,200; Hospitals = 211; Patients with SSI event = 4,532, Rate = 1.8%, (logistic) c-statistic = 0.806, HL = 0.002.					